Mental models and meaning:

An analysis into the validity of Philip
Johnson-Laird's adoption of Peircean iconism

M T Fowler

PhD   2015

MENTAL MODELS AND MEANING:
AN ANALYSIS INTO THE VALIDITY OF PHILIP JOHNSON-LAIRD'S ADOPTION OF PEIRCEAN ICONISM

MARK THOMAS FOWLER

A thesis submitted in partial fulfilment of the requirements of the Manchester Metropolitan University for the degree of Doctor of Philosophy

Department of Interdisciplinary Studies
the Manchester Metropolitan University
2015

Acknowledgements

My sincere thanks go to the following:

My supervisory team – Dr Phil Hutchinson, Dr Michael Loughlin and Dr Neil Dagnall. As well as to Dr Kathy Kinmond.

Mo Randjbar and Prof. Dick Hartley, for their support in my day job.

My mum, brother, nan and granddad.

Most of all, my thanks go to Kate.

Without the help, support, encouragement and input from all of you, I would not have been able to complete this work.

Abstract

The recent trend within cognitive science toward the adoption of a *de re* account of semantic content has been accepted by some theorists without due consideration, an oversight that has generated a number of logical inconsistencies in the literature. This thesis endeavours to shed light on this omission, and the effects thereof, by scrutinising the account of one such theorist: The psychologist Philip Johnson-Laird. Herein, Johnson-Laird's introduction of a Peircean iconism is subjected to a thorough analysis and, in this way, an important example of the problematic tensions created by a failure to address the substantial nature of a theory, before its wholesale integration into a philosophy, is revealed. The philosophic method of analysis and synthesis is applied to the arguments and hypotheses of Philip Johnson-Laird, thereby demonstrating the base assumptions that constitute his account of cognition and the relationships between these assumptions. Investigation of the results of this method identifies logical inconsistency present in the hypothesis and, following Popper and Quine, logical inconsistency is taken as a disproof of a theory. The study concludes that, although Johnson-Laird's updated philosophy was able to dissolve historical arguments against his theory, his use of Peircean iconism is inconsistent in its own right and produces serious tensions with aspects of his extant philosophy. As such, it is concluded that Johnson-Laird's externalist semantics fails and it is recommended that his account of semantics be reconsidered.

Key terms: Philip Johnson-Laird; Charles Sanders Peirce; cognitive science; the representational theory of mind; mental representation; correspondence theory of meaning; Peircean iconism; philosophic analysis and synthesis.

Table of contents

Chapter Two: Critiques of cognitive science and RTM

Chapter Three: Method

Chapter Four: Findings

Chapter Five: Discussion

Introduction

In recent years there has been a theoretical shift within the field of cognitive science toward an externalist account of semantics. That is to say, many working within the current cognitive science paradigm have, since the 1980s, been motivated to shift from the syntactical account of semantic content that was prevalent during the field's inception, toward variations of a correspondence theory of meaning. This shift and these accounts of semantic content will be the focus of this research, and this thesis will be scrutinising the current cognitive science paradigm, as well as correspondence theories of meaning, in order to determine the validity of such a referentialist move.

Definition of terms

However, before we proceed, it will first be helpful to define some of the key terms that will be used throughout this investigation, thereby providing some context for the discussion that will follow:

Cognitive Science: Here taken as a multi-approached, or interdisciplinary, investigation into cognition that arose during the latter half of the twentieth century and shares an adherence to a post-positivist paradigm and a representational theory of mind.

Representational theory of mind: A philosophy of mind that posits the manipulation of mental representations as the basis for cognition. This is a view of the mind that is characterised by an adherence to a number of theses:

1) The ontological reality of propositional attitudes and the accuracy of folk psychology.

2) That – following the Church-Turing thesis – computational processes over mental representations can account for these propositional attitudes and explain folk psychology (see Church 1936; Turing 1969).

3) That any physical system capable of instantiating the function of these computational processes will be equally intelligent (i.e., adheres to functionalism).

4) That Marr's tri-level hypothesis (where, following Marr 1982, any system – such as that instantiating cognition – needs to be described at a computational, representational and physical level) is accurate.

Mental representations: These are concrete (that is, ontologically real) information-bearing structures possessed by individuals. These structures are taken to be instantiated by a physical

symbol system and their manipulation, by computational processes, is said to produce cognition. Mental representations may be instantiated in a number of different representational formats, with the most common being propositional and analogical explanations (with the former taking a sentential symbolical format and the latter taking a congruent or simulacral symbolical format).

Correspondence theory of meaning: This is an externalist, or referentialist, account of meaning that explains the semantic content of a particular symbol in terms of the ontologically real state of affairs, that can be found 'out there' in the world, to which that symbol corresponds. As such, correspondence theories of meaning state that if symbol *A* corresponds to entity *B*, *A means B*. This correspondence between sign and signified can take two broad approaches:

1) Correspondence by correlation (where a societal, cultural or biological convention establishes a corresponding relation).
2) Correspondence by congruence (where a similarity of structure, or of logical form, between represented and representation establishes a relation).

Peircean Iconism: A congruent correspondence theory of meaning advanced by Charles Sanders Peirce to explain the semantic content of certain signs. Such signs are said to be meaningful and to represent a specific state of affairs as their own internal structure is analogous to, or mirrors, that of a certain object or a particular state of affairs.

Analysis and synthesis: One of the many methods employed in the pursuit of knowledge by the discipline of philosophy. The form of analysis and synthesis that will concern this research will be of a decompositional format, wherein an argument, hypothesis etc. is broken down into its component parts by analysis and then reconstructed through synthesis. The purpose of this two-part method is to establish the base theories that make up a position and clarify the relations and interactions between them, making the operation and the logical consistency, or inconsistency, of the hypothesis, argument etc. more perspicuous.

Research problem

Now that our terms have been defined and this research has been somewhat contextualised, we can clarify the problem with which we will be concerned. In this regard, it is important to note that within current cognitive science, and certain areas of philosophy of mind, much of the current work produced is entirely reliant upon one central philosophical framework: Computational processes operating within the representational theory of mind. Further, this framework has been employed to build one

of the most promising and potentially productive disciplines advanced in recent years (see Simon and Newell 1976). Nevertheless, although a lot of current research relies upon this paradigm (e.g. Pezzulo and Calvi 2011; Ifenthaler and Seel 2011; Eitel *et al* 2013; Orenes, Beltran and Santamaria 2014), it is not one that can be considered a confirmed, uncontroversial approach, having received many attacks from many angles in the past (see Churchland 1981; Putnam 1981; Dennett 1989; Bennett and Hacker 2003; Papineau 2004; Baz 2012; Hacker 2013). However, by far the most devastating of these attacks are a suite of arguments that originate from accusations of the representational theory of mind's inability to account for meaning. The most well-known of these arguments can be found within the philosopher John Searle's 1980 article "Minds, Brains and Programs", in which – what is commonly referred to as – the "Chinese Room argument" is advanced. This is an argument that purports to demonstrate that were an individual to go through the computations of the mind and mental representations prescribed by cognitive science, then the semantic content of the symbols manipulated would still remain hidden to that individual. Other arguments questioning cognitive science's handling of semantic content can be found in works such as Harnard (1990b), Lewis (1970) and Searle (1980), where respective charges of a transgression of the homunculus and symbolic fallacies are made.[1] Nevertheless, all such arguments reason that all the theory under question is achieving is merely the replacement of one set of symbols with another, and that mental representations can therefore only become meaningful in the presence (and corollary mental computations) of some form of conscious agent, whose own mental representations can only become meaningful in the presence (and corollary mental computations) of some form of conscious agent, whose… and so on, *ad infinitum*.

In response to these criticisms, many cognitive scientists have proposed an update to their hypotheses over the last few decades; an update which is purported to avert these arguments and to once again place cognitive science and its account of meaning back on a secure footing. For example, the psychologist Philip Johnson-Laird forcefully argues that if there exists a relation between language and the world via analogical – or iconic – mental models, then the arguments of Searle and Lewis *et al* are defeated. Johnson-Laird (1988, p115) succinctly describes this hypothesis thus:

> Logicians have only related language to models in various ways; psychologists have related it only to itself.
>
> The real task, however, is to show how language relates to the world through the agency of the mind.

Johnson-Laird further argues that this putative relation between language and the world ensures that what is the anchor for meaning, that what constitute the 'language entry and exit rules' (see Putnam 1981), under his representational theory of mind, is the world and, as such, he avouches that the

---

[1] Briefly, the symbolic fallacy is the explanation of the meaning of a symbol by reference to another, equally content bearing, symbol, while the homunculus fallacy is an infinite regress of explanation. These fallacies will be covered in detail in due course.

avenues Searle and Lewis *et al* have exploited in their criticisms have been eliminated. Furthermore, this is an analysis equivalent to those that have led others operating within the same field to re-evaluate their own semantics and to similarly adopt some form of a correspondence theory as an explanation for semantic content (e.g. Brooks 1990; Harnard 1990b; Waskan 2006; Fodor 2010; Rapaport 2011; Schweizer 2012).

However, there is an issue here: Although Johnson-Laird *et al*'s semantic updates are not insubstantial, there exists limited critical analysis of their updated position in the literature (for example, only Vosgerau 2006 subjects Johnson-Laird's iconic correspondence theory to any form of detailed scrutiny and only Du *et al* 2014, Macbeth *et al* 2014, and Orenes, Beltran and Santamaria 2014 engage in any empirical investigation of central important implications of Johnson-Laird's externalism). As such, what is required before the referentialist representational theory of mind can be confidently adopted is a thorough investigation into its logical validity. Nevertheless, to design and implement a single investigation of this type (i.e., one into the referentialist shift generally), which is capable of making comment on the success of *all* the myriad positions found within the literature that express some form of this move, is far too ambitious: The nuance that exists between all the numerous such positions means that no findings concerning any one body of work will be generalizable to another. Therefore, for this ambitious aim to be met, an individual investigation must to be conducted into each referentialist position. Consequently, it is recommended that a preliminary investigation of this type be conducted to fill one specific gap in the research. Furthermore, this is an investigation this research will task itself with and which is summarised by the following research problem:

> Correspondence theories of meaning, in their own right, are substantial hypotheses that demand many commitments. As such, their inclusion into any standing research program necessitates thorough analysis to ensure no problematic tensions are produced; a level of analysis that is currently absent in much of the cognitive science literature.

Moreover, in this instance, this investigation will focus particularly on the hypotheses of the psychologist Philip Johnson-Laird (who has been chosen as he is both a vocal and a long term advocate of such a move); a focus that produces the following research aim:

> **Aim:** To assess the validity of the hypothesis that the introduction of a Peircean correspondence theory of meaning into Johnson-Laird's mental model theory of mind saves the latter from the critiques of its commentators and to test the success of the inclusion of a Peircean referentialist semantics into such a representational theory of mind.

Research importance

An investigation that addresses this research aim will be of importance and will contribute to the literature in a number of ways:

1) It will fill the identified gap in the literature in clarifying the success or otherwise of Johnson-Laird's referentialist shift.

2) It will continue the nascent discussion concerning the state of Johhnson-Laird's philosophy started by a small number of researchers (see Vosgerau 2006; Du *et al* 2014; Macbeth *et al* 2014; Orenes, Beltran and Santamaria 2014).

3) By undertaking a detailed analysis of this kind, we will inevitably be led to other related topics in the philosophy of mind, the philosophy of language and the field of cognitive science generally. The analysis and critique of such related topics, from within the framework of this research, will lead to novel and important contributions in their own right.

Research approach

To meet the above research aim, this thesis will apply a method that is both appropriate for the stated purpose and which is valid under the post-positivist paradigm to which cognitive science and Johnson-Laird adhere, i.e., a realist ontology and an adherence to a Popperian and/or a Quinean epistemology. Consequently, this investigation will apply critical philosophical analysis to the logic of the hypotheses and arguments of Philip Johnson-Laird, through which we will establish the internal and external logical consistency of the inclusion of a structural correspondence theory of meaning and its corollary assumptions into Johnson-Laird's philosophy. In this instance, and so as to successfully interrogate Johnson-Laird's philosophy in this manner, this means that the following research questions will have to be answered:

1) Does Johnson-Laird's updated account of meaning circumvent the semantic objections of the representational theory of mind's detractors?

2) Is the particular, *nuanced*, correspondence theory of meaning being implemented both internally and externally consistent?

3) Is the introduction of a correspondence theory of meaning, *generally*, into Johnson-Laird's broader philosophy of cognition logically consistent?

It will be stated that if the produced results answer in the affirmative to all three of these questions, the theory will be considered a logically consistent position and will, following Popper (1968, p264), be said to remain as a "provisional conjecture". However, once this thesis completes this processes, it will be found that, as the theory under question produces a negative response to two of these

questions (question 1) and question 2)), the inclusion of an externalist account of meaning into Johnson-Laird's theory of mind will be found to be inconsistent and the theory will be said to have been falsified on both Popperian and Quinean grounds.[2]

Finally, it needs to be noted that the method of philosophical analysis and synthesis will be deemed to be both the appropriate and a valid technique to apply in this instance for a number of reasons, namely:

1) The method of analysis and synthesis is valid under the post-positivist paradigm.
2) That the 'traditional' quantitative and qualitative methods more commonly employed within this paradigm (e.g., experimentation, observation, surveys, ethnography, interviews, case studies etc.) are not suitable given the object of study (i.e., meaning).
3) As the semantic arguments that inspired the referentialist shift that is central to this research derive from the same method, it seems only appropriate to reapply decompositional analysis when scrutinising this shift's validity.

Additionally, it is also worth clarifying that although adherence to cognitive science's post-positivist paradigm may itself be problematic, this research will not engage in any form of scrutiny of the theory at this level. Instead, this investigation will unquestioningly operate within this framework; this will serve to produce findings which are more likely to be acceptable to cognitive science and Philip Johnson-Laird.

Thesis structure

So as to present this research problem, thesis aim, the investigation itself and its findings in a detailed and cohesive manner, this thesis employs the following structure:

Chapter one: Cognitive science

This chapter will begin this research's literature review, giving an analysis and an exposition of the general theories of concern and thereby contextualising the investigation. This covers: (i) Cognitive science as a field; (ii) the representational theory of mind; (iii) mental representation; and (iv) an account of semantics under the cognitive science paradigm.

Chapter two: Critiques of cognitive science and the representational theory of mind

A continuation of the literature review, that: (i) Details the current state of affairs concerning the arguments against the paradigm detailed in Chapter one; (ii) analyses how theorists have responded

---

[2] The detail behind this investigation's conclusions will be handled in much detail later.

to these arguments; and (iii) provides a detailed scrutiny of the correspondence theories that constitute the heart of this response. From this, it will be identified that the introduction of a correspondence theory is a substantial introduction to any philosophy of mind and that a limited amount of detailed research into such an introduction into the cognitive science paradigm has been conducted. Following from this observation, we will undertake a review of the literature and determine that this observation is accurate. As such, it will be claimed that a gap in the literature has been found and that research conducted in this area would constitute a contribution to knowledge.

Chapter three: Methodology

An argument for the appropriateness and validity of the method of critical philosophical (i.e., decompositional) analysis and synthesis, in this instance. During this discussion we will be answering a number of questions, including: (i) What will be done to answer our research aim? (ii) How will this method be implemented in this instance? And (iii) why is this approach justified here?

Chapter four: Findings

A presentation of the findings that are generated from the application of philosophic analysis to the arguments of Philip Johnson-Laird. These findings will concern the three central research questions identified earlier: Do Johnson-Laird's updated semantics circumvent the semantic objections of the representational theory of mind's commentators? Is the particular correspondence theory of meaning being implemented both internally and externally consistent? And, is the introduction of a correspondence theory of meaning, generally, into Johnson-Laird's broader philosophy of cognition logically consistent? After finding and explaining the conclusion that Johnson-Laird's philosophy is currently logically inconsistent, this chapter will then evaluate the validity and contribution of this conclusion.

Chapter six: Discussion

This chapter will draw out the inferences that can be made from the Findings chapter and offer some proposals for future research based on these findings and inferences. This chapter will again be organised into three sections, each covering the research questions that similarly structured our Findings chapter (i.e., Chapter four)**.**

Summary

To clarify the intentions of this research, this Introduction has answered the following questions:

Q1) Within which context will this research fit?

A1) This research will concern itself with the field of cognitive science. More precisely, this investigation will focus on the research of referentialist explanations of meaning currently found within this field and will be particularly interested in the arguments and hypotheses of the psychologist Philip Johnson-Laird concerning his externalist representational theory of mind.

Q2) What gap in the literature will this research try to address?

A2) This investigation will be attempting to establish the success of the operation and introduction of correspondence theories of meaning within cognitive science's philosophy of mind. As many nuanced positions can be found within the literature, this research will limit its intentions to an investigation of the accuracy of only one such account: That of the psychologist Philip Johnson-Laird. Consequently, this position generates the following research aim:

> **Aim:** To assess the validity of the hypothesis that the introduction of a Peircean correspondence theory of meaning into Johnson-Laird's mental model theory of mind saves the latter from the critiques of its commentators and to test the success of the inclusion of a Peircean referentialist semantics into such a representational theory of mind.

Q3) What method will be applied when investigating the issue at hand?

A3) The method this research will employ to discover the success of Philip Johnson-Laird's externalist representational theory of mind, will be to apply philosophic analysis and synthesis to his arguments and theories. This method will be employed so as to establish the logical consistency of Johnson-Laird's philosophy of mind and to identify any areas of tension (if any such areas exist). This use of the method of philosophy in this instance is deemed to be an appropriate and valid approach as it adheres to the post-positivist paradigm to which Johnson-Laird adheres and is appropriate under both Quinean (see Quine 1960, p3; Colyvan 2001) and Popperian (see Popper 1968, p264; Popper and Miller 1983, p155; Finocchiaro 2005, pp255-256) epistemologies. From this application, if we find Johnson-Laird's theory of mind to be inconsistent then this investigation will state that there exists a problem with his externalist semantics and that the validity of his referentialist move is in doubt. However, if this investigation finds Johnson-Laird's account to be consistent, this research will be able to assert that his referentialist representational theory of mind, following Popper (1968, p264), still remains as a "provisional conjecture".

Q4) How will this research contribute to the literature?

A4) Any findings this research produces will constitute an important contribution to the literature, as:

1) There currently exists a dearth of critical analysis of the success, or otherwise, of the referentialist move in cognitive science generally and Johnson-Laird's thought particularly. That is, any findings will fill the gap in the literature identified by this thesis.

2) These findings will inform and continue the discussion begun by a limited number of writers (i.e., Vosgerau 2006; Du *et al* 2014; Macbeth *et al* 2014; Orenes, Beltran and Santamaria 2014).

3) A thorough analysis of the topic at hand will also contribute to addressing broader questions found within the literature of philosophy of mind, philosophy of language, cognitive science and consciousness studies; questions concerning the relationships between the world and language, language and mind, mind and the world and, ultimately, subject and object.

Literature Review

Chapter One:
Cognitive science

Introductory remarks

For several decades, cognitive science has been claimed to be one of the most promising and potentially productive fields of study. Indeed, it is considered so vital in some circles that its findings have been labelled a "Law of Qualitative Structure" by Simon and Newell (1976, pp115-117), who liken the position of the field's discoveries within consciousness studies to that of the cell doctrine within biology, plate tectonics within geology and the germ theory within immunology. Given this broad potential, what is of particular concern for the purposes of this thesis are the many things cognitive science has had to say regarding the foundation of meaning and semantics. However, before we are able to fully engage with the relevant literature concerning this account, it is first necessary to orientate any discussion by fully explaining the theory and field in which it will be couched. As such, this chapter will follow a structure that moves from the more general aspects of cognitive science to the more specific, moving from a broad description of the field itself toward a strict focus upon its account of semantics. With this in mind, this chapter will be structured in the following manner:

    i) An adumbration of the historical and theoretical background of cognitive science.

    ii) An investigation into the representational theory of mind that drives cognitive science.

    iii) An explanation of the role of representation in cognitive science's account of cognition (focusing specifically on propositional representations, as found in the work of the philosopher Jerry Fodor, and their analogical counterparts, as described by the psychologist Philip Johnson-Laird).

§i History and theoretical origins

This section is concerned with positioning the current chapter as a whole by initially clarifying the historical and theoretical background in which it is set. To achieve this end, this section will:

    1) Give a brief outline of the roots of cognitive science.

    2) Provide an explication of the multidisciplinary focus of the field, as well as the motivations for such a focus.

3) Detail the ontological and epistemological assumptions shared by those working within the field.

Origins

The epithet 'cognitive science' is being used here to designate an interdisciplinary investigation into cognition, perception, reasoning, language, action, emotion and consciousness. As Miller (1979) observes, this mode of investigation is a contemporary approach toward cognition, originating as recently as the 1950s, when a number of theorists from a range of disciplines collaborated on a shared cognitive concern. In fact, Miller (*Ibid.*) uses this collaboration to precisely date the birth of cognitive science to Tuesday 11th September 1956, at the second day of a symposium on information theory held at MIT. This is not a cut-and-dried issue, however, for others have put the origins of cognitive science at various points. For example, Fodor (1975, ppvii-ix) attributes the inception of cognitive science to the work of late nineteenth and early twentieth century speculative psychologists, such as William James and John Dewey. However its rudiment, it is widely acknowledged that cognitive science was truly established as a subject in its own right in the late twentieth century, with the founding of dedicated journals and institutions on the subject in the 1970s. In fact, the first use of the term 'cognitive science' in regard to an individual field is found during this period. The work of Norman and Rumelhart (1975, p409) is a contemporaneous analysis of cognition studies, which concludes that:

> the concerted efforts of a number of people from[...] linguistics, artificial intelligence, and psychology may be creating a new field: *cognitive science.*[1]

Nevertheless, although by 1975 the field of cognitive science has a name, five years later the field has still yet to crystallise into a definitive, identifiable discipline, leading Johnson-Laird (1980, p71) to remark "[c]ognitive science does not quite exist: its precursors do" and that "[*i*]*f cognitive science does not exist then it is necessary to invent it".* Therefore, although the history of the field can be traced back through to a number of points and landmarks, it only becomes a solidified subject at the very end of the twentieth century.

Throughout cognitive science's debatable history, however, the main characteristic used to define it, which has set it apart from 'traditional' disciplines and which has made it such a promising subject, has been its focus on an interdisciplinary approach. As such, this facet of the field is deserving of special attention.

---

[1] Yet, as Bechtel et al. (2001) observe, a second candidate for the first use of the term 'cognitive science' may potentially be found in the 1975 work by Bobrow and Collins, *Representation and understanding: Studies in cognitive science.*

Multidisciplinarity

The multidisciplinarity of cognitive science manifests itself in the inclusion of "philosophy, psychology, linguistics, artificial intelligence, robotics and neuroscience" (Friedenberg and Silverman 2006, p2), as well as neurophysiology, anthropology, mathematics and economics (see Osherson and Lasnik 1990, p11), within the umbrella of the subject. The rationale for employing this multidisciplined mode of investigation comes from a well-worn allegory involving a number of researchers and an elephant in a darkened room: One researcher, focused solely upon the tail of the animal, declares that they are in the room with a snake; another, who has studied only the ear, says that the first researcher is wrong, there is an animal in the room but it is some kind of bird; the final researcher, who has found only a leg, declares that both are wrong and that there is not an animal in room at all, but a tree. The moral of this being: Without communication and cooperation between researchers, who are coming at the same problem from different angles, no true understanding of the issue they are tackling as a whole can be known, be it an elephant or the mind (see *Ibid* pp13-15).[2] Consequently, any work conducted within the field of cognitive science (including this research) must be cognisant of the importance of drawing on the findings and practices of a number of disciplines (such as philosophy, artificial intelligence, psychology and linguistics).

This, though, raises a concern: If all those involved in cognitive science research are doing so from within the practices and methods of their own disciplines, how is there any guarantee that any results can validly inform one another? For this potential difficulty to be avoided, the disparate disciplines engaged in cognitive science must be in agreement on a number of fundamental epistemological and ontological concerns. Consequently, cognitive scientists must all be working within the one research paradigm. If this were not the case, then the findings of each individual discipline could not communicate with, inform nor critique the findings of any other (i.e. the results from each 'traditional' subject would not be "cumulative" with one another (Hughes and Sharrock 1997, p5)).

Ontology and epistemology

As we have seen, in order that their findings may be cumulative (and, therefore, for the field as a whole to be able to advance), practitioners of cognitive science must share a number of fundamental assumptions. From diverse scholars (e.g. Block 1978; Dennett 1991; Fodor 1975; Johnson-Laird 1993; Pylyshyn 1984; Searle 1984; Waskan 2006) it is possible to identify the discipline's theoretical premises:

---

[2] Along with this allegory, many other arguments in support of taking an interdisciplinary approach can be found. For example, see Nissini (1997).

1) **A post-positivist research paradigm:** Cognitive scientists possess a world view and research framework dependent on a synthesis between a realist ontology and a critical epistemology.

   a. **A realist ontology:** Cognitive scientists presuppose that there is a real world 'out there', a world that is independent from any one observer.

   b. **A critical epistemology:** Cognitive scientists reject the classical notion of knowledge as justified true belief (see Gettier 1963; Popper 1974; Quine 1953; 1987).[3] Instead, they view knowledge of the world as being, not only difficult to attain, but necessarily uncertain. As such, Cognitive scientists can take either one of two alternative views. The first of these is a Popperian line, wherein we can only ever be certain that knowledge is wrong, not that it is right (i.e. we can only falsify hypotheses, not unequivocally prove them). The second view is a Quinean one, in which hypotheses are taken as certain if, and only if, it fits into our pre-existing suite of ideas without disturbance (i.e. we can take an holistic approach)**.**

2) **A representational theory of the mind:** Cognitive scientists view the mind as operating on the principles of computation and physical symbol systems (i.e. symbol manipulation).[4] Alternatively, it takes the position that the strong AI hypothesis is accurate (see Searle 1980).

Summary

To orientate the following explication of cognitive science's account of cognition and semantics, it is necessary to first define what interpretation of 'cognitive science' this thesis shall be taking. As such, it is important to make clear that henceforth 'cognitive science' will mean a multi-approached, or interdisciplinary, investigation into cognition, that, as it is generally accepted, arose during the latter

---

[3] Interestingly, in the case of Gettier problems, studies conducted by the experimental philosophers Starmans & Friedman (2012, p272) show that there is a divide between attributions of knowledge to 'Gettiered' individuals by philosophers and the laity, suggesting "the lay concept of knowledge is roughly consistent with the traditional account of knowledge as justified true belief, and also point to a major difference between the epistemic intuitions of laypeople and those of philosophers."

[4] The use of the appellation 'representational theory of mind' here is, following Fodor (1997), intentionally being used to designate a philosophy of mind employed by the majority of those working in the field of cognitive science and which contains four key hypotheses:

   1) An assumption of the truth of folk psychology and propositional attitudes.
   2) An explanation of folk psychology using mental representations.
   3) Is quiet about the nature of these mental representations, only stressing that they are not mental images.
   4) Uses the computation (as specified by Turing) of mental representations to explain cognition.

As such, the term 'representational theory of mind' is intentionally being used to distance itself from, yet include, specific forms of the philosophies employed by cognitive scientists. For example, certain labels, such as the 'computational theory of mind', are used to expresses a similar view but does so with an intimate connection to a particular branch of cognitive science, in this instance the language of thought hypothesis. These specific titles therefore exclude any other, computer based, branch of cognitive science that this research wishes to maintain. An alternative appellation here may well be 'cognitivism' (i.e. "the position that intelligent behaviour can be explained (only) by appeal to internal "cognitive processes" – that is, rational thought in a broad sense" (Haugeland 2000, p9)), and this would be correct: Those involved in the representational theory of mind are engaged in a fundamentally cognitivist program (just as those involved in the computational theory of mind are engaged in a representational theory of mind). However, 'cognitivism' is being defined here as a broad methodological approach and, as such, it has been determined that, where theories of mind are concerned, 'representational theory of mind' is, for the purposes of this research, the more appropriate title. Nevertheless, during this thesis, where it has been deemed appropriate the term 'cognitivism' may still be applied to the overarching project of cognitive scientists.

half of the twentieth century and shares an adherence to a post-positivist paradigm and a representational theory of mind (hereinafter, RTM). Although an understanding of both of the preceding presuppositions (post-positivism and RTM) is vital for anyone involved in cognitive science research – and each will be covered in greater detail – the latter of the two is by far and away the most important.[5] This is because RTM is the motor that has driven much of cognitive science (it is also the theory Simon and Newell (1976, pp116-117) labelled a law of qualitative structure).[6] As such, before engaging with the cognitive science literature itself, it is first necessary to dedicate our subsequent section to the exposition of this particular philosophy of mind.

§ii The representational theory of mind

When first introduced to the RTM, it is important to realise that it is an hypothesis that, generally, is composed from, or entails, a number of clearly defined, independent theories. These theories are dependent on each other in a structure that will inform that of this section and are as follows:

1) The concept of propositional attitudes and folk psychology.
2) The theory of computation and the operation of physical symbol systems.
3) Functionalism.
4) Marr's tri-level hypothesis.

Folk psychology

When beginning to consider the RTM, it should be noted that it is a view of the mind that is a variation of, and a development from, an outlook that is commonly accepted by most non-specialists: That humans formulate and execute plans of action on current knowledge of the world and by processes hidden from one another (see Waskan 2006, p43). This concept of Folk psychology is defined as a body of everyday definitions, or generalisations, concerning behaviour, that are commonly posited in individuals by others to explain the causes of their behaviour. Technically, folk psychological explanations are distinct in their taking of causal interactions between mental states (i.e. propositional – belief/desire type – attitudes) and behaviour for granted (see Garnham 1991; Swiatek 2011; Strijbos and de Bruin 2012), such that the following principle (taken from Smith 1987, p36) can be established:

> Agent A at t has a motivating reason to Φ only if there is some Ψ such that, at t, A desires to Ψ and believes that were he [sic] to Φ he would Ψ.[7]

---

[5] The details of the field's adherence to a post-positivist paradigm will be covered in the Methodology chapter.
[6] It should be noted, however, that the representational theory of mind is not unilaterally held throughout cognitivist approaches and other related fields. For example, embodied and extended approaches to the mind can be found in the literature (see Clark 2003; Gallagher 2005; Wheeler 2005), while, similarly, technical and non-representational theories can be found within AI's (see Brooks 1990; Müller 2007; White 2011).
[7] This focus on propositional attitudes has led folk psychology to be referred to as the Belief-Desire Model (or, more simply, the BD-Model) of action theory (see Strijbos and de Bruin 2012).

Where *Φ* is a means-end belief and *Ψ* is a desire, such that the following explanation is produced:

Kate (*A*) at noon (*t*) has a motivating reason to get a drink (*Φ*) only if there is some thirst (*Ψ*) such that, at noon (*t*), Kate (A) desires to quench that thirst (*Ψ*) and that Kate (*A*) believes that were she to get a drink (*Φ*) she would quench her thirst (*Ψ*).

Or, to state it more succinctly:

Kate drank some water because she was thirsty.[8]

This, then, is ultimately a commonplace view of the causes of action, and it is one that has been held by a considerable number of people for a considerable length of time, but it has only held popular currency in the field of philosophy since Davidson (1963, pp686-686) forcefully argued that:

for us to understand how a reason of any kind rationalizes an action it is necessary and sufficient that we see, at least in essential outline, how to construct a primary reason[…] The primary reason for an action is its cause[…] A primary reason consists of a belief and an attitude.

Since these belief/desire type of explanations have become respectable, cognitive scientists, following Davidson, have taken them for granted, and, as such, are committed to the ontological reality of the propositional attitudes that drive them (see Waskan 2006, pp36-76; McGinn 1989, p120; Fodor 1990, p32). Cognitive scientists, then, are convinced of the ontological and causal reality, or truth, of propositional attitudes, but it is vital to note that they are not convinced by the positioning of these beliefs and desires as a final explanation of action, as final *explanantia*. Instead, cognitive science takes these attitudes to be *explananda* in their own right (see Fodor 1990, pp4-5). As such, cognitive scientists are not interested in these general folk explanations themselves, but in explaining the mechanisms operating behind these generalisations: They are interested in detailing what brings rise to propositional attitudes (Waskan 2006, pp50-51). To achieve this aim, those working in the field have drawn on the idea of physical symbol systems from the discipline of computation.

Computation and physical symbol systems

The central hypothesis that underlies RTM states that the mind, like a computer, is a physical symbol system. That is to say, both the mind and a computer function and produce intelligence through the operation of a physical system and that system's manipulation of symbols. Moreover, it is claimed that these manipulations originate from a finite number of pre-determined operations present in the physical system itself. For example, in the case of the most basic Turing machine, all that is required is a tape divided into frames, a read/write head and a store of operations.[9] Each frame of the tape is

---

[8] This concise stating of the folk psychological explanation implicitly assumes that Kate believes that what she is drinking is water, that drinking water quenches thirst and that she desires to quench her thirst.

[9] The Turing machine is named after its inventor and the man who, it can be considered, laid the foundations for the RTM and cognitivism: Alan Turing (1950). However, although Turing was instrumental in the development of the RTM, further important individuals and works

marked with the symbol '1', or the symbol '0', or is left blank. When the read/write head then reads a frame it performs a certain pre-defined action determinate on the symbol (if any) it has found and the commands present in its store. The commands present in the machine defined by Turing are:

1) Delete the current symbol (if any) and write down a '1'

2) Delete the current symbol (if any) and write down a '0'

3) Move the tape one step to the right

4) Move the tape one step to the left

Or, as Turing (1936, p231) explains the operation of his machine:

> We may compare a man in the process of computing a real number to a machine which is only capable of a finite number of conditions q_1; q_2; ... q_R which will be called "*m*-configurations", The machine is supplied with a "tape" (the analogue of paper) running through it, and divided into sections (called "squares") each capable of bearing a "symbol", At any moment there is just one square, say the *r*-th, bearing the symbol $\exists(r)$ which is "in the machine", We may call this square the "scanned square", The symbol on the scanned square may be called the "scanned symbol", The "scanned symbol" is the only one of which the machine is, so to speak, "directly aware", However, by altering its *m*-configuration the machine can effectively remember some of the symbols which it has "seen" (scanned) previously. The possible behaviour of the machine at any moment is determined by the *m*-configuration q_n and the scanned symbol $\exists(r)$. This pair q_n, $\exists(r)$ will be called the "configuration": thus the configuration determines the possible behaviour of the machine. In some of the configurations in which the scanned square is blank (i.e. bears no symbol) the machine writes down a new symbol on the scanned square: in other configurations it erases the scanned symbol. The machine may also change the square which is being scanned, but only by shifting it one place to right or left.

With this finite number of symbols and operations, basic Turing machines are, as Turing (1936) demonstrates, capable of performing a vast array of variegated tasks. For example, if furnished with the following tape:

0	1	1	1	0			0	1	1	0

And the operation table:

		Blank	0	1
	1	16	EL	1
	2	L2	E3	?

---

that contributed to its foundation were Noam Chomsky's 1955 review of Skinner's *Verbal behaviour*, and the 1960 article "Minds and machines" by Hilary Putnam.
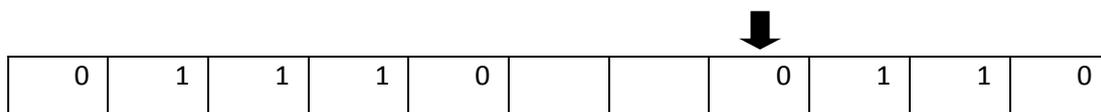
17

marked with the symbol '1', or the symbol '0', or is left blank. When the read/write head then reads a frame it performs a certain pre-defined action determinate on the symbol (if any) it has found and the commands present in its store. The commands present in the machine defined by Turing are:

1) Delete the current symbol (if any) and write down a '1'

2) Delete the current symbol (if any) and write down a '0'

3) Move the tape one step to the right

4) Move the tape one step to the left

Or, as Turing (1936, p231) explains the operation of his machine:

> We may compare a man in the process of computing a real number to a machine which is only capable of a finite number of conditions q_1; q_2; ... q_R which will be called "*m*-configurations", The machine is supplied with a "tape" (the analogue of paper) running through it, and divided into sections (called "squares") each capable of bearing a "symbol", At any moment there is just one square, say the *r*-th, bearing the symbol $\exists(r)$ which is "in the machine", We may call this square the "scanned square", The symbol on the scanned square may be called the "scanned symbol", The "scanned symbol" is the only one of which the machine is, so to speak, "directly aware", However, by altering its *m*-configuration the machine can effectively remember some of the symbols which it has "seen" (scanned) previously. The possible behaviour of the machine at any moment is determined by the *m*-configuration q_n and the scanned symbol $\exists(r)$. This pair q_n, $\exists(r)$ will be called the "configuration": thus the configuration determines the possible behaviour of the machine. In some of the configurations in which the scanned square is blank (i.e. bears no symbol) the machine writes down a new symbol on the scanned square: in other configurations it erases the scanned symbol. The machine may also change the square which is being scanned, but only by shifting it one place to right or left.

With this finite number of symbols and operations, basic Turing machines are, as Turing (1936) demonstrates, capable of performing a vast array of variegated tasks. For example, if furnished with the following tape:

0	1	1	1	0			0	1	1	0

And the operation table:

		Blank	0	1
	1	16	EL	1
	2	L2	E3	?

---

that contributed to its foundation were Noam Chomsky's 1955 review of Skinner's *Verbal behaviour*, and the 1960 article "Minds and machines" by Hilary Putnam.

3	L3	E4	E5
4	R4	?	L6
5	R5	?	L1
6	06	!	L3

Wherein, the tape represents two numbers (a three and a two divided by zeroes), the arrow denotes the position of the read/write head, the top row of the table signifies the content of the current cell being read by the read/write head, the left column of the table represents the possible states to which the machine can be set and the contents of the cells in the table signify the instructions sent to the machine, where:

'1' = draw a one.

'0' = draw a zero.

'R' = move the tape one cell to the right.

'L' = move the tape one cell to the left.

'E' = erase whatever is currently found in the cell.

'?' = state should not be reached.

'!' = end process and output result.

Number: 'Move to state _'

With this tape and store of operations, the machine will start in state **1**, read that the cell contains a zero, erase whatever is in the current cell and move the tape one step to the left and move to state **2**. From here, the tape is again moved one step to the left and reset to **2**. The machine then follows the stored operations to effectively move each '1' from the left number on the strip to the right number, thereby adding the two numbers together. Once this is finished, the machine is commanded to insert a final '0' to the end of the new number to signify completion. In this instance, the Turing machine will produce the following response:

				0	1	1	1	1	1	0
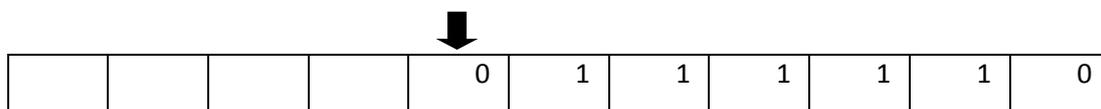
Signifying the number five. This should demonstrate that with nothing but a simple system, an automatic process of addition can be achieved.[10] In fact, with any two sets of numbers placed onto the strip (arranged in accordance with the correct symbolism) and applying the same stored operations, the machine would successfully add them together. But this is not the end of the power of the Turing machine and computation, for the store of operations shown above is only one of the

---

[10] This example is based on one found in Waskin (2006, pp16-17).

many potential sets a machine may be loaded with, and these can be operations designed to perform all manner of calculations (see Turing 1936).

As alluded to, the power of this theory of computation has been taken further than demonstrated above. The likes of Turing (1969) and Church (1936) have stated that *any* process is – potentially – algorithmically computable, or, in the terminology of Church (*Ibid.*, p356), "effectively calculable". Meaning, any process that converts an input into an output can be instantiated through the process of computation if the correct symbols and operations are provided, *including the operations of any other Turing machine*. This ability to store the program of special purpose machines in the memory of another, thereby emulating it, creates the concept of a 'universal' Turing machine that is capable of reproducing the behaviour of any, and all, possible special purpose Turing machines, such as the addition machine detailed above (see Turing 1969; Preston 2002, p5).[11] If one accepts this Church-Turing thesis, as it has become known, one will therefore be led to the conclusion that it is theoretically possible that the processes of folk psychology's beliefs and desires (which, effectively, convert inputs (sense data etc.) into outputs (actions, propositional attitudes etc.)) can be computationally represented, just like the process of addition shown above. Although not formally provable, the Church-Turing thesis has so far not been refuted and still stands as motivation for the conviction that the beliefs and desires of folk psychology are effectively calculable by a computer (see Johnson-Laird 2006, p2). The RTM, then, uses this Church-Turing thesis as an argument in favour of the potential explanatory power of computation and physical symbol systems in regard to cognition, arguing that an increase in the complexity of a computational system would come with an exponential growth in the power of the computer and the types of executable operations it is capable of performing: From simple calculations, to induction, deduction, linguistic capabilities and further actions of cognition.[12] This then is how cognitive science views the brain: As a processor that performs operations on hidden symbols, which derive from sensory input, to produce – or output – beliefs, desires and behaviour. That is to say, cognitive scientists, and any other RTM theorist, take cognition to operate on the same principles as those of a computer. Consequently, all that is required is to provide a suitable level of description, like Turing did with his machine, and we will have a fully detailed account of the mental entities and propositional attitudes that are claimed to drive mentality.

With this reliance on the principles of computation made explicit, it becomes apparent that cognitive scientists are claiming that cognition can arise through the operation of *any* physical symbol system,

---

[11] The digital computers and handheld devices with which we are all familiar are examples of such 'universal' Turing machines.
[12] However, some argue that this position is based on a misreading of the Church-Turing thesis (see Goldin and Wegner 2008; Abramson 2011).

including computers themselves (this is a position known as the strong AI hypothesis, a position that states that "the computer is not merely a tool in the study of the mind; rather, the appropriately programmed computer really is a mind, in the sense that computers given the right programs can be literally said to understand and have other cognitive states" (Searle 1980, p417)).[13] What follows from this is the belief that many physical systems, as long as they are capable of instantiating the correct operations and symbols, are capable of cognition (see Polger 2012, pp337-338). Or, as this interdependence of system and intelligence is articulated by Simon and Newell (1976, p120):

> physical symbol systems are capable of intelligent action, and[…] general intelligent action calls for a physical symbol system.

Consequently, cognitive science and the RTM cannot be accused of what Cuda (1985) terms "neural chauvinism", in that they conclude that any system, not just the neurons of brain matter, can instantiate intelligence.[14] Instead, the RTM states that mentality is multiply realisable. As such, RTM is a branch of functionalism (see Johnson-Laird 1983, pp8-10; Fodor 1990, pp9-10), an explanation of cognition that merits its own exposition.

Functionalism

Although many different breeds of functionalism can currently be found within the literature, the type of functionalism we are concerned with here is a metaphysical position that provides a theory on the nature of the operation and foundation of mind.[15] This theory states that what constitutes a mental state (e.g. a desire to quench a thirst, the belief that drinking water will quench a thirst) is a functionalist state. In turn, functionalist states are defined by their processing – or handling – of the interactions between inputs (perceptions), other mental states (beliefs/desires) and outputs (behaviours and other propositional attitudes) (see Putnam 1975, pp362-440). Alternatively, as Heil (2004, p99) has it, "a state is a functional state of a particular sort in the event that it answers to a particular job description". That is to say, functionalists assert that mental states are not instantiated by a specific collection of neurons within a brain, but rather by the way a physical system operates or functions, the role it plays, and how it responds to stimuli. For example, in regard to our earlier example, Kate's desire to drink water is only constituted by an appropriate input stimulus (e.g. a thirst), the production of appropriate beliefs in the stimulus (e.g. how to quench a thirst and how to behave if you are thirsty) and the activation of an appropriate action (e.g. to go pour a drink), and *only* because of them. Under a functionalist view, therefore, only systems with internal states that can

---

[13] The strong AI hypothesis is opposed to the weak, or cautious, AI hypothesis: "According to weak AI, the principal value of the computer in the study of the mind is that it gives us a very powerful tool. For example, it enables us to formulate and test hypotheses in a more rigorous and precise fashion" (Searle 1980, p417).

[14] We will be ignoring the mereological fallacy during the following discussion and will be equating the mind with the brain throughout. However, this position is being adopted only for ease of exposition and will be reconsidered later.

[15] Examples of these alternate forms of functionalism include "machine functionalism, semantic functionalism, etiological functionalism, methodological functionalism and so on" (Shapiro 2008, p7).

adequately meet these conditions, or that can play these roles, are capable of quenching a thirst (or holding any other functionalist state). Consequently, functionalist states are said to be multiply realisable, i.e. can be instantiated by any adequate physical realiser. These physical realisers can be produced from any physical system as long as it is adequate for the operation of the functionalist state. Consequently, functionalist states and, *eo ipso*, functionalism is said to be "hardware neutral" (Shapiro 2010, p8).

For the purposes of clarification, the technical term 'realiser' deserves extra attention. For it should be made clear that a physical realiser is not the same thing as the functional state that the physical system realises: A diamond is not the same as an element of carbon, and your computer software is not the same as your computer hardware (even though in both instances the latter is a necessity for the presence of the former). Consequently, no realiser should be mistaken for the functional state it realises (see Putnam 1975, pp362ff). That is to say, functional states supervene upon (emerge from, are enabled by, or are "correlated with" (*Ibid.*, p337)) physical realisers. And, as Polger (2007) observes, this non-identical/supervenience relation follows through to minds and brains (or AI and computer hardware): Brains (or computer hardware) are what realise a functionalist state, but they, in themselves, are not minds (or AI). The power of this facet of functionalism has been observed by Polger (2012, p341):

> The main traditional argument for functionalism is precisely that brain states appear to be in an any-to-one relation to psychological states—many different kinds of brains (or nonbrains, potentially) seem to be plausible candidates for having psychological states. According to this line of reasoning, we have empirical reason to think that psychological states are not uniquely made up (one-to-one) but are rather 'multiply realized' by different brain states in different creatures. This was Putnam's original argument for his functionalist hypothesis[.]

But this now seems to raise an issue: Have cognitive scientists been led into some form of dualism by simultaneously positing both a physical realiser and a functional state as an explanation of the mind? As we shall see, this is not the case, cognitive scientists are not expanding their ontology with such a distinction, but are detailing the one system from divergent contexts. This is an aspect of cognitive science's view of cognition that is deserving of special attention.

Levels of analysis

When interpreting the answers that cognitive scientists have posited over the years as an explanation to the questions of cognition, it must be realised that they are doing so at three separate levels of abstraction (see Pylyshyn 1984). This is not to say that cognitive scientists have posited three

independent varieties of mental operation, just that they are explaining the same thing from three different perspectives. From this fact it can be recognised that cognitive science's view of the mind rests upon an ontological monism but a conceptual triad or ternary (i.e. it posits one substance – matter – but explains its behaviour through three different lenses). Consequently, these three separate levels are not merely taken to be synonymous, they are said to be *identical*. This conceptual ternary is a view that has become known (after its originator David Marr 1982) as Marr's tri-level hypothesis (see Beebe 2004). This division into three levels of abstraction is done to manage the complexity of the mind, for, as Marr (1982, pp19-20) asserts:

> Almost never can a complex system of any kind be understood as a simple extrapolation from the properties of its elementary components. Consider, for example, some gas in a bottle. A description of thermodynamic effects – temperature, pressure, density, and the relationships among these factors – is not formulated by using a large set of equations, one for each of the particles involved. Such effects are described at their own level, that of an enormous collection of particles; the effort is to show that in principle the microscopic and macroscopic descriptions are consistent with one another. If one hopes to achieve a full understanding of a system as complicated as a nervous system[…] or even a large computer program, then one must be prepared to contemplate different kinds of explanation at different levels of description that are linked, at least in principle, into a cohesive whole[.][16]

Therefore, cognitive scientists must describe the mind at all three levels of abstraction before they can claim its complete understanding. Consequently, the levels of abstraction in computation are akin to those concerning the organisation of economic laws, which must be separately described and studied at both the macro and micro levels before a system can be said to be fully detailed (see Pylyshyn 1999, p5). Prior to explaining why this is the case, it is first necessary to enumerate the levels of abstraction that Marr and his peers have put forward:

1) **Computational abstraction:** This is the most abstract level of description at which cognitive scientists describe the mind. This level is concerned with the broad intentions and operations of the system under description. As such, the computational level is described as being concerned with the *what* and the *why* of systems and their operation (*Ibid*., p22).

2) **Representational abstraction:** This level attempts to explain the actions posited at the computational level: It answers the *how* of the computational level's *what* and *why*. This is done by answering the question of how inputs are processed and converted into outputs (see Marr 1982, pp19-27).

---

[16] This division into layers of abstraction can be thought of as an approach akin to that employed by the field of electrical engineering, wherein electrical phenomena are treated and described at a number of levels, from Maxwell's equations, through the logic gate abstraction to a physical circuit system, because, as Agarwal and Lang (2005, p4) have it, "electrical engineering creates further abstractions to manage the complexity of building large systems".

3) Physical abstraction: This is the least abstract level at which the mind is described. This level of description concerns the particular physical realiser that is being used to produce mentality (i.e. the neurons and tissue that make up the brain, or the specific hardware used to run AI) and details "[t]he structure and the principles by which the physical object functions" (Pylyshyn 1989, p57).[17]

As an example for why these discrete levels of abstraction are needed, it will be enlightening to return to our previous example of the Turing machine, which was only fully described once we knew what it was doing, how it was doing it and what it was doing it on. In that instance, the computational description was provided by the mathematical concept of addition; the representational level was populated by the binary symbol system of 0s and 1s as well as the stored operations that determined the actions of the machine; and the physical level was completed by the strip of paper, the read/write head, the hardware that stored the operations and the entire cabling etc. that connected the equipment and enabled it to run. From this, the power and importance of David Marr's analysis should become apparent: Through a description from all three level of Marr's tri-level hypothesis, Alan Turing has provided a full explanation of a computing machine for addition. This in turn should then explain our earlier observation that before we can truly understand certain phenomena, such as cognition, we must first provide a description of that phenomena from the perspective of all three levels of abstraction. This is the case because we surely cannot be said to fully comprehend a system until we can detail what that system is doing, state why it is doing what it is doing, specify how it is doing it and detail the physical realiser that is instantiating it.

Finally, from this analysis, it should become apparent that the RTM is, fundamentally, a physicalist account of the mind, i.e. it holds that all that exists is physical even those things that do not appear as such (e.g. cognition). This can be seen in its adherence to the monist ontology that constitutes the base abstraction of Marr's tri-level hypothesis, i.e. the physical. As such, the position is committed to the view that phenomena that do not appear to be physical, such as the mind and mentality, are generated through a supervenience relation with concrete, physical matter.

Summary

To summarise, then, the RTM is an explanation of mentality that is founded on the following principles:

---

[17] These levels of analysis are also referred to, respectively, as 1) the semantic or knowledge level; 2) the symbol level and 3) the physical or biological level (see Pylyshyn 1989, p57).

Folk psychology: It takes the common view of action causation in stating that the propositional attitudes of beliefs and desires are ontologically real and are both necessary and sufficient for instigating an individual's behaviour.

Computation and physical symbol systems: It follows the Church-Turing thesis, from the field of computation in stating that the operation of a physical system upon a finite number of symbols is capable of processing data and producing an infinite variety of complex behaviours. Taking this thesis, cognitive science uses it to explain the operation of the propositional attitudes central to folk psychology.

Functionalism: It is led to the view that if *any* physical system, not just brain tissue, can play the role of the processes of cognition, i.e. that can *function* in the manner defined at the three levels of abstraction, then it is said to be engaged in cognition as equally as brains are claimed to be.

The tri-level Hypothesis: It states that for any complex system, such as cognition, to be fully explained, it has to have been done so at three separate levels of abstraction:
1) *The computational level*, which fully states what a system is doing.
2) *The representational level*, which involves the symbols and operations that explain how a system is doing what it is doing.
3) *The physical level*, which details the physical makeup of the system.

That is to say, the RTM follows the folk conception of psychology in assuming the ontological reality of propositional attitudes such as beliefs and desires, as well as the causal role in action of these propositional attitudes, and takes these beliefs and desires as requiring explanation in their own right. As such, following the concept of a physical symbol system from the field of computing, cognitive science has provided a deeper explanation of the causal power of mental elements with the use of a tripartite abstraction that explains the behaviour of these beliefs and desires at three distinct levels: Computational, representational and physical. This is a view that results in a functionalist position, wherein any physical system capable of playing the roles adumbrated in these three levels (be it composed from neurons or microchips) is capable of cognition (see Fodor 1990, pp1-30).

Now that we have clarified the theoretical positions that underlie the RTM, we are in a position to take a closer look at the level of abstraction that is most in need of an explanation. For, where cognition and the RTM are concerned, the physical and computational levels should already be

apparent: The physical realiser is our brain tissue (or, if we were to avoid the mereological fallacy for a second, the biology and electro-chemical processes that makes up a person) and the questions of the what and the why of mentality are answered by folk psychology. However, questions arise once we consider the middle, representational, level. For, as Marr (1981, p23) observes, when considering the representational level it is important to keep in mind that "there is usually a wide choice of representation". In other words, for any computational system there is any number of representational descriptions (i.e. symbols and operations) that could successfully account for its behaviour. Accordingly, cognitive scientists have posited a number of competing, and antithetical, symbol systems to account for mentality over the years. Consequently, we need to continue by providing a brief account of these systems.

§iii Forms of representation

Now that the assumptions operating behind cognitive science research have been clarified, it would be apposite to focus on those structures theorists have posited at the representational level of cognition, and explain how those structures are able to produce phenomena such as reasoning and semantics. To achieve this end, this section will: 1) Explain the concept behind the symbols, or mental representations, cognitive scientists have hypothesised at the representational level of cognition; 2) elucidate the propositional account of representations, covering that account's explanation of meaning; before 3) giving a detailed description of analogical representations, including a look at its historical and theoretical influences and explaining by what means these models are able to account for reasoning.[18]

Mental Representation

Where cognitive science and cognition is concerned, theorists have conceived of a number of ontologically real formal information-bearing structures, of one kind or another, at the representational abstraction, all of which are combined and manipulated by pre-determined specified rules, computations or operations to elucidate folk psychological explanations of behaviour. An important note here, however, is that over the years, psychologists, philosophers and cognitive scientists have all posited several possible, and antithetical, varieties of structural form for these mental representations. The most prevalent conceptions concerning these mental representations can be divided into two broad trends: The first toward approaches based on a mental logic that use formal rules of inference between representations to produce reasoning and that are based on a

---

[18] It shall become clear that, in the context of this thesis, how these analogical models account for the foundations of meaning is deserving of special attention. Consequently, this facet of mental representation will be handled in its own, succeeding, section (§iv Procedural semantics).

sentential symbolical format (i.e. propositional representations); and the other toward an isomorphic or analogical approach, which explain reasoning through subjecting representations to continued testing and use a model or simulacral symbolical structure (i.e. analogical representations or mental models). As such, it is apposite of us to shed some light on both of these conceptions.

Propositional representations

The syntactic approach to the structure of mental representation is arguably the traditional and most prevalent view of the operation of cognition found within the cognitive science literature (see McGinn 1989, p172; Fodor 2001, p38).[19] It is a theory of mind that is most intimately connected with the philosopher Jerry Fodor, who advances the view that:

> Representation presupposes a medium of representation, and there is no symbolization without symbols. In particular, there is no internal representation without an internal language. (see Fodor 1975, p55)

That is to say, adherents of propositional representations, such as Fodor and others, explain cognition by postulating an internal Language of Thought (hereinafter, LoT) – a language that has subsequently become known as 'Mentalese' (Fodor 2000, pp63-74) – that should be viewed in the strict and literal sense. That is, within the propositional representation paradigm, there exist a finite number of mental, atomic 'words' that combine, following a strict grammar, to produce an infinite number of meaningful mental sentences and, further, that inferences between these mental sentences account for reasoning (see Fodor 1975, pp79-97).[20, 21] However, before we continue, there is one expositional note to be stressed here: Although there exist numerous, nuanced permutations of the propositional doctrine, espoused by a number of cognitive scientists (e.g. Pinker 1999; Plotkin 2000; Rapaport 2007), for the purposes of exegesis this thesis will focus solely on the work on the subject by Jerry Fodor, who defends his syntactic conception of mental representation via a couple of arguments:

1) That as language is infinitely productive (i.e., is capable of producing an infinity of meaningful propositions), then thought must itself be capable of the same level of production: It must be able of representing an infinity of meaningful thoughts. As such, mental representation must be in the form of a discrete combinatorial system, such as a language (see Fodor 1975, p31; 1990, pp16-19).

2) Fodor points to the field's reliance on Turing machines as a basis for cognition, highlighting the fact that the machines detailed by Turing function on syntactic representations and

---

[19] In fact, concerning the syntactic view, Fodor (2000, p42) effusively states that "[c]ognitive science minus the syntactic theory of computation is Hamlet without the Prince".

[20] Atomic in the sense that the meaning of any one components is not determined by any other. If this were not the case, the theory would be led to semantic holism and issues concerning the inability of individuals to share concepts (see Fodor 2000, p69). See the subsequent chapter for a more detailed discussion of this argument

[21] It is important to bear in mind that this language of thought is not identical to any natural language. This is because, as Fodor (1975, p56) stresses, "[t]he obvious (and, I should have thought, sufficient) refutation of the claim that natural languages are the medium of thought is that there are nonverbal organisms that think."

argues that a theory of mind based on such machines must, *eo ipso*, also rely on the same syntactic, or "discursive", representations of its paradigm case (Fodor 2010, p62).[22]

When considering the philosophy of Jerry Fodor it needs to be recognised that it has undergone some adaptations since its first articulations in the 1970s, and that there are, consequently, essentially two Fodorian LoT hypotheses: The early version of the 1970s and 1980s, as advanced in *The Language of Thought*, and the later, current, incarnation, as seen in *LOT2: The language of thought revisited*.[23] Because of this fact, the following discussion of Fodor's ideas will be organised chronologically: It will move from its first articulations to the theory as we find it today.

LoT

When Fodor first expresses his LoT hypothesis, he maintains the basic facts of any propositional theory of representation: That there exists a mental language that is composed from a finite number of atomic components (see Fodor 1975, pp125-126) and that the processing of these components produces inductive and deductive reasoning (Fodor 2000; 2001; 2007; 2010). Additionally, Fodor (*Ibid.*, p133; pp95-97) states that these components are both native (i.e. unlearned) and as rich, or nearly as rich, as natural languages. As such, he contends that language learning essentially consists in learning (i.e., correlating) the appropriate translations of mental propositions into their natural language counterparts, with a mental "dictionary" mediating between the expressions of the two languages (*Ibid.*, pp78-79; p125). This is a view that entails some rather daunting assumptions, however, especially concerning the "very extreme nativism" it necessitates and presupposes (*Ibid.*, p96). For, Fodor must now accept that *every* term in language, including neologisms, corresponds to a concept that comes preloaded in every individual (i.e. that Plato was born with the concept HELICOPTER). This is an issue Fodor (*Ibid.*) attempts to avoid by arguing that "complex concepts (like, say, 'airplane') decompose into simpler concepts (like 'flying machine')".[24]

Were semantic content is concerned, Fodor (*Ibid.*, pp150-151) uses this atomic nativism as a base to expound the view that meaning is based on a two-part process, affirming:

> A theory of the hearer can reasonably be expected to contain two distinguishable components. The first of these is concerned with explaining sentence comprehension proper; i.e., with characterizing the

---

[22] It should be highlighted that there are issues with both these arguments. For example, although 1) is necessary for representation to be language-like, it is not sufficient for mentality to be in such a format, as opposed to an equally productive form, (e.g., analogical representations). Also, and as we have seen in our prior discussion of Marr's tri-level hypothesis, many representational systems can feasibly occupy the representational level. Therefore, to unquestioningly employ the traditional system used by Turing is fallacious.

[23] Nevertheless, Fodor's work should not be presumed to comprise two discrete, clearly defined stages. Rather, there is a continuum in his work between the publication of *The language of thought* in 1975 and *LOT2* in 2010.

[24] Unfortunately, however, this defence seems to fail in capturing the semantic nuance found between related terms, such as those found within the complex concepts 'helicopter', 'glider', and 'jumbo jet' (cf. Putnam 1988, pp15-18).

computations which effect the correspondence between wave forms and messages; i.e., with specifying those mental operations which eventuate in a display of the information that utterances of sentences convey; i.e., with showing how hearers reconstruct the communicative intentions of speakers. Call this component a 'sentence understander'. The second component is concerned with representing the data processes (including drawing inferences) which are *defined over* the information that utterers of sentences convey; i.e., those data processes which mediate the hearer's use of information he gleans from the utterances he hears. Call this component a *logic*. Then, roughly (abstracting from feedback and the like) the output of the sentence understander is the input to the logic. Equivalently, the (or a) function of the sentence understander is to represent utterances in the normal form for which operations in the logic are defined.

That is to say, for early LoT, there are two stages to sentence comprehension: 1) A conversion of natural language propositions into their Mentalese counterparts (a process accomplished by a 'sentence understander'); and 2) A parsing of these Mentalese sentence, via inference, to recover meaning (completed by a 'logic'). Within the early LoT, the sentence understander is fulfilled by truth rules and the logic by meaning postulates, both concepts that require further investigation.

Truth rules

Truth rules are formulae for converting predicates in a natural language into a LoT and are what establish the 'definitions' in an individual's mental dictionary. They are of the form ⌜*y* is *a philosopher*⌝ *is true iff x is a philosopher*, where the terms on the left of the *modus ponens* are words in a LoT and those on the right are in a natural language (see *Ibid.* pp59-60). Consequently, the example used here would convert the word 'philosopher' into the mental concept PHILOSOPHER. Although these rules may seem trivial or even empty, this is not the case, for, as Fodor (2007, pp1-2) explains:

> "[Truth rules are not trivial or empty…] if you assume that the semantic properties of English sentences are described in some language *other than* English. That 'the cat said 'meow'' is true if and only if 'le chat a dit 'meow' is patently *not* trivial. To the contrary, it's just the sort of fact that a French speaker who is trying to learn English (or an English speaker who is trying to learn French) would need to know".[25]

This is necessary as, for LoT, meaning is only present in mental predicates. Therefore, natural language is devoid of meaning and needs to be translated into a language where it, semantics, is present, i.e. Mentalese (see Fodor 1975, p64). With this in mind, a question seems to be being begged here: If the meaning of a natural language predicate is determined by its translation into its equivalent in the LoT, what determines the meaning of these Mentalese predicates? Is there a further metalanguage of thought that is capable of representing the extensions of predicates (a meta-metalanguage)? In his initial work on the LoT hypothesis, Fodor, working with Jerrold Katz, answers this question with the

---

[25] This is founded on the distinction between mention and use. On the left of the truth rule an input sentence is *named/mentioned*, while the output sense on the right is not named but *used*.

introduction of 'semantic markers and distinguishers' into the mental dictionary, with the meaning of Mentalese terms being accounted for syntactically. As Katz and Fodor (1963, pp185-186) aver:

> The semantic markers and distinguishers are the means by which we can decompose the meaning of one sense of a lexical item into its atomic concepts, and thus exhibit the semantic structure IN a dictionary entry and the semantic relations BETWEEN dictionary entries. That is, the semantic relations among the various senses of a lexical item and among the various senses of different lexical items are represented by formal relations between markers and distinguishers.

Nevertheless, by the time of the publication of *The language of thought,* Fodor has abandoned this syntactic account and instead answers the above question by stating that Mentalese is capable of representing the meanings of predicates as it is capable of representing the physical extensions of those predicates. Further, Fodor argues that it is a brute fact that Mentalese is capable of doing this: That it is an innate ability of individuals that their internal language is structured such that it represents the extension of a predicate. Or, as Fodor (1975, p66), asserts:

> there are two ways in which it can come about that a device (including, presumably, a person) understands a predicate. In one case, the device has and employs a representation of the extension of the predicate, where the representation is itself given in some language that the device understands. In the second case, the device is so constructed that its use of the predicate (e.g., in computations) comport with the conditions that such a representation would specify. I want to say that the first is true of predicates in the natural languages people learn and the second of predicates in the internal language in which you think.

As such, although the early LoT does not have a causal theory of reference, it still, in part, relies on an – innate or non-causal – correspondence between the world and Mentalese for the provision of semantic content.[26] However, this reliance on extensions of mental content is not said to be the whole story of the provision of semantic content, merely a necessary part of it. For Fodor (1987) argues that the mapping of Mentalese tokens onto their appropriate extensions within contexts is also a determiner of meaning, i.e., early LoT relies on both broad and narrow content.[27] Fodor (1987, pp27-54) argues for this reliance on narrow content with the following – here formally stated – argument:

1) **Premise1:** Mental states are individuated by their causal powers. Explicitly, the specific content of a mental state is the specific cause of behaviour. (*Ibid.*, pp33-34; p44).[28]

---

[26] Nevertheless, this position is not uncontroversial. For example, compare Fodor's position with that of Putnam (1975, pp261-262) or Pietroski (2005, p259).

[27] This conception of narrow content is later qualified by Fodor (1991). In "A modal argument for narrow content" he argues for a more nuanced account of narrow content, wherein only entities that share contingent, over conceptual, properties share causal powers. Therefore, only a difference in contingent properties between two entities establishes a difference in their causal powers. As such, any two entities that share contingent properties, even if they differ in their conceptual properties and broad content, are said to have the same narrow content.

[28] However, this does not imply that any two or more mental states that lead to, cause, identical behaviour, will be classified as identical. It is only mental states whose causal powers are identical across counterfactuals and contexts that are said to be the same (Fodor 1987, p35; p48).

2) **Premise2:** An entity's causal powers do not alter with a change in broad content. (*Ibid.*, pp34-35; pp40-41).

∴

3) Causal powers must be an intrinsic property of an entity. That is, context must not affect the causal powers of an entity (*Ibid.*, pp35-36).

∴

4) Mental content must be intrinsic; must be narrow (*Ibid.*, pp44-46).

As such, early Fodor proposes a two-part conception of semantic content, thereby employing a moderate internalism that necessitates both narrow content (i.e., the internal functions performed on mental states that move "from contexts onto truth conditions" (*Ibid.* 1987, p53)) as well as broad content (i.e., the truth conditions and extensions themselves) (see also Fodor 1991, p6; p10).

Meaning postulates

As mentioned earlier, the logic, or that which manipulates the LoT predicates created by truth rules to parse Mentalese sentences, is fulfilled in the early LoT by meaning postulates, which are, as Fodor and Fodor (1980, p761) have it, "rules of logical inference". That is to say, they are the formulae that link terms in the mental language and derive the correct mental content intended by a natural language sentence, or, as Fodor and Fodor (*Ibid.*, p764) have it:

> A meaning postulate[…] is like other rules of inference in that it takes as input a structure representing the *semantic* properties of a construction, and yields another semantic structure as output.

For example the natural language sentence 'John ate' will be translated literally into the Mentalese predicate ⌜*John ate*⌝, which is, strictly, a meaningless, incomplete sentence (as it contains an empty functional argument). With this data as input, the logic, or meaning postulates, will infer the missing functional argument position and deduce, therefore, that the real state of affairs is that ⌜*John ate something*⌝. Meaning postulates – after the abandonment of semantic markers and distinguishers – also handle such semantic functions as synonymy, knowledge by description and generally inform how sentences are to be parsed by defining over the terms, and the associations between those terms, found within an individual's mental dictionary (see Fodor 1975, pp150-152).

LoT2

Over the intervening three decades between the publication of the original *Language of Thought* and *LoT2*, the hypotheses at the heart of Fodor's philosophy has undergone a number of revisions. Although this is the case, the modifications are not so vast as to make the theory of today unrecognisable from its early articulations. For example, the theory remains an explanation of

cognition that is based on finite, atomistic mental representations that are syntactically organised in a manner akin to natural languages and which use inferences between representations to account for reasoning (Fodor 2000; 2001; 2007; 2010). Additionally, LoT2 also adheres to the view of these atomistic representations as being both as profligate, or nearly as profligate, as natural languages and as being innate (2010, pp129-168)).[29] Finally, Fodor (2007, pp1-2) still maintains a 'sentence understander', which translate natural language sentences into their Mentalese equivalents (although the appellation 'truth rules' has subsequently been replaced with the term 'Tarski sentences').

However, where major revisions in the Language of Thought hypothesis has occurred has been within that theory's account of semantics, with a move away from a moderate internalism, which necessitates both a non-causal reference and internal functions (i.e., broad and narrow content), to a purely causal theory of reference, organised within a new structure and abandoning any form of narrow content (see *Ibid.*, 1994, pp1-26; 2010, pp17-18; p199).[30, 31] As such, and to clarify, narrow content, the logic of meaning postulates and the mental dictionary of LoT have been abandoned since the hypothesis' first publication in favour of a causal theory of reference.

The first of these major modifications to Fodor's LoT is a movement away from a two-part moderate internalism involving narrow content (see *Ibid.,* 1994).[32] In its place Fodor (*Ibid.*, pp140-141) has introduced a causal theory of reference, wherein mental concepts derive their meaning through reference with actual and possible properties and individuals in the world alone. That is to say, Fodor (*Ibid.*, p199) has introduced a one-part theory of reference into LoT2 in which the correspondence between representation and states of affairs is established through an individual's experience with those states, rather than having the correspondence as an innate facet of cognition (as was present in the early LoT). This causal basis derives from an individual's experience with the world leading to certain states of affairs "locking on" (*Ibid.*, pp141-144) to nascent Mentalese concepts (nascent in the

---

[29] In fact, LoT2 holds a complex position on the innateness of concepts. While still defining himself as a "raving nativist" (*Ibid.,* p153) and arguing that concepts cannot be learned (*Ibid.*, p139), Fodor (*Ibid.*, p130) also avers that "it doesn't quite follow that any concepts are innate[… because] 'learned' and 'innate' do not exhaust the options". Instead concepts are said to be *acquired* via experience, but what contributes to an individual's ability to acquire a concept or not is determined by their "innate endowment" (*Ibid.*, p145). Even though this is the case, Fodor (*Ibid.*, p146) concludes that:
> Concepts are, as it were, there from the beginning. We have the concepts we do because we have the neurology we do; we have the neurology we do because we have the phenotype we do; and we have the phenotype we do because we have the genotype we do. We have [the concept] CARBURETOR for the same sort of reason that we have ten fingers.

[30] Interestingly, Fodor (2010, p199) was motivated to adopt this causal reference to avert such concerns that include the homunculus fallacy, which, as we shall see, are issues that have motivated adaptations within the semantic theory of others.

[31] As previously remarked, the intervening years between the publication of *The language of thought* and *LOT2* hold a developing continuum of positions within Fodor's LoT hypothesis. For example, a causal theory of reference fixation and, therefore, meaning can be found in Fodor's philosophy as early as 1990, where it is declared that:
> there are the right kinds of causal linkages between the symbols that the device manipulates and things in the world[...] It is entirely reasonable (indeed it must be true) that the right kind of causal relation is the kind that holds between our brains and our transducer mechanisms (on the one hand) and between our brains and distal objects (on the other). (Fodor 1990, p33)

[32] For arguments that motivate Fodor's abandonment of narrow content, see Putnam (1988, pp19-56).

sense that they are always present in an individual but only play any role in cognition once the locking on process has been "triggered" (*Ibid.*, pp141-144)). In other words, under LoT2, concept *acquisition* is *triggered* by experiences of physical instantiations of that concept.[33] Finally, it should be highlighted that the mechanics of this 'locking on' process is not actually detailed by Fodor, it is merely concluded that it is simply the "kind of thing that our brain tissue just does" (*Ibid.*, p152).

The final revision Fodor has made to the semantics of LoT is a move away from a mental 'dictionary' towards that of a mental 'filing cabinet' (*Ibid.*, p94). This move constitutes a restructuring of the mental concepts, or representations, an individual possesses: From a flat system to one of a more relational, dynamic structure. Fodor (*Ibid.*, pp94-95) describes this new structure as follows:

> Think of your head as containing (*inter alia*) an arbitrarily large filing cabinet, which can in turn contain an arbitrarily large set of files, which can in turn contain an arbitrarily large number of memos. We can think of these files, and of the memos that they contain, as quite like real files that contain real memos in the real world (except, of course, that mental memos are written in Mentalese; and, since there has to be room for them in people's heads, mental filing cabinets can't occupy much space). The basic idea is this: When you are introduced to John (or otherwise become apprised of him), you assign him a Mentalese name and you open a mental file, and the same Mentalese expression (M(John)) *serves both as John's Mentalese name and as the name of the file that contains your information about John*; just as, in the ontologically untendentious files that one comes across in ontologically untendentious filing cabinets, the file labelled 'John' is likely to be the one where you'll find stuff pertaining to John.

As such, now, under this file analogy, each mental representation has its own named file, a name that also doubles as that representation's tokening in Mentalese expressions, meaning that "*we think in file names*" (*Ibid.*). These files contain Mentalese 'memos', or pieces of information, pertaining to what is represented and which determine an individual's propositional attitudes towards that which is represented (*Ibid.*, p97; p143). By defining your beliefs etc. concerning entities, these memos also delimit the associations (including synonymy) you hold between representations. For example, your file HOUSE may contain a memo stating that houses tend to be paired with the concept WINDOW, consequently, when you hear "house" you will think "window". Furthermore, as reference exhausts semantic content, the organisation of mental representations explain the relationships between meaning. For example, in the case of Frege (or, substitutivity) cases, where the one referent has two senses (e.g. 'the evening star' and 'the morning star' are filled by the one entity, Venus), is accounted for by a mistake in the filing system: The one entity has, for one reason or another, two different files, containing different memos (i.e. different, and even contradictory, belief/desire states). That is, an individual may possess two concepts: $VENUS_1$ and $VENUS_2$, one that contains a memo concerning a

---

[33] For a full definition of the nature of 'concept acquisition', as defined by LoT2, see footnote [29] and Fodor (2010, pp129-168).

belief in Venus' appearance in the morning, the other a memo concerning Venus' appearance in the evening (see *Ibid.,* p99). Consequently, in Frege cases, although there are confusions over the organisation and propositional attitudes (i.e., beliefs and desires) concerning mental representations, confusions that create linguistic misunderstandings, the meanings of the two terms are identical (*Ibid.,* p82). Although this is a view that necessitates the presence of cognitive dissonance between an individual's belief and desire states, this is not problematic, for, as Fodor (1990, p174) asserts:

> it couldn't conceivably be required that the *propositional objects* of all the attitudes attributed to a guy [sic] at any one time should be to any extent mutually consistent: There's nothing wrong with hoping that *P* while fearing that *not-P*; and believing that *P* while wishing that *not-P* practically defines the human condition.[34]

Analogical representations

In contrast to these propositional representations, stands the mental model theory (or, simply, the model theory). This theory is traditionally viewed as the psychologist's approach to mental representation (see McGinn 1989, p172; Fodor 2008, p197), most likely due to the large amount of explanatory and experimental work produced on the subject within that field (most notably by Philip Johnson-Laird and David Marr). This is a theory based upon the hypothesis that the world is represented in the mind of an individual via internal simulations, isomorphs or models, which are then used as a basis for cognition and action (Johnson-Laird 1977, pp206-207; 1993, p275; 2005, p185). Johnson-Laird (1999, p116) summarises the position thus:

> The theory of mental models[…] postulates that reasoning is based not on syntactic derivations from logical forms but on manipulations of mental models representing situations. Models can represent the world, simulate a process, and yield inductive or deductive inferences.

It is important to note that although it is easy to assume that these mental representations or models are, in form and function, essentially nothing more than a mental image, akin to those with which we are all conscious and familiar, a mental representation should not be confused as being constituted in such a format. This is the case, for, as Johnson-Laird (*Ibid.*) highlights, "models are distinct from images, they can contain abstract elements, such as negation, that cannot be visualized" (see also Johnson-Laird 1980; 1993; 2010). To support this position on mental representation, adherents of the model theory have employed a number of arguments over the years (see Johnson-Laird and Khemlani 2013 for an example of a raft of experiments and empirical data used to support analogical models).

---

[34] This position has also been articulated by McDowell (1977, p173), concerning his own de re account of meaning, as follows:
> [T]he belief which makes the [denial of co-reference] intelligible is a (false) second-order belief to the effect that the subject has, and is expressing, a first-order, belief[…] This second-order belief is manifested by the subjects action, not expressed by his words. No belief is expressed by his words: they purport to express a belief[…] but since no appropriate belief could be thus described, there is no such belief as the belief which they purport to express.

However, one common defence, seen in the likes of Johnson-Laird (1980; 1993; 2006), can be found in a series of famous psychological experiments performed during the 1970s. During these studies, it was found that a subject's response time in a task involving the mental manipulation of a presented figure varied according to said figure's spatial properties (e.g., size, orientation, complexity, etc.), strongly suggesting that participants need to manipulate some form of mental representation which themselves have something *like* or *akin to* spatial properties (see Kosslyn 1980; Shepard 1978).

Now that these preliminaries have been established, we are in a position to explicate the details of the model theory and its operation in more detail. However, one caveat before we begin: Although the analogical mental model theory is a nuanced position and a variety of different permutations can be found within the literature (e.g. McGinn 1989; Waskan 2006), this research will be focusing only upon the particular analogical framework held by the British psychologist Philip Johnson-Laird. Consequently, when first introduced to this particular theory of analogical mental representation, it is important to note that the main influences on the structure and operation of Johnson-Laird's posited mental symbols' come from two sources: The first arising in the first half of the twentieth century by the Scottish psychologist and philosopher Kenneth Craik (see Johnson-Laird 1980; 1983; 1993) and the second from the late nineteenth and early twentieth centuries by the pragmatist philosopher and scientist Charles Sanders Peirce (see Johnson-Laird 2002; 2005; 2006; 2008a; 2010; 2013). As such, to uncover the nature of the models that function at the representational abstraction of Johnson-Laird's approach to cognitive science, it will be valuable for us to examine the work and influence of these thinkers in turn. (N.B. These influences can be seen as defining two clear stages in the development of Johnson-Laird's philosophy of mind, with an early period, covering the 1980s-2000s, defined by a Craikean influence and a later period, from 2002 onwards, characterised by a Peircean aspect. Nevertheless – and although the proceeding section (§iv Procedural semantics) will follow this development – this section will be ignoring this distinction and will only be looking at what the work of both Craik and Peirce tells us about Johnson-Laird's model theory as we find it today.)

Craik and the nature of analogical models

Writing seven years before the publication of Turing's "Computing machines and intelligence" (with *The nature of explanation* being first published in 1943), Kenneth Craik (1914-1945) puts forward a model-based view of the mind founded on the concept of analogue computers, rather than the digital computing of Turing (1950).[35] These analogue computers "are devices which embody functions between their inputs and outputs to produce the requisite solution without going through the steps

---

[35] Interestingly, writing in 1943, Craik (1966, p38) is also concerned with designing machines capable of behaving like the brain, a project akin to that which concerned Turing (1950, p433), i.e. "Can machines think?"

of mathematics or logic necessary for deriving answers by built-in mathematical or logical procedures" (Zangwill 1980, p6). That is to say, analogue computers use physical, mechanical or chemical systems to process inputs into outputs (Craik 1967, p51), whereas, as we have seen in our discussion of Turing (1936), digital computers electronically apply mathematics and logic to do the same (compare analogue and digital watches). As such, when viewing the influence of Craik's work upon modern cognitive science, it must be remembered that it is not a direct influence. Rather, Craik's reasoning is being applied to a digital RTM in place of an analogical one.

Within his own work on modelling, Craik (1967, p54) considers that what is crucial to the constitution of any model is that its structure, though not necessarily its appearance (see *Ibid*., p51), is analogous to the state of affairs it is modelling. In other words, according to Craik and his followers, for a model to model it must share the same 'relation-structure' as the part of the world that is of interest, i.e. "the fact that it is a physical working model which works in the same way as the process it parallels, in the aspects under consideration at any moment" (*Ibid*.). Therefore, for any mechanical model of a situation (e.g. "[a] calculating machine, an anti-aircraft 'predictor', and Kelvin's tidal predictor" (*Ibid*., p51)) to be considered as such, it must be able to be used in place of the original state of affairs for a number of reasons (e.g. expense, availability, efficiency or self-preservation) (see *Ibid*., p52) and it must, like the original, operate by "the 'translation' of the external processes into their representatives (positions of gears, etc.) in the model; the arrival at other positions of gears etc. by mechanical processes in the instrument; and finally, the retranslation of these into physical processes of the original type" (*Ibid*. p51)). Ergo, for Craik, a model *qua* model, must take an input, process it and produce an output, in a natural manner that preserves the structure of the state of affairs to which it refers. This naturalism constraint translates as an equally important factor in the operation of mental models, for it imposes

> an important structural property deriving from a constraint on the set of possible mental models: a *natural* mental model of discourse has a structure that corresponds directly to the structure of the state of affairs that the discourse describe. (Johnson-Laird 1983, p125)

As such, this naturalism, termed "structural preservation and naturalness" by Vosgerau (2006, p255), limits the representational scope of models, ensuring that they only include those things found in a state of affairs and are absent from such things as "sophisticated mathematical notations" (Johnson-Laird 1983, p93).

Following from Craik's reasoning, we are led to the conclusion that cognition itself should be considered as some form of model of the world. For, the two phenomena (cognition and modelling) share a lot of the same attributes. For example, it appears that the major purpose of cognition is to

run a test in place of an actuality. Furthermore, Craik (*Ibid.*, pp50-51) observes that "[d]uring the process of reasoning[…] there are three essential processes:

> (1) 'Translation' of external process into words, numbers or other symbols,
>
> (2) Arrival at other symbols by a process of reasoning, deduction, inference, etc., and
>
> (3) 'Retranslation' of these symbols into external processes (as in building a bridge to a design) or at least recognition of the correspondence between these symbols and external events (as in realising that a prediction is fulfilled)" (*Ibid.*, p50).

This is a line of reasoning that concludes in the insight of the fundamentally isomorphic nature of cognition; an influential conclusion Craik (1967, p61) summarises as follows:

> If the organism carries a 'small-scale model' of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilise the knowledge of past events in dealing with the present and future, and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it. Most of the greatest advances of modem technology have been instruments which extended the scope of our sense-organs, our brains or our limbs. Such are telescopes and microscopes, wireless, calculating machines, typewriters, motor cars, ships and aeroplanes. Is it not possible, therefore, that our brains themselves utilise comparable mechanisms to achieve the same ends and that these mechanisms can parallel phenomena in the external world as a calculating machine can parallel the development of strains in a bridge?

A conclusion that has had a direct influence on the philosophy of Philip Johnson-Laird (1980, p98), leading him to similarly conclude:

> human beings do *not* apprehend the world directly; they possess only internal representations of it[…] A model *represents* a state of affairs and accordingly its structure is not arbitrary like that of a propositional representation, but plays a direct representational or analogical role. Its structure mirrors the relevant aspects of the corresponding state of affairs in the world.

Peirce and the nature of analogical models

However, since the early 2000s, this idea of Craikean modelling and relation-structure finds expression in the work of Johnson-Laird via his employment of the terms 'icon' and 'iconicity' to describe the structure of his representations. Johnson-Laird has borrowed both of these terms from the work of the American psychologist and pragmatist philosopher Charles Sanders Peirce. The use and meaning of these terms derive from Peirce's analysis of, and investigation into, the structure of signs (i.e. into the foundations of a symbol's semantic content). From this investigation, Peirce (1906; 1931-1958) defines a sign as – not necessarily a physical – something that signifies a worldly state of affairs to an

interpreter (e.g. a conscious agent). But he also identifies that this signification occurs in three different ways. As such, Peirce (§§274-284)[36] discerns between three types of sign:

1) Icon
2) Index

and

3) Symbol[37]

The first of these, icons, represent through an analogy between the relation of the parts of the sign and the state of affairs it signifies (e.g. a picture indicates the scene it does through having corresponding elements in corresponding relationships); the second mode of a sign, an index, represents via a physical existential relation between it and the relevant state of affairs (e.g. a sundial signifies the time of day through a physical connection between that which is modelled, the sun, and the sign, the sundial); the final form of signs, symbols, represent via formal rules (e.g. a proposition represents a scene via the formalised grammatical rules of its structure). Peirce (1906, p495) clarifies his typology thus:

> an analysis of the essence of a sign, (stretching that word to its widest limits, as anything which, being determined by an object, determines an interpretation to determination, through it, by the same object,) leads to a proof that every sign is determined by its object, either first, by partaking in the characters of the object, when I call the sign an Icon; secondly, by being really and in its individual existence connected with the individual object, when I call the sign an Index; thirdly, by more or less approximate certainty that it will be interpreted as denoting the object, in consequence of a habit (which term I use as including a natural disposition), when I call the sign a Symbol.

With our previous discussion of Craikean models in mind, it can quickly be perceived how Peirce's analysis influences Johnson-Laird's theory of analogical, or iconic, models: It both strengthens and supports his contention that mental representation functions through an analogy – or a shared (i.e. natural and structure preserving) relation-structure – with a specific state of affairs.

Therefore, and to summarise, to define the fundamental nature of mental representations, Johnson-Laird draws on the work of both Kenneth Craik and Charles Sanders Peirce. With a reliance on these sources, it is argued that the internal structure of a mental representation (i.e. the elements of the model and their relations) is a mirror of, or is identical to, the state of affairs or situation being

---

[36] References made in this fashion refer, unless specified, to the numbered paragraphs found within Peirce's collected works, 1931-1958.

[37] It needs to be clarified that Peirce drew a number of different analyses into the operation and manifestation of signs throughout his work. The one that concerns this thesis is, as Peirce (1906, pp496-497) affirms, "only one of ten different divisions of Signs which I have found it necessary more especially to study", but it is the one which occupies Johnson-Laird (2002; 2006; 2008; 2010) and which has been described by Burks (1949, p673) as "the simplest of his [Peirce's] many classifications of signs, and is, moreover, the most important of them all, for it contains the essence of even the most complicated of them without sharing their repetitive and unwieldy character. Historically, the more complicated classifications developed as expansions of the simpler one". As such, this will be the only division of signs made by Peirce to which this research will give attention.

modelled and preserves or maintains this relation if submitted to the same operations. This iconic characteristic of mental models is a facet of Johnson-Laird's philosophy that continually manifests itself throughout his work via the use of synonyms for iconicity to express the structure of his mental models (synonyms such as 'analogical', 'isomorphic', 'parallel' and 'simulation' (see Johnson-Laird 1980; 1983; 2008a; 2008b; 2010)) and can be seen in quotes such as the following:

> Mental models are iconic insofar as possible. What "iconic" means is that their structure corresponds to the structure of what they represent[.] (Johnson-Laird and Khemlani 2013, p9).

Now that we, via its influences, have explored the nature of the mental models Johnson-Laird has posited in the representational level of Marr's tri-level hypothesis, we can now move on to the procedures or computations that are said to be performed on these types of representation to process them, i.e. we are able to investigate how reasoning is said to be produced under Johnson-Laird's analogical mental model framework.

Reasoning and mental models

The distinctive feature of the operations performed on Johnson-Laird's mental models, when used to explain reasoning, is that they do not rely upon any formal rules of inference, or any other axiomatic formal logic.[38] Instead, in place of these formal inferential rules, mental models rely upon their iconic nature and computations to construct, investigate, compare, conjugate, remove and update models to explain the two kinds of reasoning employed by individuals: Induction and deduction (see Johnson-Laird 1980; 1983; 1993; 1994a; 1994b; 1999; 2006; 2008a; 2008b; 2010).

Deduction

In the case of deduction – which moves from the general to the specific – Johnson-Laird relies on the aforementioned Peircean iconicity of mental models. The motivation for this reliance can be found in Peirce (§279), who, during his investigation into icons, identifies that:

> a great distinguishing property of the icon is that by the direct observation of it other truths concerning its object can be discovered than those which suffice to determine its construction. Thus, by means of two photographs a map can be drawn, etc[...] This capacity of revealing unexpected truth is precisely that wherein the utility of algebraical formulae consists, so that the iconic character is the prevailing one.

This means that the iconic nature of mental representations allows conclusions that are not explicitly stated in any original premises to be drawn from subsequently constructed models. For example, if given the premises:

---

[38] For Johnson-Laird's reasoning behind this abandonment of an innate formal logic see the work of Wason (1966), Wason and Shapiro (1971) and Wason and Johnson-Laird (1972). These studies demonstrate that, when trying to solve a problem, the content of that problem affects participants' reasoning, a phenomena that could not possibly occur when operating under a formal logic, which sees no difference between 'If *x* then *y*' and 'If I go to Manchester, I will take the train' while, as Wason et al's experiments demonstrate, people do.

1) Katie is to the left of Phil

and

2) Kath is to the left of Katie

A mental model akin to the following can be produced:

Kath   Katie   Phil

After investigation of which, it can be deduced that:

3) Kath is to the left of Phil

That is, a valid conclusion not explicitly stated in the premises can be found or *deduced*.

This reliance upon iconicity, though, is just one strategy that can be used to produce conclusions through a deductive process. Where more complex propositions – such as logical and conditional puzzles – are concerned, additional strategies can be employed. These rely on suppositions, which are mental models that:

> are assumed for the sake of argument and which must be "discharged" if a derivation is to yield a conclusion. One way to discharge a supposition is to make it explicit in a conditional conclusion (conditional proof), and another way is to show that it leads to a contradiction and must therefore be false (reductio ad absurdum) (Johnson-Laird and Yang 2008, p341).

Meaning, that we can reach valid conclusions by creating an initial model consistent with one of the proposition's premises (taking it to be true for the sake of argument) and employing computational processes to update that suppositional model or conjugating it with another model consistent with one or more of the other premises. Once this has been achieved, we can then interrogate the resulting model to find if it is valid and, if it is, what conclusions can be read off from it, or, if it is not valid, conclude that a contradiction has been reached. Once a conclusion, of any kind, has been attained, models constituting counterexamples consistent with the original premises can be created to ensure the validity of any such conclusions (see *Ibid.*, p345; Johnson-Laird and Khemlani 2013, p14).

Before we continue, however, it is important to highlight the fact that when a model is constructed from a proposition, it constitutes a generalisation of that proposition's premises. For an infinity of nuanced models of, to take our previous example, "Katie is to the left of Phil and Kath is to the left of Katie" can be drawn. For example:

Kath Katie   Phil

and

Kath   Katie   Phil

or an infinity of other, subtly different, representations. As such, when we perceive a premise we draw only the one model of it, and that generalised model "captures what is common to [the] swarm of possibilities" (Johnson-Laird and Yang 2008, p345).[39, 40]

Induction

So far, all that these procedures and models have explained is the process of deduction, but there is another mode of reasoning requiring an explanation before Johnson-Laird's models can be said to have fully accounted for the phenomena: That of induction (or reasoning which moves from the specific to the general). Johnson-Laird (1994a) explains this form of reasoning with a central reliance on the addition (i.e. update) procedure or computation we have just encountered (although, as we shall see, it must also make use of procedures for the deletion and conjugation computations). This central operation, addition, updates an existing representation with information garnered from subsequent premises or prior experiences in a process that can create two types of induction: Specific and general:

> **Specific induction:** These concern a particular event, process or object. For example, if you have trouble opening your car door on a frozen winter morning, and have had previous experience of that particular car's lock freezing and sticking in cold temperatures, then the information from the models of prior experience (of the lock's likelihood to stick) will be added (more accurately, conjugated) to the current model of the inoperable door (i.e. the representation will be updated with information from another mental model) to create the conclusion, *induction*, that the lock has again frozen in this instance (see *Ibid*, pp14-16).

> **General induction:** These concern a class of events, processes or objects and occur through a slight adaptation of the process we have just witnessed during specific induction. In the case of generalised inductions, information is added to a model that invalidates it, removing it from the suite of available representations, thereby leaving only the one, now general, representation (see *Ibid*, pp16-20). For example, if it is found that a variety of particles (either

---

[39] This creation of only one generalised model is a fact taken to be one of the central assumptions of Johnson-Laird's mental model theory (see Johnson-Laird 2010, p2). The theoretical difficulty of how one instance can stand for an infinity of cases was, until relevantly recently, explained by Johnson-Laird's reading of the *Philosophic Investigations* and by reference to Wittgenstein's discussion of 'family resemblance', stating that a single model is capable of representing a multitude through the fact that it is a schema which possesses the criteria of the concerned state of affairs (see Johnson-Laird 1983, pp191). Recently, Johnson-Laird (2002, p83) has updated his response to this theoretical question, adding to the concept of Wittgensteinian schema with an explanation based on the work of Charles Sanders Peirce and 'iconicity'. Stating "a mental model captures what is common to all the different ways in which a possibility might occur. Hence, a mental model of the spatial proposition just given ['The circle is to the right of the triangle'] represents what is common to any situation in which a circle is to the right of a triangle, but it represents nothing about size, colour, distance apart, and other such "accidents." The model can be updated to take into account information about these properties and relations. But no matter how many details are added to a description, it is always consistent with many distinct states of affairs." As such, it is now claimed that we create the one mental schema from a proposition, a schema from which we are able to "distinguish between accidental properties[…] and essential properties" (Johnson-Laird 2006, p139).

[40] The motivating reasons behind the creation of only the one mental model of a state of affairs in such cases, instead of a plurality, is one of processing power, i.e. it is a strategy designed to reduce the load and processing complexity placed on our finite brains during reasoning tasks (i.e., a finite brain cannot hold an infinitude of models).

A- or B-type particles) released from a nuclear reaction are damaging to tissue, a number of potential conclusions can be drawn, including:

 (1) Any one of the particle types (either particle A or particle B) is damaging to tissue.

and

 (2) Particle A and particle B only damage tissue when operating in unison.

Information can then be added to the mental representations of these conclusions, such as: "Particle A damages tissue when released on its own", which would support one model (model (1) in this instance) and eliminate the other (model (2) here). For:

> an assertion of the form:
>
> If A & B then C
>
> corresponds to a set of models that includes:
>
> A [not]B [not]C
>
> When this model is eliminated, the resulting set is equivalent to:
>
> If A then C…
>
> The operation of eliminating a model – by adding information that contradicts it – suffices for any generalisation, because generalisation is nothing more than the elimination of possible states of affairs (*Ibid.*, p18).

Again, as in deduction, when a conclusion is reached through any kind of induction, counterexamples consistent with the original premises will be created to ensure that any conclusions reached cannot be refuted and are therefore valid (see Johnson-Laird 2010, p7; Johnson-Laird and Khemlani 2013, p21).

Human error

Now the process of deduction and induction have been accounted for, all that is necessary to explain is the troubling fact that humans regularly err in making logical deductions and inductions. A fact that is not apparent in the strategies described above. As such, Johnson-Laird needs to clarify how this possibility is said to occur during human reasoning under an analogical representational framework. What is central to his answer here is the realisation that the human mind is not unlimited in its capacities; a limitation which leads to a preference, by memory, of dealing with a limited number of models at any one time (Johnson-Laird 1994a, p25). Due to this shortcoming, the mind is forced to cut corners when creating and manipulating mental models during problem solving (see Johnson-Laird 1994a; 2006; 2008a; 2008b; 2010). We have already seen one of the strategies employed by the mind for lessening the burden put on our memory: Creating only one generalised model where a plurality can be produced. An additional tactic is referred to in the work of Johnson-Laird as the principle of

truth (see Johnson-Laird 1980; 2010; 2013; Johnson-Laird and Khemlani 2013).[41, 42] This principle states that, to minimise load, the brain represents only what is true in a proposition's premises, that:

> mental models represent only those states of affairs that are possible given an assertion, and that within each of these possibilities they represent a clause in the assertion, whether it is affirmative or negative.
> (Johnson-Laird 2010, p209)

For example, if given the disjunction 'either he took the bus or he walked' and the fact that 'he walked', only a representation of 'him walking' will be produced. The false clause, 'he took the bus', will be completely abandoned. This simple strategy to save load by reducing information achieves its end, but its implementation has a devastating impact in the mind's ability to unerringly perform certain inferences. To explain this, Johnson-Laird (2008b, p5) gives the following example:

> Consider this problem, for instance:
>> • Either Jane is kneeling by the fire and she is looking at the TV, or else Mark is standing at the window and he is peering into the garden.
>> • Jane is kneeling by the fire.
>
> Does it follow that she is looking at the TV?
>
> Most individuals say yes[…] They overlook that when the second premise conjunction is true, the first conjunction is false, and that one way in which it can be false is when only its first clause is true, i.e. Jane is kneeling by the fire but *not* looking at the TV.[43]

That is to say, the "principle of truth reduces load on our memory, and it seems innocuous. Yet, it can lead us into the illusion that we understand a description that, in fact, is beyond us" (Johnson-Laird 2011, p480). Consequently, it is discovered that the heuristics the mind uses to facilitate reasoning, as well as to manage its inherent limitations, are the underlying causes of common errors (see Johnson-Laird and Khemlani 2013, pp11-13).

Summary

To review, how the analogical mental model theory operates at the representational level of cognition is akin to that of Turing's machine: It relies upon a suite of formal symbols and a finite set of operations to process them and to convert sensory input into intelligent action.[44] In the case of a Turing machine, this symbol system was constituted by 0s, 1s and the four commands to the read/write head, which

---

[41] This principle of truth is another of the central assumptions of Johnson-Laird's mental model theory (see Johnson-Laird 2010, p2).

[42] This inclusion of the principle of truth in Johnson-Laird's model theory should not be confused with the assertion that only what is the case can be modelled, rather than what is said to be true *within a proposition*. If the former state of affairs were the case, then no counterfactual could be modelled nor any discussion of counterfactuals be meaningful; both things which are accounted for within the model theory as a whole. Furthermore, as Goldvarg and Johnson-Laird (2001, p584) state, in certain instances the brain can overcome its prejudice toward the principle of truth and create 'complete models' that are capable of "overcoming the principle of truth" and representing states of affairs that are stated to be false within a proposition. These 'complete models' are also referred to as "fully explicit models" (Khemlani, Orenes and Johnson-Laird 2012, p543; Johnson-Laird and Khemlani 2013, pp11-13).

[43] Johnson-Laird is appealing to evidence produced from experiments he performed with Walsh for these claims (see Walsh and Johnson-Laird 2004).

[44] As such, Johnson-Laird's theory entails, as Fodor (2000, p12) identifies, a "nativism about[…] computational mechanisms".

cover how to manipulate them appropriately; with minds, for analogical cognitive scientists, the following facets constitute the symbol system:

Representations: The mind employs analogical, or iconic, mental models akin to those posited by Craik (1967) and Peirce (1906; 1931-58), wherein the elements of the representation, and the relations between them, mirror those of the state of affairs which is being represented. Furthermore, these models are natural and structure preserving, in that only those components and relations found in the worldly state of affairs are found in the model and, further, that the model reacts to changes, stimuli and operations as does those states of affair (see Johnson-Laird 1980; 1983; 1988; 1993; 1994a; 1999; 2002; 2006; 2008a; 2008b; 2010; 2013).

Computations/Operations: To process its analogical representations, thereby converting sensory input into an appropriate output (i.e. an intelligent action or propositional attitude), the mind employs computational procedures to construct, investigate, update, conjugate, compare and remove mental models (see *Ibid.*; Johnson-Laird and Yang 2008).

Now we have explicated analogical representations and their explanation of the processes operating behind reasoning, it will be appropriate to detail how this system accounts for the appearance of meaning in natural language propositions, an account that Johnson-Laird has termed procedural semantics.[45] For reasons that will soon become apparent, this account of semantic content will play a crucial role within the context of this research. Consequently, Johnson-Laird's procedural semantics will be given a greater attention over the similar work of others, such as Jerry Fodor, and, therefore, will be given its own dedicated section.

§iv Procedural semantics

Throughout its history, procedural semantics has not been a steady, unchanging theory. Instead it is one that has developed and been built upon over its lifetime (this was an evolution acknowledged during our earlier analysis and explication of the influence of Craik and Peirce on Johnson-Laird). However, this does not mean that the theory of today is unrecognisable from the theory of the 1970s. Rather, it has developed along with Johnson-Laird's wider philosophy, maintaining many of its original positions but adapting to subsequent discoveries. Consequently, this section will follow Johnson-Laird's work chronologically: Initially concerning itself with the first articulations of the theory and

---

[45] It should be noted, however, that in recent years (i.e., since his shift of focus from meaning toward reasoning), Johnson-Laird has dropped his use of the epithet 'procedural semantics' for his account of meaning. However, even though this is the case, natural language sentences are still said to be meaningful due to the procedures mental translations perform on mental representations under Johnson-Laird's contemporary philosophy (Johnson-Laird 2006, pp23-24). As such, the title 'procedural semantics' is still appropriate and will be applied throughout this thesis.

then demonstrating its development over the years (e.g., adapting to such things as the theory of analogical models), until it becomes the complete correspondence theory of meaning we find today. As such, this section will:

1) Defend the sole focus on mental representations taken when explaining meaning under the cognitive science paradigm.

2) Undertake a discussion of the earliest 'compile and execute' version of Johnson-Laird's theory.

3) Give a report of the changes made to this early theory along with Johnson-Laird's development of analogical mental models.

4) Analyse how these analogical mental models are themselves meaningful through a Peircean structural correspondence theory.

Representational abstraction

Before commencing our discussion of Johnson-Laird's account of meaning, it is first necessary to clarify why this discussion will focus solely upon the middle tier of Marr's tri-level hypothesis (as did our earlier analysis of Fodor's LoT). This is necessary because, as Marr (1982, p25) identifies, "a correct explanation of some psychophysical observation must be formulated at the appropriate level [of the tri-level hypothesis]". As such, when considering the problems faced when explaining semantic content under the cognitive science framework, it is crucial that any putative answers are described at the correct level of abstraction. Consequently, it is informative to review comments made by Pylyshyn (1989, p57) concerning the definition of the representational (or, as he terms it, the symbol) level:

> The semantic content of knowledge and goals is assumed to be encoded by symbolic expressions. Such structured expressions have parts, each of which also encodes some semantic content. The codes and their structure, as well as the regularities by which they are manipulated, are another level of organization of the system.

Therefore, any explanation of meaning, within the tri-level paradigm, should be concerned only with the operation of the representational abstraction to the detriment of both the computational and physical abstractions. Why this is so becomes apparent after consideration of the operation of the three levels of abstraction: The physical level must surely remain mute on the topic of semantics (on pain of imbuing inert pieces of matter with semantic content) and the computational level can only conclude in an awareness of the phenomena's presence, an answer which would merely beg the question. As such, Johnson-Laird's discussion of meaning (as well as Fodor's and any future discussion within this thesis) will be solely pitched at the representational tier of Marr's tri-level hypothesis.

Early 'compile and execute' semantics

Returning to Johnson-Laird (1977), his procedural semantics was defined during its first articulations in the 1970s, as a theory of meaning based on a metaphor with the 'compile and execute' strategy employed by computers when interpreting programming languages. Under this metaphor, semantics is explained by the initial translation of natural language propositions into a 'program' in a hearer's mental 'processor' (*Ibid*., p191) (i.e. propositions need first to be compiled before being understood). These stored 'programs' can then be converted into a state of affairs, such as an action, by its translation into some relevant and appropriate response by the hearer (i.e. the compiled proposition can be executed) (*Ibid*., pp206-207). For example, if given the command "Shut that door", an individual will first translate the command into some mental state understandable by the mind, before employing that mentally understandable state to initiate the correct series of bodily movements to close the door, i.e. the command will be compiled and then executed.[46]

It is important to note, however, that this first expression of procedural semantics suffers from some serious deficits. For, during this period, Johnson-Laird (1976, p635) ultimately explains semantic content by using the lexical decompositional paradigm prevalent at the time, wherein:

> The most elementary constituents of meaning are[… taken to be] various ineffable mental operation and states: these are the fundamental "particles" that make up the everyday use of language.

This means that, for the Johnson-Laird of this period, a word is only meaningful once it has been mentally broken down (i.e. compiled) into the 'fundamental mental particles' from which it is comprised. This is a theoretical (and syntactical) position that, as Johnson-Laird (1993, p348) later highlights, begs the question, for all it has achieved is to explain language by merely relating it "to itself" (see also *Ibid.* 1988).

Craikean mental models

Due to such difficulties, Johnson-Laird's hypothesis undertakes a number of subsequent improvements, the most significant of which is the development of a fully functioning analogical representational system in the 1980s (i.e. Johnson-Laird's adoption of the theory of Craikean analogical mental models). This evolution is crucial to the growth of procedural semantics away from its earlier unsupportable incarnation, with the meaning of propositions now being translated (i.e. 'compiled') operations or computations that are performed on mental models (as described in this

---

[46] It is important to note that the mind must 'compile' the natural language command into a state understandable by the mind before it can be executed, as the mind cannot be said to operate using natural language. If this were asserted, then the question of how a vast number of natural languages arose must be answered. This position would also lead any adherent into accepting the Sapir-Whorf hypothesis and its attendant difficulties (see Gumpertz & Levinson 1996). See also Fodor (1975, p56), who makes similar comments regarding his own, comparable, 'sentence understander and logic'.

chapter's prior section) rather than merely broken down into their 'fundamental elements.' Or, as Johnson-Laird (1980, p91) has it:

> the meanings of words are decompositional procedures that relate to mental models of the world, and, in particular, on the use of lexical procedures that interact with the general procedures for constructing, manipulating and evaluating mental models.

However, with this adoption of Craikean analogical models in the 1980s, Philip Johnson-Laird's procedural semantics undergoes a further adaptation. For, the translation (or compilation of propositions into procedures to manipulate mental models) now requires an additional step between compiled propositions and action: That of a propositional representation (or 'lexical procedure') that is understandable and executable by the mind and that determines which are the correct procedures to affect a hearer's analogical models. As such, procedural semantics is now:

> a theory of comprehension in which there are two stages. In the first stage, a superficial understanding of an utterance gives rise to a propositional representation, which is close to the surface form of a sentence. This symbolic representation is constructed in a mental language that has a vocabulary of comparable richness to that of natural language[…] The second stage of comprehension, which is optional, makes use of propositional representations as a basis for the construction of a mental model, whose structure is analogous to the state of affairs described by the discourse. Hence the recovery of a propositional representation is a necessary precursor to the construction of a mental model (Johnson-Laird 1983, p244).

Although this intermediate stage of propositional representations links words to mental models and explains comprehension and how a sentence is interpreted, it does not account for the semantic content present in the original proposition. All that has currently been achieved is the translation of one set of symbols, language, into another, mental models, thereby only relating signs to further signs. This final, crucial, step of grounding meaning is accomplished by linking mental models to the world. For, as Johnson-Laird (1987, p199) affirms:

> What has sometimes been underestimated[…] is the importance of reference, or more properly, its psychological correlate: the representation of specific referents, real or imaginary.

Therefore, the early, Craik influenced, Johnson-Laird provides an explanation of the meaning of propositions based on a correspondence to a state of affairs in the world. Or, as Johnson-Laird (1983, p265) articulates this position:

> once truth conditions have[…] been taken care of there is no need to give a separate account of semantic properties of expressions. They are emergent properties of the truth conditions (*Ibid*., p265).

As such, and to clarify, with the development of Craikean analogical mental models there is a sea change in Johnson-Laird's account of meaning and the operation of procedural semantics between the 1970s and 1980s: That of the abandonment of lexical decomposition for the adoption of a prototype, or embryonic, correspondence, reference-based, theory of meaning.[47, 48]

Moreover, the adoption of such an embryonic correspondence theory of meaning not only introduces a structural referentialist, externalist, account of meaning into Johnson-Laird's philosophy of mind, it also provides a number of benefits to his account of semantic content as a whole. For a structural correspondence theory of meaning is capable of explaining such things as: 1) The productivity of language and thought; 2) the systematicity of language and thought; and 3) concept learning (see Fodor 1975, p31; 1990, pp16-19; 2010, pp17-20). For:

1) **The productivity of language and thought:** Productivity, as a concept, covers the phenomenon that language and the mind are systems capable of producing an infinite number of unique and distinct propositions and thoughts. This potential of finite individuals to produce an infinity of representations can only be explained if this potentiality derives through the application of a finite number of procedures. As such, just as a strict grammar can produce an infinity of sentences, so a finite number of computations can be applied, following strict operations, so as to produce an infinity of mental models, which constitute an infinity of thoughts that can generate an infinity of propositions.

2) **The systematicity of language and thought:** Systematicity covers the phenomenon that individuals are able to infer that a system possesses certain representations from the fact that other representations are found in its repertoire. For example, if an individual is capable of producing the representation 'John loves Mary', it can also be inferred that that individual is also able to generate 'Mary loves John'. A grounded theory of meaning is capable of explaining this as, under this account, the representations underlying both 'John loves Mary' and 'Mary loves John' are generated by a strict grammar, a grammar that, if capable of generating the one, necessitates the ability to generate the other.

---

[47] Interestingly, this hypothesis of Johnson-Laird's (that meaning under a mental model framework must operate via some form of correspondence theory), is one which was foreshadowed by Craik (1967, p100), when he first articulated the analogical model theory.

[48] As such, holding such a referentialist semantics inevitably leads Johnson-Laird to the position that singular terms and thoughts, i.e., those that concern an ontologically real state of affairs, are Russellian in nature, i.e., must follow Russell in holding that "someone who uttered a sentence containing an empty singular term would fail to say anything, in the sense that he would fail to express a thought" (Evans 1982, p12). For reference necessitates a referent and representation necessitates something to represent, therefore, if semantics is wholly based on reference and representation, propositions and thoughts need a referent to be meaningful.

3) **Concept learning:** The ability to learn a concept prevents an individual having to be born with the mental representations for all the concepts with which they will be acquainted and can meaningfully represent, i.e. it prevents the need for every concept to be native**.** A correspondence theory of meaning can account for learning new concepts by the generation of a mental model that corresponds to a state of affairs that instantiates a concepts. This, therefore, avoids the necessity of a brute nativism within Johnson-Laird's philosophy. (However, it still implies some level of nativism to account for the fundamental procedures themselves, as well as, seemingly, the building blocks of mental models, from which new representations are constructed.)

Peircean mental models

The use of the terms 'prototype' and 'embryonic' above, in reference to Johnson-Laird's procedural semantics of the 1980s, is deserving of explanation. Although the procedural semantics of this period is clear in its reliance on reference and truth conditions (e.g., "the truth conditions of words are intimately dependent on the mind's ability to relate representations to the world" (Johnson-Laird 1987, p207) and "A model *represents* a state of affairs and accordingly its structure[…] plays a direct representational or analogical role. Its structure mirrors the relevant aspects of the corresponding state of affairs in the world" (Johnson-Laird 1980, p98)), Johnson-Laird's position at this time is not yet clear on is the complete details of the nature and operation of this reference, i.e., it is not yet fully clear how, exactly, the mind relates *all* types of mental model to a state of affairs. This prototypical character of Johnson-Laird's semantics from the 1980s, and through the 1990s, is especially apparent when considering the dearth of information given by Johnson-Laird during this time concerning negation and counterfactuals, or those mental representations of false states of affairs, where, *eo ipso*, a representation does not correspond to a physically real state of affairs.[49]

This situation (wherein the intricacies of the operation of reference between mental model and world is unfortunately left unexplained) remains within Johnson-Laird's theory of meaning through the 1980s and 1990s, and is only resolved when Johnson-Laird (2002) connects his work to that of Charles Sanders Peirce's, with his adoption of the terms 'icon' and 'iconicity'. With this move, Johnson-Laird therefore ties the operation of reference under his theory of meaning to Peirce's own and completes his previously nebulous reference based procedural semantics. Now, taking from Peirce (§§2.282-2.279), procedural semantics relate models to the world "in respect to the relations of their parts[… That is,] an analogy between the relations of the parts of each", but this also, following Peirce, include

---

[49] However, the commitment of Johnson-Laird (1983, p173; 1994, pp196-197) to ontologically real modal universes could, potentially, be used to explain counterfactuals through possible world semantics.

"signs[...] used for expressing negation" (§3.385) (however, this strategy has subsequently been extended, so as to include "other abstract concepts, such as "possibility, "truth", and "obligation" that transcend iconicity" (Johnson-Laird and Khemlani 2013, p11)). Consequently, mental models are now not only said to derive their meaning as an emergent property from the fact that their parts, and the relationship between those parts, mirror a physical state of affairs found in the world (e.g. through a congruent ""correspondence" theory" of meaning (Johnson-Laird 2006, p231-232)) but also via mental signs attached to these mental models. For, with an influence from the work of Peirce (§§3.385-3.386), Johnson-Laird is able to provide a meaningful account of false propositions based on the concept of negation (see Johnson-Laird 2002; 2006; Johnson-Laird and Khemlani 2013). For example, and as Barres and Johnson-Laird (2003, p2) aver:

> Individuals have the option of using negation as an operator on a model. Hence, they can represent the possibilities in which the assertion ['the square is above the triangle'] is false by constructing a model of the form:
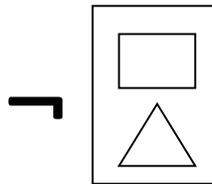
> where "¬" denotes negation.

Therefore, when confronted by the apparently meaningful, yet false, proposition 'You are not reading this sentence', you do not understand it by converting it into a representation of yourself not reading the proposition, which does not correspond to any state of affairs, but by converting it into a representation of yourself reading the sentence, which does correspond to a state of affairs, and then, so it is argued, attaching a mental symbol for negation to reverse its truth values. Thereby producing a meaningful mental model that possesses the correct truth values as the false proposition yet still corresponds to a part of the world (see Khemlani, Orenes & Johnson-Laird 2012, p546). Therefore, and to clarify, where counterfactuals are concerned, under a Peircean correspondence theory of meaning, a proposition may result in the production of a model that represents the pertinent parts of the world concerned in the appropriate relation, but with an extra, purely mental, symbol to signify that the state of affairs are not the case and to reverse the truth values (it should be remembered, however, that this tactic of the inclusion of mental symbols to iconic representations is also applied

to "other abstract concepts" (Johnson-Laird and Khemlani 2013, p11), e.g. "The [model] theory allows that models can be tagged with numerals denoting their probabilities" (*Ibid.*, p23)).[50, 51]

Summary

Therefore, and to summarise, procedural semantics has, fundamentally, remained true to the compile and execute strategy it embodied in the 1970s, but has developed over the years along with Johnson-Laird's wider philosophy. This process of development has left us with the following complete account of semantic content:

1) A proposition first needs to be interpreted, needs to be compiled. This is achieved by the following:

   a. A proposition is translated into a propositional representation (i.e. it is compiled into a rich mental language, executable by the mind).

   b. This propositional representation is used to execute relevant procedures upon the suite of mental models carried by the hearer, i.e. the rich mental language of the mind is used as the basis for creating, updating, conflating, verifying and deleting mental models.

2) Although Johnson-Laird's accounts for the meaning of propositions by relating them to analogical mental models, this begs the question: 'how are mental models themselves meaningful?' Johnson-Laird answers this question by explaining that a mental model receives its meaning via its reference to the world. This is achieved via several means:

   a) In the case of true states of affairs and proper names (i.e. that which is the case), the model is connected to the world via its iconic status. That is to say, the model is meaningful as it corresponds to the parts, and the relationship between these parts, of a real section of the world.

   b) Where counterfactuals are concerned (i.e. what is not the case) Johnson-Laird appeals to the same process of iconicity used to explain factual comments but with the addition of the mental symbol for negation, to reverse the truth values of a state of affairs.

---

[50] In the case of compound, or multiple-clause, negative expressions, such as 'You are not reading this sentence or you are not seated', "individuals[…] have to enumerate the various possibilities one at a time" (*Ibid.*, p543. See also Khemlani, Orenes and Johnson-Laird 2014). That is to say, that "individuals represent a disjunction such as: *Viv is in Rome or Pat is in Rome or both*, by considering each of its possibilities separately. That is, they construct three mental models representing these possibilities" (Johnson-Laird 2002, p86).

[51] Note, however, that this theory of negation and its related account of counterfactuals does not explain the occurrence of *misrepresentation* (i.e. it cannot account for the occurrence of inaccurate representations of a specific and determinable state of affairs). Furthermore, with the use of isomorphism to naturalise reference, as is done so under this Peircean model, Johnson-Laird is also tasked with explaining why, if a mental model is an icon of a state of affairs, then only the former is a representation of the latter and not *vice versa*. This latter difficulty is founded on a problem with symmetry in isomorphism and, therefore, requires an asymmetrical account of iconism – wherein, a mental model is reliant on a state of affairs, but not *vice versa* – for a solution. However, this discussion of the naturalisation of reference is beyond the scope of the current discourse, but this difficulty, and others, will be handled in due course.

3) Once a proposition has been interpreted (compiled), it can be acted upon (executed). As such, the resulting mental representation of a proposition – and its resulting intermediate propositional representation – is used as a basis for action.

Consequently, Johnson-Laird's current and full account of the meaning of sentences (here captured by the label of procedural semantics) is a multi-stage process: Firstly, a proposition is converted into an intermediary propositional representation (or lexical procedure) understandable by the mind. This, in turn, is employed by the mind to appropriately update, by the means of set operations or computations (details of which are found in the preceding section), its suite of mental models that correspond to states of affairs in the world.[52]

Conclusion

This chapter has moved from the more general aspects of cognitive science to specific accounts of cognition and meaning: Investigating the overarching goal and approach of cognitive science, as well as individual representational explanations of cognition and meaning that can be found within that approach. We have seen that the field as a whole places a focus on a post-positivist paradigm and an interdisciplinary research approach when investigating the workings of a folk psychological conception of cognition and employs a functionalist RTM and a tripartite system of explanation when describing these workings. After covering this broader project of the field, we then saw in detail the workings of the two major explanations of mentality currently operating within this paradigm: 1) Propositional; and 2) analogical representations. We exemplified the former of these two approaches, propositional representations, by the work of the philosopher Jerry Fodor. Finding that such approaches posit a finite number of atomistic mental entities that are organised syntactically in a manner akin to natural languages (so much so that these entities are referred to collectively as Mentalese). The semantic content of such representations are accounted for by a causal theory of reference, wherein an individual mental entity 'locks on' to its real world referent once it is learnt. The latter, analogical, approach was characterised with an investigation into the work of psychologist Philip Johnson-Laird. It was found that this approach operates at the middle, representational, abstraction of cognitive science's tri-level hypothesis and accounts for cognition through the operation of a suite of 'iconic' representations posited at the representational abstraction. By this Peircean term, Johnson-Laird (2006, p427) means **"A representation [of the world] with relations among its parts that correspond

---

[52] It should be clarified that all these explanations ignore the obvious question of what, exactly, are these 'elements' or 'parts' to which icons correspond? Are they the medium sized objects we commonly find in the world, such as a chair? The components that make up medium sized objects, such as the chair's legs, back, cushion etc? The atoms that make up these components? Sub-atomic particles? It should also be noted that these same questions arise with Fodor's referentialist account of mental representations.

to those in what it represents". Consequently, Johnson-Laird (2002) argues that the meaning of natural language propositions are ultimately derived via a congruent correspondence theory of meaning, with natural language being translated into operations to generate iconic mental models that refer to worldly states of affairs via a shared logical form.

Chapter Two:
Critiques of cognitive science and RTM

Introductory remarks

Although the claims found in the previous chapter's exposition of cognitive science is common currency within that field (and others, such as philosophy of mind and AI) and can be (or, similar claims can be) readily found within the relevant literature, it is necessary to highlight the fact that these views and positions are not uncontroversial or uncontested. In fact, adherents of cognitive science, as previously defined, have found that the majority of their foundational claims have been challenged at some point. For example, commentators have criticised the position due to its commitment to:

> **Folk psychology and propositional attitudes:** As we have seen, the RTM is based on the common sense assumption that propositional attitudes (i.e. belief and desire states) have the causal and semantic properties we suppose them to have and, further, takes these states to be the *explanandum* of its research. However, some challenge this claim and argue that propositional attitudes do not have the reality cognitive scientists claim. Two examples of this anti-realism concerning propositional attitudes are the instrumentalism of Daniel Dennett and the eliminitivism of Paul and Patricia Churchland. The former of these, Dennett (1989; 2000), argues that, strictly speaking, there are no propositional attitudes, yet they are, from the point of view of an observer in the 'intentional stance', useful fictions for predicting behaviour and are akin to concepts such as a centre of gravity.[1] In a, superficially, similar way, while ultimately having a quite different agenda, Churchland (1981), appreciates the predictive success of propositional attitudes but argues that that criterion alone is not enough to establish a

---

[1] Although it will be said in the main text that the example arguments portrayed here are not the focus of this research and, as such, will not be critically analysed or given their deserved attention, this research will take advantage of this space outside of the main text to make comment upon them when is deemed appropriate. For example, it should be noted that Dennett's instrumentalism can be tasked with answering the charge that, surely, predictive success implies or infers a theory's accuracy, i.e. with convincingly answering the question 'if propositional representations are not true, why do they work?' (cf. Dennett 1989; 2000). Furthermore, Dennett's position inevitably concludes in a behaviourist view of the mind, for its ascription of mental content to a system is based solely on behavioural matrices (*Ibid.* 2000 pp59-61). However, Dennett (*Ibid*, p72) himself claims that this conclusion is "perverse", arguing that the internal complexity and variegated sensory connection humans have to the world "yields a robust version of belief after all" (*Ibid.*). Nevertheless, this does not seem convincing, for the internal complexity of a system does not alter the foundational claim of instrumentalism that:

> I [Dennett] will argue that any object – or as I shall say, any system – whose *behaviour* is well predicted by this strategy [the intentional strategy or adopting the intentional stance] is in the fullest sense of the word a believer. What it is to be a true believer is to be an intentional system, a system whose *behaviour* is reliably and voluminously predictable via the intentional strategy [Emphasis mine]. (*Ibid.*, p59)

theory's scientific accuracy. Rather, if other measures, such as "coherence and continuity with fertile and well-established theories in adjacent and overlapping domains [i.e. with the 'traditional' sciences]" (*Ibid*, p73, see also *Ibid.*, p76), are also considered, propositional attitudes and folk psychology will be found, so it is argued and following the language of Lakatos (1984), to be a 'degenerative', 'decadent' or 'stagnant' science akin to alchemy, "and has been for millennia" (*Ibid.* p75). As such, what is needed is a replacement (or, more accurately, a 'displacement' or 'transcendence' (*Ibid*, p74)) of propositional attitudes with a scientifically respectable alternative, such as neuroscience.[2]

Representations: As the previous chapter highlighted, a central pillar of the RTM is that our belief/desire states are explained by mental representations. However, some argue that this positioning of representation within cognition is fallacious. For example, one line of criticism uses the posited necessary and sufficient conditions of a representation that, in general, it is required to make the distinction between vehicle and content, with the former being the representational form of the picture etc. (e.g., the vehicle of this proposition is ink and paper or pixels on a VDU), and the latter the message that the picture etc. conveys. However, where thoughts are concerned, some argue that there is no vehicle, just content, i.e. thoughts are said to be all message and no medium. Consequently, we cannot make the distinction between vehicle and content with mental representations. Consequently, thoughts cannot meet the necessary and sufficient conditions to be considered a form of representation. Consequently, so some conclude, Marr's tri-level hypothesis and the RTM are incorrect (see Hacker 2013).[3]

Physicalism: During our earlier discussion of Marr's tri-level hypothesis, it was identified that the RTM is a physicalist interpretation of the mind. That is, cognitive scientists assert that all phenomena (including the intelligent action of humans, such as speech etc.) must be physical in origin. However, some assert that this position leads those who hold it into difficulties. For example, it has been argued that if it is the case that physical processes are the cause of all our behaviour, then our mental, qualitative, experience of the world must now be nothing more than an inert phenomenon aligned with our actions. For, so it is argued, if it is the electrochemical actions of the brain that alone have potency, our propositional attitudes and belief/desire states can be removed and 'intelligent' action would still continue. That is to say,

---

[2] It should be noted, that this elimintivism of the Churchlands derives from their anti-functionalism as well as an anti-realism concerning propositional representation and is also a paradigmatic target of Bennett and Hacker's *mereological fallacy* (see below).
[3] However, that thoughts have no form is a crucial principle to this argument, a position the likes of Fodor and Johnson-Laird would dispute.

physicalism is said to lead to epiphenomenalism as our actions are now overdetermined and our mental (belief/desire) states are nothing but 'causal danglers' (Papineau 2004).[4] This is a type of argument that has also been discussed by Fodor (1993).

The hypothesis that the brain and neuron firings are the seat of cognition: There seems to be an implicit assumption operating behind the RTM that the seat of the mind is the brain; that propositional attitudes, such as beliefs and desires are instantiated by the neurons etc. of brain tissue alone. However, such a view has been criticised by the likes of Bennett and Hacker (2003) for its treatment of a part of cognition (the brain) as if it were the whole (a complete individual) and, as such, is said to have committed the *mereological fallacy*. This fallacy, whose name derives from the logic of the relations between parts and wholes, "consists in ascribing to a part of a creature attributes which logically can be ascribed only to a creature as a whole" (*Ibid.*, p29). For, we should not say that a *brain* has propositional attitudes, as cognitive scientists are apparently doing, but that a *human being* believes, desires, hopes etc., as:

> These are all attributes of human beings. Is it a new *discovery* that brains also engage in such human activities? Or is it a linguistic innovation, introduced by neuroscientists, psychologists and cognitive scientists, extending the ordinary use of these psychological expressions for good theoretical reasons? Or, more ominously, is it a conceptual confusion? Might it be the case that there is simply *no such thing* as the brain's thinking or knowing, seeing or hearing, believing or guessing, possessing and using information, constructing hypotheses, etc. – that is, that these forms of words make no sense? (*Ibid.*, pp70-71)

That is, that cognitive scientists' supposed view of the mind as being isolated in only one part of a human being (their brain) is a logical confusion that has set them off on the wrong path and has only led them to postulate confused answers to a confused question (just as Descartes' isolation of mind from body has done (see Damasio 2005)).[5, 6]

---

[4] Nevertheless, it needs to be made clear that cognitive scientists, following Marr's tri-level hypotheses, do not posit physical actions and/or propositional attitudes as the cause of behaviour, as though the two phenomena were distinct entities as this argument supposes. Rather, cognitive scientists see propositional attitudes, mental representations and the physical actions of the brain/body as the same thing: They hold a conceptual ternary but an ontological monism, i.e. the three phenomena are seen as being *identical*. Therefore, to say that mental states are epiphenomena of physical action, as this argument does, is senseless, for mental states *are* physical actions. However, it should be noted that if issue is conjointly taken with Marr's tri-level hypothesis along with this epiphenomenalist argument, the defence presented here will not be available.

[5] However, to attribute this position to *all* of cognitive science and cognitive scientists may be mistaken. As we shall see (Chapter Five: Discussion), there is an awareness within some of the literature that an isolated brain, or 'bed-ridden' AI, is not capable of propositional attitudes and cognition proper, and that it is only within 'embodied' systems, within which a 'brain' is still central, that intelligence is capable (see, for example, McGinn 1989).

[6] For empirical evidence that supports these arguments, see Byrge *et al* 2014 and Pfeifer *et al* 2014.

The hypothesis that language is systematically analysable: Some, such as contemporary radical contextualist ordinary language philosophers (such as Avner Baz, Charles Travis etc.), say that language is 'radically contextualised' (or, alternatively, 'occasion sensitive') and therefore cannot be formally analysed within the monolithic format cognitive scientists wish. These philosophers argue that locutions only make sense once we have taken note of the context within which they are uttered. That is to say, there exists a view which:

> proceeds from an assumption of what speakers must know in knowing the meaning of a word, at least when it comes to 'singular substantives': namely, how to 'apply' the word to cases, how to refer to (denote, name, classify, pick out) cases or items by means of it, apart from any context of significant use[... Even] such philosophically innocent predicates as 'weighs 79 kilos' or 'is green'[...], contextualists have argued, are [not] fit for (either true or false) application to cases apart from contexts of significant use. (Baz 2012, p82)

For example, as Travis (1989) observes, the truth conditions for the proposition 'There is milk in the refrigerator', in reference to a mere puddle of milk at the bottom of said fridge, changes depending on the occasion of its locution: Whether it is said to someone cleaning the fridge or someone making a cup of coffee (where it will be true in the former case and false in the latter). Therefore, to embark on a systematic project of linguistic analysis that is not sensitive to context, as cognitive scientists propose, cannot succeed. In fact, so it is argued, an accurate and full analysis of language in all contexts is not possible, all that can be achieved is the analysis of propositions on a case-by-case basis.[7]

Functionalism: Some argue that the RTM's foundational position that any physical system that is capable of instantiating the functions of cognition is capable of producing mentality is mistaken. For example, some contest that functionalism is unable to account for qualia – the 'what-its-likeness' of consciousness – or answer the 'inverted qualia argument' found in the likes of Putnam (1981, pp79-81). In these types of argument, we are asked to imagine two individuals looking at a vase that they both believe to be blue, but that one individual's experience of blue is qualitatively identical with the other's experience of red, and *vice versa*. In such an instance the causal or functional role of the vase's colour is the same as if it were being experienced identically by both persons (it causes belief in the objects 'blueness'), but the intervening mental state, the phenomenal character, is not the same. Therefore, so it is

---

[7] For a similar argument, based on reasoning capacities rather than radical contextualism, see Putnam (1988, pp87-89).

argued, not all mental states (i.e. qualitative states) can be causal/functional in origin and, therefore, cannot be explained by functionalism.[8, 9]

Marr's tri-level hypothesis: It should be remembered that cognitive scientists, in respect to cognition, hold a conceptual triad but an ontological monism. That is, they view propositional attitudes, mental representations and physical brain states to be *identical*. However, this may lead those who hold the position into theoretical difficulties. For example, during Saul Kripke's discussion of metaphysical necessity (i.e. facts that are *true in all possible worlds*, e.g. 'Barak Obama is human') and contingency (i.e. facts that are *only possible in alternate worlds*, e.g. 'Barak Obama is President') in relation to identity statements, in his *Naming and necessity*, a number of issues are brought to light. For example, if we take a propositional attitude (i.e. the belief that it will rain), a mental representation (i.e. the propositional or analogical representation of the possibility that it will rain) and a biological state (i.e. the electrochemical state of the body that realises our representation and our belief) to be rigid designators (i.e., to refer to the same state of affairs in all possible worlds) then we seem to be lead into some form of neural, or biological, chauvinism. For, as Kripke (1972, pp133-150) argues, if we take Mental state M to be identical to biological state B, and both B and M are rigid designators, then $B=M$ must be a necessary truth, i.e. $B=M$ must be true in all possible worlds. Therefore, any world in which B exists, M must also exist. This is unproblematic for followers of the RTM and is something they may freely concede, however identity statements go both ways: If it is a necessary truth that $B=M$ then it is an equally necessary truth that $M=B$ and any world in which the mental state exists then the *electrochemical biological state must also exist and be correlated with it*. (N.B. the conceptual ternary of cognitive scientists is here being simplified to a dualism, this is merely for the sake of exposition and will have no effect on the argument when extended so as to accommodate Marr's intervening representational level, i.e., so as to capture the true position of Marr's tri-level hypothesis that *propositional attitude=mental representation=physical realiser*.) Kripke (*Ibid.*, p154) summarises this argument as follows:

> In sum, the correspondence between a brain[/biological] state and a mental state seems to have
> a certain obvious element of contingency. We have seen that identity is not a relation which can

[8] Interestingly, as Putnam (1981, P81) observes, this form of argument is only successful "if such cases [as inverted qualia] are really possible". However, Wittgenstein, within *Remarks on colour*, conducts an analysis of this very prospect, during paragraphs 84-88. During this investigation, it is observed that the qualitative nature of a colour is not determined independently, that a colour's qualitative aspect is instead determined in relation to all the other colours in the spectrum, i.e. in its position within the colour wheel. Therefore, to assert that some of a person's colour qualia can be inverted (e.g. from blue to red and *vice versa*) is nonsensical and such arguments as employed by Putnam fail before they begin. Nevertheless, Wittgenstein's analysis does not preclude the possibility of an inversion of the whole colour wheel, i.e. that Person$_1$'s qualitative perception of the world is the colour negative of Person$_2$'s. As such, inverse qualia arguments may be logically possible, but only if whole colour systems are included within the account.
[9] For alternative arguments against functionalism, from the point of view that denies the multiple realisibiliy of cognition or which presumes semantic holism, see Churchland (1981) and Putnam (1988) respectively.

hold contingently between objects. Therefore, if the identity thesis were correct, the element of contingency would not lie in the relation between the mental and physical states. (Kripke 1972, p154)

Therefore, it seems to hold mental states and physical states/realisers as being identical, as Marr's tri-level hypothesis seems to, is to hold the unintuitive and neurally, i.e., biologically, chauvinistic position that only brain states can cause mental states.[10] Furthermore, there are also objections to specific levels of the Marr's tri-level hypothesis. See, for example, the folk psychology, representational and physicalism sections in this list.

The above should demonstrate that cognitive science has been, and is, subject to a number of potentially devastating criticisms that attack the gamut of the hypotheses it puts forward. (However, the above should not be considered an exhaustive list of the hypotheses held by the RTM, let alone an exhaustive list of arguments against the RTM.[11]) Nevertheless, this investigation will identify that some of those criticisms that have been the most interesting and that have had the greatest impact on cognitive science (or, more accurately, the RTM), are those that derive from that approach's account of semantics. As such, we will be singling out those criticisms that charge the RTM, and, therefore, cognitive science, with an ineffectual handling of meaning and which, thereby, conclude that computation is not suitable for cognition.[12] Consequently, this chapter will be giving its full attention – or, alternately, will be limiting its attention – to cognitivist criticisms of cognitive science and the RTM that are of a semantic nature, to the detriment of those directed toward propositional attitudes, functionalism etc.[13] Accordingly, this discussion will be structured around:

i)   An individual examination of a number of semantic arguments against cognitive science's RTM and that can be found in the literature.

ii)   An investigation into the impact such arguments have had on cognitive science's theory of mind.

---

[10] However, there is an issue here. For, adherents of the RTM *eo ipso* hold a functionalist view of cognition. Therefore, such theorists do not contest that mental representations and propositional attitudes are identical to any *physical state*, or even any *physical realiser*, as portrayed in the above argument. Instead, they assert, following Putnam (1975, p291), that:

> Two systems are functionally isomorphic if *there is a correspondence between the states of one and the states of the other that preserves functional relations*. To start with computing machine examples, if the functional relations are just sequence relations, e.g. *state* A *is always followed by state* B, then, for F to be a functional isomorphism, it must be the case that state *A* is followed by state *B* in system 1 if and only if state *F*(*A*) is followed by state *F*(*B*) in system 2 [emphasis in original].

Consequently, those who subscribe to RTM adhere to the following identity statements: *Functional relations=mental states* and *mental states=functional relations*. And, as functional relations can be realised by a plethora of physical systems, the charges of neural chauvinism are avoided. However, as this defence is based on the functionalist view of cognition, if that stance is simultaneously attacked, this position cannot be taken.

[11] For an alternative list of this kind, see McGinn (1989, pp120-132).

[12] As we shall see, this attribution of significance and impact to such arguments derive from both the recognition of their import by those working within the relevant fields (e.g. Johnson-Laird 1988, p99; Harnard 1991, p47; Preston 2002, pp1-2; Müller 2007, p102), as well as the affect these arguments have had on the account of cognition held by such theorists (e.g. Brooks 1990, p3; Harnard 1990b, p335; Johnson-Laird 1993, p34; Fodor 2010, p199; Rapaport 2011; Schweizer 2012, pp199-204).

[13] Moreover, to give every argument against cognitive science and/or the RTM its deserved attention, individually analysing each in turn, would lead this thesis into volumes.

iii) A full exposition of an alternate theory of meaning many theorists have been adopting in response to such arguments.

iv) A study into the literature for evidence of the scale of analysis of such a response.

v) A statement of the research aim of this thesis.[14]

§i Semantic arguments

Before beginning this investigation into the semantic arguments levelled against the RTM, it needs to be clarified that this section will be focused on the suite of such arguments that can be encapsulated by the philosopher John Searle's Chinese room argument.[15, 16] As such, this section shall follow the proceeding order:

1) An exposition and examination of Searle's Chinese room argument, its relation to the symbol grounding problem and its impact in relation to the RTM and those hypotheses within cognitive science based upon it.

2) An investigation into the homunculus fallacy and the potential impact that it has on the RTM and related philosophies of mind.

3) An analysis of the symbolic fallacy and its possible impact on the RTM and those theories within cognitive science that rely upon it.

4) An exposition of the RTM's and cognitive science's issues concerning semantic holism.

The Chinese room argument

The most well-known argument against the concept that computation is a successful explanation of meaning, and therefore cognition, is what is known as the Chinese room argument (hereinafter, the CRA). This prominent thought (*gedenken-*)experiment has been highly influential since its first publication by Searle (1980, pp417-418) and follows the – now familiar – ensuing narrative, a narrative which deserves being quoted at length:

> Suppose that I'm locked in a room and given a large batch of Chinese writing. Suppose furthermore (as is indeed the case) that I know no Chinese, either written or spoken, and that I'm not even confident that I could recognize Chinese writing as Chinese writing distinct from, say, Japanese writing or meaningless squiggles. To me, Chinese writing is just so many meaningless squiggles. Now suppose further that after this first batch of Chinese writing I am given a second batch of Chinese script together with a set of rules for correlating the second batch with the first batch. The rules are in English, and I understand these rules

---

[14] It needs to be highlighted, however, that although the following investigation will be focused on the semantic arguments made against much of cognitive science and the RTM, this thesis will not be addressing their success or otherwise. Its attention will merely be focused upon the impact said arguments have had on cognitive scientists' account of cognition and on the success of the updates they have inspired.

[15] Although Searle's Chinese room argument in "Minds, brains, and programs" is explicitly targeted toward strong AI, the RTM's reliance on computation for cognition means that the argument is equally applicable to that project as well as that of those involved in strong AI.

[16] For an example of other semantic criticisms not encapsulated by Searle's Chinese room argument, and, therefore, not included in these pages, see the discussion surrounding misrepresentation and the 'disjunction problem' in Dretske (1986), Fodor (1990) and Vosgerau (2010).

as well as any other native speaker of English. They enable me to correlate one set of formal symbols with another set of formal symbols, and all that "formal" means here is that I can identify the symbols entirely by their shapes. Now suppose also that I am given a third batch of Chinese symbols together with some instructions, again in English, that enable me to correlate elements of this third batch with the first two batches, and these rules instruct me how to give back certain Chinese symbols with certain sorts of shapes in response to certain sorts of shapes given me in the third batch. Unknown to me, the people who are giving me all of these symbols call the first batch "a script" they call the second batch a "story" and they call the third batch "questions". Furthermore, they call the symbols I give them back in response to the third batch "answers to the questions" and the set of rules in English that they gave me, they call "the program". Now just to complicate the story a little, imagine that these people also give me stories in English, which I understand, and they then ask me questions in English about these stories, and I give them back answers in English. Suppose also that after a while I get so good at following the instructions for manipulating the Chinese symbols and the programmers get so good at writing the programs that from the external point of view – that is, from the point of view of somebody outside the room in which I am locked – my answers to the questions are absolutely indistinguishable from those of native Chinese speakers. Nobody just looking at my answers can tell that I don't speak a word of Chinese. Let us also suppose that my answers to the English questions are, as they no doubt would be, indistinguishable from those of other native English speakers, for the simple reason that I am a native English speaker. From the external point of view - from the point of view of someone reading my "answers" – the answers to the Chinese questions and the English questions are equally good. But in the Chinese case, unlike the English case, I produce the answers by manipulating uninterpreted formal symbols. As far as the Chinese is concerned, I simply behave like a computer; I perform computational operations on formally specified elements. For the purposes of the Chinese, I am simply an instantiation of the computer program.

Now the claims made by strong AI [and the RTM] are that the programmed computer understands the stories and that the program in some sense explains human understanding. But we are now in a position to examine these claims in light of our thought experiment:

1. As regards the first claim, it seems to me quite obvious in the example that I do not understand a word of the Chinese stories. I have inputs and outputs that are indistinguishable from those of the native Chinese speaker, and I can have any formal program you like, but I still understand nothing. For the same reasons, Schank's [story interpreting] computer understands nothing of any stories, whether in Chinese, English, or whatever, since in the Chinese case the computer is me, and in cases where the computer is not me, the computer has nothing more than I have in the case where I understand nothing.

2. As regards the second claim, that the program explains human understanding, we can see that the computer and its program do not provide sufficient conditions of understanding since the computer and the program are functioning, and there is no understanding.[17, 18]

To summarise: The CRA purports to demonstrate that were an individual to go through the computations of the mind prescribed by strong AI, the RTM, and much of cognitive science, then the semantic content of the symbols and mental representation being manipulated would still remain hidden to that individual. Ergo, it must be concluded that the semantic content of mental symbols are equally hidden during their purely syntactic manipulation by machine 'intelligence' under the strong AI paradigm, or human intelligence under the RTM paradigm. As Searle (*Ibid.*, p417) identifies, the issue here is that a sign's lack of intentionality (i.e. "that feature of certain mental states by which they are directed at or about objects and states of affairs in the world" (*Ibid.*, p424)) means a lack of meaning (see Fodor 1990, p28), a deficit that no amount of syntax can rectify. Or, in other words, that *syntax is not sufficient for semantics* (see Searle 1990, 31).

Since its first publication, however, the CRA has been subjected to a vast amount of criticism that has attempted to prove that the argument's conclusions are incorrect (e.g. Block 2002; Rey 2002; Rapaport 2007; Schwartz 2012). In fact, the first publication of "Minds, brains, and programs" in *The behavioural and brain sciences* was immediately followed by twenty-six responses, or commentaries, covering twenty-five pages, from some of the most prominent figures in cognitive science and AI, including Dennett (1980), Fodor (1980), Minsky (1980) and Pylyshyn (1984), all of which criticise some aspect of its structure and argument. Nevertheless, although many commentators have tried to invalidate the CRA, attempting to show that Searle-in-the-room's – or the room's – failure to understand Chinese is either incorrect or unimportant, many of these objections, as observed by Searle (1987, p301; 1990, p30), focus on the form of the argument rather than on its logic; i.e. commentators tend to assume that the narrative of the CRA exhausts its purpose.[19] This state of affairs has meant that much effort has been spent attempting to show that, as the form of the CRA is not

---

[17] Though this argument is here being attributed to Searle (1980), it is one that can be seen to derive from the 'Chinese Nation' thought experiment from Block (1978) (and may, arguably, be one prefigured by Wittgenstein's grocer in PI 1 (see Hutchinson 2007, p707-708)).

[18] Although the CRA, in the form presented here, is only applicable to traditional, linear, computational accounts of mind, Searle (1990, p28) reformulates this thought experience into the 'Chinese Gym' argument, to show that the CRA's central contention (that syntax is not sufficient for semantics) is equally applicable to parallel, connectionist accounts.

[19] Articles that critically tackle the logic that makes up the CRA can be found in Hauser (1997) and Haugeland (2002). Such theorists contend that the CRA fails against the RTM, for, so it is reasoned, that view of the mind does not state that syntax exhausts semantic; rather, it states that syntax *in addition to* execution, exhausts semantics (see *Ibid.*; Haugeland 2002). However, this is an argument based on the ontological assumption that a syntax and a program are distinct from their execution; that there exists a "*difference between inert instantiation[s] and dynamic instantiation[s]*" (Hauser 1997, p211). Nevertheless, it seems implausible to suggest that an inert 'syntax' and an inert 'program' are syntax or programs at all; an inert 'program' is merely code and an inert syntax is nothing. As such, the arguments of such commentators of the CRA seem to be based on fallacious assumptions and, as such, fail. (Unless, that is, they are able to disprove the intuitive claim that you cannot have syntax *in addition to* execution, i.e., that syntax *entails* execution).

accurate, neither are its conclusions. This has usually meant pointing out that in the thought experiment, as seen above, Searle is not an analogue of a computational machine (to which strong AI adherents ascribe understanding), rather he is merely the analogue of a computational machine's central processing unit (to which strong AI adherents ascribe no such understanding) (see *Ibid.* 1980, pp419-420). This is a response that has become known as the 'systems reply', as it is being claimed that it is the system (i.e. Searle-in-the-room, the room itself, the books, the operations, etc. all together) that understands Chinese, not Searle-in-the-room/the CPU. However, as Searle (*Ibid.*, pp420-422; 2002, pp53-54) avers, you can update his original, vanilla, thought experiment so that the system has been memorised by Searle-in-the-room, who can then go through the process in his head away from any closed room or other system. In such an instance, it is Searle-in-the-room, and only Searle-in-the-room, which comprises the whole system. However, even in such a case, although Searle-in-the-room is the system, he still understands no Chinese. Nevertheless, commentators have been no more convinced by the form of this updated argument than by the original CRA and continue to argue its imprecision (see Block 2002; Rapaport 2007). This back and forth, and the existence of a similar state of affairs concerning what has become known as the Robot reply (see Dennett 1980; Schweizer 2012), has led to the observation that the "consensus among cognitive scientists to the effect that the CRA is, and has been shown to be, bankrupt" (Preston 2002, p46).[20]

Nevertheless, the point of interest for this research is that our discussion so far has surrounded the inappropriateness of the *form*, or the logical consistency, of the CRA, not its central idea or message: That without intentionality, signs cannot become meaningful symbols (Searle 1980, p417). For, one crucial sense in which the CRA is still both important and relevant in relation to this thesis, despite its many failures, is not as a formal *argument* at all, but as a '*reminder*' that syntax alone is not sufficient for semantics (see Searle 1984, p33; 1989, p701). For, so it is being argued, the CRA should be considered as a piece of conceptual analysis that makes certain of our concepts and their operations more redolent, rather than as a rigorously structured argument in its own right. Under this guise, the CRA is seen as expressing, what Searle (1989, p703) terms, the "brutally simple" argument (hereinafter, the BSA) and its real aim is "merely to *illustrate* or remind us of the truth" (Preston 2002, p28) that syntax is not sufficient for semantics. This BSA is a more fundamental argument which Searle (1989, p703) formally states as follows:

1. Programs are syntactical.
2. Minds have semantics.
3. Syntax by itself is neither sufficient for nor constitutive of semantics.

---

[20] This is not an ascription of such a view to Preston himself.

Therefore,

4. Programs by themselves are not minds.

However, though it has been made clear that the CRA should be viewed as more illustration of the truth of the BSA than as argument, it is not yet clear *how* it is an illustration, i.e. how it legitimately reminds us that syntax does not exhaust semantics. To answer this, we shall argue that the CRA is an illustration that without an account of intentionality, explanations of cognition are led into a number of fallacies: 1) Stating that signs only become symbols to already conscious systems and 2) explaining the content of a symbol by reference to another, equally meaningful, symbol. These are two issues Searle explicitly presents within his account of the CRA, and which he articulates as follows:

1) "[F]ormal symbol manipulation by themselves don't have any intentionality; they are quite meaningless, they aren't even *symbol* manipulations, since the symbols don't symbolize anything". (*Ibid.* 1980, p422). As such, "[a]ny semantic interpretation has to be from outside the system of the symbol manipulation". (*Ibid.* 1989, p701).

2) "The computer, to repeat, has a syntax but no semantics. Thus, if you type into the computer "2 plus 2 equals?" it will type out "4." But it has no idea that "4" means 4 or that it means anything at all. And the point is not that it lacks some second-order information about the interpretation of its first-order symbols, but rather that its first-order symbols don't have any interpretations as far as the computer is concerned. All the computer has is more symbols". (*Ibid.* 1980, p423)

Therefore, the CRA reminds/shows us that without intentionality a cognitive system will be embarrassed as it will inevitably be led into certain destructive fallacies, including 1) the *homunculus* and 2) what Johnson-Laird (1988, p107) terms the *symbolic* fallacy. (Furthermore, and as we shall see, the transgression of these two fallacies implies the transgression of yet another: Semantic holism.) As such, Searle's formal definition of the BSA (see Searle 1989, p703) can be reformulated as follows, so as to remove any hidden ellipses and make its operation more perspicuous:

1. Programs are syntactical.

2. Minds have semantics.

3. Syntax is not sufficient for nor constitutive of intentionality (*Ibid.* 1980, p422).

4. Without intentionality, the meaning of a symbol can only be another, or be derived externally.

∴

5. Without intentionality there is no understanding, no semantics (*Ibid.*, pp422-423).

∴

6. Syntax is not sufficient for nor constitutive of semantics

∴

7. Programs by themselves are not minds.

Interestingly, and with the above in mind, it can now be seen that the CRA – and the BSA which it communicates – is an instantiation of what some in the field of artificial intelligence have termed 'the symbol grounding problem' (hereinafter, the SGP). This is a problem that still poses problems to those working within AI today (see White 2011, p179) and that Harnard (1990b, p335) defines as relating to the following questions:

> How can the semantic interpretation of a formal symbol system be made intrinsic to the system, rather than just parasitic on the meanings in our heads? How can the meanings of the meaningless symbol tokens, manipulated solely on the basis of their (arbitrary) shapes, be grounded in anything but other meaningless symbols?

Consequently, the CRA can be seen as not only an illustration of the BSA, but also, *eo ipso*, a reminder of the truth of the SGP.[21]

Now that this view of the CRA has been proposed (i.e. that it is an illustration of the BSA and the SGP and that any symbolic account of cognition that fails to provide for intentionality, or fails to address the symbol grounding problem, will be led into both the homunculus and symbolic fallacies and the difficulties they themselves entail), it would now be prudent to investigate the issues this thought experiment raises in their own right.[22]

Homunculus Fallacy

The homunculus fallacy (from the Latin *homo*, 'man, human being', and *–culus*, the diminutive suffix, and so, therefore, literally meaning 'the little man fallacy') is the error that an explanation of a

---

[21] However, it should be noted that not all those in the literature agree with this one-to-one relationship between the CRA and the SGP (c.f. Rodríguez 2012).

[22] The use of 'ground' and 'grounding' made throughout this discussion deserves definition. It should be clarified that 'ground' is a broader topic of study, of which the semantic grounding concerning this research is just a facet. This broader topic of study, following Fine (2012, p40) "stands to philosophy as cause stands to science". As such, ground, according to Correia and Schnieder (2012, p29) concerns "a particular sort of non-causal priority" and attempts to "ground all of the facts which appear to presuppose the reality of the mental or of value in terms of facts which do not presuppose their reality" (Fine 2012, p41). Consequently, the topic of ground not only includes semantics, as covered in these pages, but also covers "a list of priority claims which have been held in different areas of philosophy [e.g.]:
  1. Mental facts obtain because of neurophysiological facts.
  2. Legal fact are grounded in non-legal, e.g. social, facts.
  3. Normative facts are based on natural facts.
  4. meaning is due to non-semantic facts.
  5. Dispositional properties are possessed in virtue of categorical properties.
  6. what accounts for the existence of a whole is the existence and arrangement of its parts.
  7. A set of things is less fundamental than its members.
  8. What makes something beautiful are certain facts about the reception of its beholders.
  9. A substance is prior to its tropes or modes.
  10. That snow is white is true because snow is white. (Correia and Schnieder 2012, p1)
Although this broader topic of study is deserving of further attention, there is unfortunately no space within this study for such a provision. As such, any interested in the contemorary discussion surrounding ground should be directed to Correia and Schnieder's (eds.), 2012, *Metaphysical grounding: Understanding the structure of reality*.

cognitive process relies, ultimately, on the same cognitive process. This state of affairs is considered an error as it leads to the necessity of some other (or a 'little man' in the head of a cognitive agent) to perform the action in which we are interested. For example, if we explained the operation of vision with some sort of 'Cartesian theatre' (described by Dennett 1991, p39, as the belief "that somewhere, conveniently hidden in the obscure "center" of the mind/brain, there is a "Cartesian theatre", a place where "it all comes together""), wherein the external world is 'projected' as mental images within the brain, the following question arises 'But how are these internal images understood and processed?' Thus resulting in the logical conclusion that some homunculus is present in the brain's Cartesian theatre and is processing any projected mental images. The problem with committing such a fallacy is not only the intuitively incorrect positing of some conscious homunculi within our heads, but that it leads to a question begging infinite regress. For, if, continuing our example, vision is explained by an homunculus observing projections within a Cartesian theatre, what explains the vision of this 'little man'? Unless a new explanation of vision can be offered, it seems that he must also have an homunculus watching images in *his* Cartesian theatre. However, what then explains this second homunculus' vision? A third homunculus? And so on, in a chain of reasoning that continues *ad infinitum*. Furthermore, the parasitic nature of semantic content in homunculus scenarios produces one further issue for ungrounded cognitive systems. For, as Searle (2002, p65) observes:

> Human success and failures exist only relative to human interests. And indeed computer success and failures exist only relative to human interests because the machine does not have any psychologically real or observer-independent interests. In the case of Deep Blue, the machine did not know that it was playing chess, evaluating possible moves, or even winning and losing. It did not know any of these things, because it does not know anything. All psychological attributions made of it were observer-dependent. Indeed it did not even know that it was number-crunching or carrying out a program. In an observer-independent sense, the only things going on in the machine were very rapid state transitions in electronic circuits.

In other words, without intrinsic intentionality a system is unable to imbue its own actions with meaning. Therefore, ungrounded cognitive systems are unable to have goals or aims, let alone engage in goal-orientated behaviour. Any attribution of such purpose can only originate from outside the system or from an homunculus.[23]

In regard to cognitivist systems, cognitive science and the RTM, this argument is a potential threat, for without intentionality "[a]ny semantic interpretation [of a mental representation] has to be from

---

[23] This argument is one labelled by Harnard (1990a) as the 'hermeneutic hall of mirrors', which he articulates as follows:
"One need only recall that a Turing Machine, be it ever so "equivalent" to a furnace or airplane, cannot heat or fly: It merely manipulates symbols in a way that is formally equivalent to, and hence systematically interpretable as, heating or flying. But in the hermeneutic hall of mirrors created by projecting our own interpretations onto symbol manipulation, this distinction vanishes".

outside the system of the symbol manipulation" (*Ibid.* 1989, p701). Or to state this argument alternatively, within physical symbol systems:

> The symbols and the symbol manipulation, being all based on shape rather than meaning, are systematically interpretable as having meaning -- that, after all, is what it is to be a symbol system, according to our definition. But the interpretation will not be intrinsic to the symbol system itself: It will be parasitic on the fact that the symbols have meaning for us, in exactly the same way that the meanings of the symbols in a book are not intrinsic, but derive from the meanings in our heads. Hence, if the meanings of symbols in a symbol system are extrinsic, rather than intrinsic like the meanings in our heads, then they are not a viable model for the meanings in our heads[.] (Harnard 1990b, p338)

Therefore, with an absence of intentionality, as the CRA diagnoses RTM, cognitive systems necessitate the involvement of, what Peirce (§§274-284) terms, an 'interpretant' (i.e. an external conscious entity that imbues a sign with meaning) in order to convert a meaningless sign, or any other representations it employs, into a meaningful symbol. Consequently, without intentionality, for the RTM to have a mental symbol *qua* symbol necessitates an external Interpretant to imbue a mental sign with meaning, i.e. it necessitates an homunculus (see also McGinn 1989, p200).

Now that the homunculus fallacy and its relation to the RTM has been investigated, we now progress to reviewing the symbolic fallacy.

The symbolic Fallacy

One of the first clear articulations of what this research, following Johnson-Laird (1988, p107), is calling the symbolic fallacy, can be seen in the discussion on the validity of 'semantic markers' (wherein a word's meaning is exhausted by the semantic markers ascribed to it within a mental 'dictionary' and which we earlier saw in our introduction to Fodor's LoT (see Katz and Fodor 1963; Katz and Postal 1964)) by Lewis (1970 pp18-19), who argues:

> Semantic markers are *symbols*: items in the vocabulary of an artificial language we may call *Semantic Markerese*. Semantic interpretation by means of them amounts merely to a translation algorithm from the object language to the auxiliary language Markerese. But we can know the Markerese translation of an English sentence without knowing the first thing about the meaning of the English sentence: namely, the conditions under which it would be true. Semantics with no treatment of truth conditions is not semantics. Translation into Markerese is at best a substitute for real semantics, relying either on our tacit competence (at some future date) as speakers of Markerese or on our ability to do real semantics at least for the one language Markerese). Translation into Latin might serve as well, except insofar as the designers of Markerese may choose to build into it useful features – freedom from ambiguity, grammar based on symbolic logic – that might make it easier to do real semantics for Markerese than for Latin.[24]

---

[24] An alternative articulation of this symbolic fallacy, in this instance in relation to the SGP, can be found in Harnard (1990b, p339):

That is to say, an account of semantics commits the symbolic fallacy if it merely explains one meaningful symbol/suite of symbols by replacing it/them with another (for example, explaining a Chinese sentence to a monolingual English speaker by translating it into Latin or even further Chinese). This state of affairs is problematic for any semantic theory for it begs the following question: What explains the meaning of the second symbol/suite of symbols? That is, it merely regresses the problem such theories are attempting to resolve. Moreover, as Searle (1980, p422) argues, further difficulties dog any hypotheses that commit the symbolic fallacy, for:

> formal symbol manipulation by themselves don't have any intentionality; they are quite meaningless, they aren't even *symbol* manipulations, since the symbols don't symbolize anything.

Consequently, any system that commits the symbolic fallacy must either:

1) Enter into an infinite regress of symbol translation.

2) Commit the homunculus fallacy (for whom a sign can symbolise).

3) Completely fail as an account of meaning.

As such, any theory of cognition that transgresses the symbolic fallacy is, if not disproved, on extremely dubious logical ground.

In regard to the syntactic account of the RTM used within cognitive science, it is being argued that such a system commits the symbolic fallacy as its account of the semantic content of natural language is based solely in its translation into a mental symbol (i.e. a mental representation, be it either propositional or analogical), and, further, that the meaning of these mental symbols/representations are explained by their replacement, via computations, with other mental symbols. Therefore, so it is argued, the RTM and the fields, such as cognitive science, that employ the hypothesis commit the symbolic fallacy on two counts see (Searle 1980; 1984; 1990; 2002):

1) It accounts for the meaning of natural language via its translation into a suite of mental symbols.

2) It accounts for the meaning of these mental symbols via their translation into other mental symbols.

My own example of the symbol grounding problem has two versions, one difficult, and one, I think, impossible. The difficult version is: Suppose you had to learn Chinese as a second language and the only source of information you had was a Chinese/Chinese dictionary. The trip through the dictionary would amount to a merry-go-round, passing endlessly from one meaningless symbol or symbol-string (the definientes) to another (the definienda), never coming to a halt on what anything meant[…] The only reason cryptologists of ancient languages and secret codes seem to be able to successfully accomplish something very like this is that their efforts are grounded in a first language and in real world experience and knowledge. The second variant of the Dictionary-Go-Round, however, goes far beyond the conceivable resources of cryptology: Suppose you had to learn Chinese as a first language and the only source of information you had was a Chinese/Chinese dictionary! This is more like the actual task faced by a purely symbolic model of the mind: How can you ever get off the symbol/symbol merry-go-round? How is symbol meaning to be grounded in something other than just more meaningless symbols?

Summary

To summarise this discussion: The semantic critiques of the RTM, encapsulated by the CRA as BSA and SGP, contend that any system that is based on a physical symbol system, such as the mind under RTM and as found in a digital computer, will be unable to account for meaning if those symbols are not grounded, i.e. if such a system has no intentionality. This lack of grounding on the part of the system causes issues to arise, for without a final ground for semantic content to reside, no atomistic interpretation of symbols used by the system can take place, all that can be achieved in such instances is the transgression of the homunculus and symbolic fallacies (that is, an infinite regress of question begging explanations) and an holistic account of meaning that reduces the whole theory to absurdity. As such, these semantic arguments against the RTM (and, therefore, cognitive science and similar cognitivist projects) state that as such systems fail to ground their symbols and to successfully account for meaning, such devices, and the philosophy of mind that motivate them, will be led into an unrecoverable logical position. This is a conclusion akin to Putnam's analysis of functionalist theories of mind, especially in regard to his twin earth argument (see Putnam 1975, pp215ff.) – on which more later – from which it is deduced that:

> a more general reason why computational models of the brain/mind will not suffice for cognitive psychology[:] We cannot individuate concepts and beliefs without reference to the *environment*. (Putnam 1988, p73)

Semantic holism

However, the above discussion leads us to a final semantic argument that has so far only been implicit in our handling of the CRA and SGP: That of the problems of semantic holism. For, in light of the symbolic fallacy, it can now be inferred that any account of meaning that fails to ground its symbols must rely on the full suite of such symbols to determine the meaning of any one. For, if meaning is purely a product of the relations between, and operations on, symbols, then the meaning of any one symbol in isolation will not be determinable. As such, it is only in the context of the network in which a symbol resides that meaning can be inferred.

This state of affairs causes difficulties for a system, for without an atomic, individually defined form of semantic content meaning will be holistic and the semantic content of any one symbol in the system will be dependent on *all* the symbols and operations found within it. In such a system, if you make a change in the suite of symbols or the operations between them, then the meaning of *all* the symbols changes. However, as it must be self-evident that no two individuals possess the exact same suite of symbols and make the same inferences between them, then no two individuals have the same complete system. Consequently, any account of meaning that implies an holistic approach will be led

into the *reductio ad absurdum* that no two individuals imbue the same symbols with the exact same meaning and that, therefore, no true communication, as it is commonly perceived, is possible and has never occurred.[25] Or as Fodor (2000, p69), articulates this argument:

> If semantics is holistic, then the content of each of your thoughts depends on the content of each of your others; and, since no two people (indeed no two time slices of the same person) have all their thoughts in common, semantic holism implies that there are no shared intentional states.

§ii Response to semantic arguments

Now that these semantic arguments have been clarified, we are in a position to appreciate their significance, and explicate why they occupy a position of importance and interest within this thesis. This putative notability not only comes from the recognition they are given and the import attributed to them within the relevant literature (for example, Müller 2007, p102, who is concerned with the related field of AI, states that "The impression that AI has tacitly abandoned its original aims is strengthened by the widespread belief that there are arguments which have shown a fundamental flaw in all present AI, particularly that its symbols do not refer or represent, that they are not "grounded", as one now says" – see also Johnson-Laird 1988, p99; Harnard 1991, p47; Preston 2002, pp1-2), but also from the reaction many theorists working within the RTM paradigm have had to such semantic arguments (e.g. Harnard 1990, p335, who is again only concerned with the related field of AI, states that he is concerned with the SGB, concluding: "The problem is analogous to trying to learn Chinese from a Chinese/Chinese dictionary", and proposes that "A candidate solution is sketched [within the article]: Symbolic representation must be grounded bottom-up" – see also Brooks 1990, p3; Rapaport 2011, p7; Schweizer 2012, pp199-204). For detailed examples of this influence and reaction, we can look to the development of the two theories we examined in our prior chapter (i.e. Fodor's LoT and Johnson-Laird's mental models). For, a concern and response to the critiques summarised by the CRA, BSA and SGP can clearly be witnessed within the evolution of both theorists'

---

[25] However, compare this argument with those of Quine (1960) and Putnam (1988). Nevertheless, these authors' arguments, to avoid the above *reduction* – with which they agree (see Putnam 1981, p22) – state that the meaning of a term is based on its relation to "the other statements of the theory[… or a] corporate body of statements" (*ibid.*, pp8-9), i.e. not to all the theories or statements but to a clearly defined subset. Yet, this demarcation of groups of statements from one another seems arbitrary and baseless. This can be seen in one of Putnam's own examples:

> If someone tells you, for example, that *the thief entered through that window, and there is muddy ground outside the window*, you will "deduce" that *there are footprints in the mud*. But this is not a *logical consequence* of the facts stated, for you obviously have made use of an unstated auxiliary hypothesis to the effect that *if the thief entered through that window, he walked on the ground to get to that window*[…] What has experiential import is the corporate body of statements, and this import is not the simple sum of the experiential imports of the individual statements. (*Ibid.*)

But, if this is correct, this auxiliary hypothesis is based on at least one further 'unstated auxiliary hypothesis', that *the thief was subject to gravity*, which in turn is based on a large number of further hypotheses, including the behaviour of celestial bodies. That is, following this reasoning, statements seem to be related to hypotheses far removed from their 'own corporate body' and which are, following the logic of Quine and Putnam, themselves related to a large number of 'unstated auxiliary hypotheses', and which are, following the logic of Quine and Putnam, in turn themselves related to a large number of 'unstated auxiliary hypotheses', which are... As such, it seems that Putnam, and Quine, are unable to convincingly demark sections from the whole and, as such, must be led into the *reduction* argument drawn in the main text.

work, with each thinker's early syntactical accounts of cognition, which explained the semantic content of natural languages by mere translation into ungrounded mental symbols (via either semantic markers (e.g. Katz and Fodor 1963) or lexical decomposition (e.g. Johnson-Laird 1976; 1977; 1978a)), having today been abandoned in favour of a causal correspondence theory of meaning, in which the meaning of natural language propositions derive, ultimately, via reference to ontologically real entities (e.g. Fodor 2010; Johnson-Laird 2002; 2006).[26] As such, these theorists acquiesce to semantic arguments against functionalist accounts of cognition, such as are found in Putnam (1975, p227; 1981, p19; 1988, p73) that they fail to capture the fact that "'meanings' just ain't in the *head*". Moreover, this externalist move can be seen as, in part, a response to the homunculus and symbolic fallacies we have just encountered. That this conjectured motivation has some accuracy can be seen in the concerns of these writers when expounding their hypotheses, with Fodor (1990, p162) stating:

> denotation is presumably a word/world relation purely. There is thus some hope that an extensional semantics can avoid the holism that plagues use theories.

and subsequently averring:

> there's another, more parochial reason why it would be nice to have a causal account of the semantics of LOT. It's an argument that's been raised against the LOT picture very many times (indeed, against RTM as such) that Mentalese formulas themselves require an internal interpreter if they're to be meaningful. The immediate implication would be a homuncular regress, which is not a pretty thought[…] Patently, however, it is defanged if the content of a mental representation is determined not by the results of its interpretation but by its causal connections to things in the world. (Fodor 2010 p199)

While Johnson-Laird (1988, p107) has asserted that:

> one decisive objection to all the theories that I have described [i.e. meaning postulates] is that they say nothing about how words relate to the world. They contain no machinery that explains how my warning about [an unseen] table could guide your movements. Networks can tell you that two words are related, or that one sentence is a paraphrase of another, but they cannot tell you anything about the state of the world. They are as circular as dictionaries; they commit the "symbolic fallacy".

Adding, that:

> computers have no principles for relating the numerals they use to the external world[…] as to the meaning of their operations and results, the interpretation is left to the people who use them. Philosophers sometimes conclude from this fact that mental processes cannot be computational. This argument will be less persuasive, perhaps, when computers are provided with the missing link – with processes that relate their symbols to the world. (Johnson-Laird 1993, p34. See also, *Ibid.* 1988)

---

[26] See also Kripke (1972) and Putnam (1981).

Furthermore, this introduction of referentialism is a theoretical move in response to the concerns of the CRA, BSA and SGP that is not idiomatic to just these two thinkers, many other cognitive scientists and others working within the RTM framework have used such arguments as stimuli toward adopting a similar referentialist account of meaning (this is an observation that has also been made by Cole 2009, p439, who states that "Fodor, like most everyone else, has been influenced by the rise of externalism"). For example, Schweizer (2012, p201), while considering those arguments encapsulated by CRA and SGP, deduces that:

> if meanings ain't in the heads of individual human agents, then they're certainly not in the data bases of computational artefacts[...] On the foregoing widely accepted model of 'direct' reference, there is an essential causal and chronological link that semantically tethers an individual's linguistic behaviour to its environmental context.

And Rapaport (2011, p7) discussing his response to Searle's CRA, concludes that:

> Yes, our internal concepts (neuron firings, etc.) are, presumably, causally generated by objects and events in the external world, and our beliefs are true insofar as our internal ''semantic networks'' match the ''ontology'' of the external world.[27] (See also *Ibid.* 2007 and Ford 2011.)

In light of the above, it can be seen that semantic arguments hold a notable position for the RTM paradigm within cognitive science, as such criticisms highlight the necessity of intentionality and grounded symbols. In fact, this is an issue, as has been demonstrated, that has contributed to numerous theorists updating their syntactic accounts of meaning to the referentialist, correspondence based semantics that is prevalent within the field today. Nevertheless, although such a strategy may be common, it is not free from concern. For, the introduction of such a referentialist semantics into any hypothesis should not be considered an insubstantial update, even if a full blown correspondence theory of meaning is proven to sidestep the CRA, the BSA, the SGP and other related difficulties. This is because, such an account of meaning entails a large number of assumptions, any one of which may conflict with those already present in a theorist's existent philosophy of mind, and, therefore, demands considerable analysis before its logically consistent introduction can be confidently asserted. However, to see why this concern is genuine, it is apposite of this chapter to first continue with an analysis of the nature and history of these descriptions of semantic content themselves. As such, this thesis shall continue with an investigation into correspondence theories of meaning.

---

[27] Interestingly, the influence of the CRA/SGP that is being sketched here may have led to Fodor's observation that "in their deepest heart of hearts, the great majority of psychologists think that[…] they already know how to naturalise reference. Namely, it's got to turn out to be some sort of association[…] between ideas and the world" (Fodor 2010, ppp197-198).

§iii Correspondence theories of meaning

To put it succinctly, these accounts of language assert that meanings "aren't in the head" (Putnam 1988, p73), instead, semantic content is viewed as deriving from the world. That is, correspondence theorists assert that signs (mental or otherwise) gain their meaning through a connection to a particular state of affairs found in an ontologically real world 'out there' in some observer-independent reality. As such, symbols are viewed as deriving their semantic content by this connection and *mean* that part of the world to which they refer, denote or correspond. This is an account of meaning that has a rich history in the philosophical literature, arguably being expressed as early as Aristotle's *On Interpretation* (§1) in the 4th century BC, wherein it is expressed that:

> Spoken words are the symbols of mental experience and written words are the symbols of spoken words. Just as all men have not the same writing, so all men have not the same speech sounds, but the mental experiences, which these directly symbolize, are the same for all, as also are *those things of which our experiences are the images* [emphasis mine].[28]

This position on semantic content is one that has continued throughout philosophy's development since this time, with such notable thinkers as John Locke, Bertrand Russell, Ludwig Wittgenstein (under certain readings), Saul Kripke and Hilary Putnam adhering to a similar view at some point in their thought. Throughout this development, although substantial divergences may exist between individual instantiations of the theory, this hypothesis is characterised by its focus on an external semantics. As such, correspondence theories of meaning are innately antithetical to internalist theories of meaning that contend that semantic content is accounted for purely mentally, i.e. theories that state that we do not have to go outside the head to explain meaning. These theorists (e.g. McGilvery 1998, Chomsky 2000, Pietroski 2005) alternately argue that meanings are mental tokens that are derived and individuated syntactically. Or, as Pietroski (2005, p254) has it:

> knowing what sentence S means—that is, understanding S—is not a matter of somehow associating S with a function from contexts to truth-conditions.
>
> Rather, the meaning of S is a compositionally determined intrinsic property of S that constrains and guides without determining how S can be used to make true or false assertions in various conversational situations.

Importantly, however, an externalist semantics is also antithetical to the arguments such internalists employ to support their position. There exist a number of such arguments within the literature, however this thesis will consider only a limited number:

---

[28] Where 'those things of which our experiences are the images' is being taken to mean substance. For, as it is expressed in Aristotle's *On the soul* (2006b, p3):
> when we are able to give an account conformable to experience of all or most of the properties of a substance, we shall be in the most favourable position to say something worth saying about the essential nature of that subject[.]

1) **Negative facts:** Some (e.g. Pietroski 2005) argue that, although it is capable of explaining the positive facts of language (i.e., that the word 'table' means TABLE), externalism is incapable of explaining negative ones. For example, and as Belletti and Rizzi (2002, pp6-7) highlight, when we consider the simple word 'he', as used in:

      a) John said he was happy

and

      b) He said john was happy

it becomes apparent that there are complex grammatical structures involved that changes the meaning of the word 'he': In a) 'he' can potentially mean 'John', i.e. co-reference is possible, while in b) this is not so, i.e., co-reference is not possible and 'he' could never mean 'John'. What is of interest here, however, when looking at a) and b), is that externalism may be able to account for the meaning of 'he' (i.e. the positive fact of semantic content; the state of affairs to which 'he' refers), but it is not able to tell us why the grammar of a) allows for co-reference while the grammar of b) does not (i.e. the negative fact of the restrictions of syntax). Consequently, some internalists claim that externalist semantics are unable to account for all aspects of language and, therefore, fail.

2) **Tractability:** Secondly, internalists appeal to putative tractability issues for externalist computational accounts of cognitions to support their position. For, so internalists contest, externalists are guilty of "insist[ing] on relating a computational theory's domain to users of cognitive states and to things outside the head" (McGilvery 1998, p232). That is, externalists are said to make the brain responsible for computation over both mental states and *all* the entities in the world (for it is these objects that account for meaning). Further, this is said to demand too much, in fact it is said to demand a "theory of everything" (*Ibid.,* p237). Consequently, such an externalist system would not be constructible, i.e., could not be implemented by the brain or any other physical realiser, as it is to "extend the domain of a computational theory of a cognitive competence to things outside the head[… and] is to broaden the task of the theory to the point that it becomes unmanageable; it undermines the effort to construct an honest theory" (*Ibid.*, p234).

3) **Poverty of the stimulus for language learning:** Additionally, it is a truism that any theory of meaning must also account for how and why children are able to acquire and understand the meanings of natural language locutions as well as how they operate. However, internalists

contest that children are not exposed to a sufficient level of language use before they can accurately determine and use the meaning of natural language terms within all their possible, complex and nuanced, grammatical positions. For example, if we return to the complex syntactical structures at play in the use of 'he' between a) and b) above, there exists, so internalists argue, a level of nuance and difficulty within language that young children cannot grasp given the level of exposure they are able to have received, especially when we consider that this feat is achieved "at a time when they cannot even learn elementary arithmetic" (Chomsky 2002, p81). Consequently, certain theorists conclude that:

> A careful look at the interpretation of expressions reveals very quickly that from the earliest stages, the child knows vastly more than experience has provided. That is true even of simple words. (*Ibid.* 2000, p6)

That is, it is argued that acquiring language cannot be a process of establishing relations to states of affairs as children can learn and use meaningful language before they could possibly have had the opportunity to found such correlations, so meaningful language must, therefore, be an internal and innate capacity of humans (see also McGilvery 1998, p236).

Forms of reference fixation

Nevertheless, and ignoring the challenges from internalism for now, externalism faces a number of questions in its own right, the main one of which being: 'How is the reference between sign and signified reliably fixed?' For, if signs cannot be fixed across both individuals and time then no form of communication is impossible. Furthermore, if reference fixation is not successfully established, your thoughts now will not have the same meaning as your previous or future thoughts, making cognition, as we suppose it, to be impossible. As such, for a correspondence theory of meaning to succeed, it first needs a determinate and reliable method of reference fixation. Within the literature, and following Kirkham (1997, pp119-139), who is working in the related area of correspondence theories of truth (cf. Putnam 1975, p269), there are two broad strategies to answer this question: Correspondence by congruence and correspondence by correlation.[29]

Correspondence as congruence

The first broad strategy for reference fixation is that of congruence, an approach which argues that what relates a sign to a state of affairs is a structural relation between the two, i.e. that there exists an equivalence of structure between sign and signified causing the former to relate to the latter. Or, as Kirkham (1997, p119) avouches, this approach:

---

[29] It should be noted, however, that although this typology of correspondence by congruence and correspondence by correlation is useful, it does gloss over much detail found within the individual theories of correspondence following either broad strategy.

claim[s] that there is a structural isomorphism between truth [or semantic] bearers and the facts to which they correspond when the truth[/semantic] bearer is true[/meaningful]. Like the two halves of a torn piece of paper, the parts of the truth[/semantic] bearer fit with the parts of the fact.

As such, this strategy explains reliable and consistent reference through a sign possessing the same structure to a state of affairs. Examples of this approach include Wittgenstein's use of logical form within the Ineffable interpretation of the *Tractatus logico-philosophicus* (see Hacker 1997) and Johnson-Laird and Peirce's use of iconicity.

However, before we continue, it should be noted that there are a number of issues unique to holding such a view of reference fixation. For, as Daitz (1953) and Wittgenstein (again in his *Tractatus*) identify, such a form of reference is limited by, what are respectively termed, the referee's 'genre' or 'pictorial form'. A property of a sign respectively explained as follows:

> As icons signify by being like what they signify, their range of signification is limited. The special restriction of a reflection to what is in spatial proximity to it is unconnected with its iconicity, and due solely to the fact that it is a causal phenomenon. Both reflections and pictures, however, can show only spatial and coloured relations, i.e. they can show only what is in (what I earlier called) the same genre as themselves. A song or a taste can be neither reflected nor pictured. (Daitz 1953, p195)

And:

> **2.161** There must be something identical in a picture and what it depicts, to enable the one to be a picture of the other at all.
>
> **2.17** What a picture must have in common with reality, in order to be able to depict it —correctly or incorrectly — in the way that it does, is its pictorial form.[30]

That is to say, the genre or pictorial form of a sign determines which aspects of a state of affairs it can mirror and delimits to what it can refer. For example, a picture's genre/pictorial form is that of the visual (i.e. size, shape and physical relations). As such, as a picture can only show the visual attributes of a state of affairs and nothing else, it can only show us what a duck, for example, looks like, but not what it sounds or tastes like. However, given this restriction, this is a limitation that does not seem to apply to language and thought, as both systems are able to meaningfully represent a seemingly limitless variety of properties, e.g. they can both not only refer to what a duck looks like, but also what it tastes and sounds like. Therefore, those who hold a congruent or structural account of reference fixation within a correspondence theory of meaning are tasked with explaining this apparent lack of limitation to the genre/pictorial form of language and mental representation. That is, with identifying the special genre/pictorial form of language/mental representation that allows it to effectively denote

---

[30] Again, this is taking an Ineffable reading of Wittgenstein.

a state of affair's differing attributes and qualities (e.g. its visual, structural, audible, emotional etc. states) at the same time.

Finally, there is a further issue with such approaches in explaining why, if both sign and signified share the same structure, the former refers to the latter but not *vice versa*. That is to say, congruent correspondence theorists have to account for the one-way nature of reference within a strategy that seems to rely on the two-way phenomena of equivalence for correspondence. As such, congruent correspondence theorists need to explain this asymmetrical relation, i.e. why a sign refers to a state of affairs whilst a state of affairs does not refer to a sign.[31]

Correspondence as correlation

The other broad approach to reference fixation is to deny that correspondence is due to any innate property of the sign and the relevant fact, but is due to some form of societal, cultural or biological convention that correlates a sign to a state of affairs. There exists a vast number of these posited conventions to explain this correlation, however the difficulty with each approach is in ensuring the method is strict enough to account for the apparent consistency of a sign's meaning. Examples of posited conventions that are said to meet this difficulty include brute biological correlation (see Fodor 1975); the knowledge of experts (see Putnam 1988); initial baptism and 'causal chains' (see Kripke 1972); functional teleological/Darwinist explanations (see McGinn 1989); experience of facts *causing* the correlation, or the 'locking on', of representation to represented (see Fodor 1990; 2010); and Wittgensteinian enculturation/Learning (see Hutchinson 2008).

However, both the positions detailed above (i.e., correspondence by congruence and correlation) entail a raft of assumptions, both broad and specific, apart from those already clarified or alluded to (i.e. a mechanism of reference fixation, an antithesis to internalism and a realist ontology concerning states of affairs/facts (see McGinn 1989, pp10-14)). As such, for the purposes of this thesis (i.e., to demonstrate why the introduction of a correspondence theory of meaning into an hypothesis is not an insubstantial addition), it is necessary to enumerate and detail a number of the most pressing of these assumptions. To complete this, the following section will be broken down into a number of subheadings, each of which will analyse an assumption in turn. Consequently, this section will continue with a discussion of a raft of issues along the proceeding order:

1) Brentano's problem.
2) Epistemic scepticism of meaning.

---

[31] This asymmetrical relation between sign and signified has also been labelled, following Strawson's "On referring", as 'individuation-dependence' by McGinn (1989, p6).

3) Metaphysical issues.

4) Twin Earth cases.

5) Substitutivity/Frege cases.

6) The implementation problem.

7) Slingshot arguments.[32]

Brentano's problem

Further to the question of reliable and consistent reference fixation, another query arises if a correspondence theorist also holds a physicalist ontology (i.e. if they state that all there is, is the physical). If these two positions are simultaneously held, such a theorist is charged with explaining how this referential mechanism scientifically reduces to matter. That is, they are faced with answering 'Brentano's thesis' or 'Brentano's problem' (see Putnam 1988 pp1-3). This is a thesis derived from certain readings of the German psychologist and philosopher Franz Brentano's work on intentionality in *Psychology from an empirical standpoint*. Within this work, it is averred that:

> Every mental phenomenon includes something as object within itself, although they do not all do so in the same way. In presentation something is presented, in judgement something is affirmed or denied, in love loved, in hate hated, in desire desired and so on. This intentional in-existence is characteristic exclusively of mental phenomena. *No physical phenomenon exhibits anything like it*. We could, therefore, define mental phenomena by saying that they are those phenomena which contain an object intentionally within themselves [emphasis mine]. (Brentano 2009, pp88-89)

Some, such as Chisholm (1955), take this to be arguing that intentionality could not be a physical process, that no mere collection of matter is capable of reference or of being *about* another state of affairs on its own.[33] Instead, it is only the mind, or mentality, that is capable of creating such a phenomena as reference, just as an ant's track marks are only a picture of a state of affairs once a conscious observer recognises it as such (see Putnam 1981, p1). Consequently, if a physicalist stance is simultaneously occupied along with a correspondence based theory of meaning, as is done so within the forms of RTM investigated here, then this apparent irreducible nature of intentionality/correspondence needs to be rejected and Brentano's challenge met. (See also Putnam 1988, pp73-89.)

Scepticism

Next, there is an issue in that referentialism seems to conclude in an epistemic scepticism of meaning. For, as our terms are now Russellian and as we do not have idealised epistemic access to the world,

---

[32] Although this catalogue of issues faced by correspondence theories of meaning has attempted to be as exhaustive as possible, it is inevitable that projects of this nature fall short of completeness. As such, it should be clarified that this discussion makes no claims of definitiveness.

[33] It should be highlighted that the ascription to Brentano of this thesis is not universally accepted (cf. Moran 1996).

we must conclude that we cannot be certain about the meaning of any of our terms. That is, a clear issue is now raised: If the possibility exists that what we believe to be meaningful utterances are nothing more than Russellian singular terms which possess no referent (or which possess a different referent than we suppose), how can we differentiate between when our expressions are significant and when they are not (or when we are mistaken in their meaning)? In the literature, one response to this query is the introduction of an accurately informed third-party, an individual capable of correctly discerning between expressions which have a referent – and therefore semantic value – and those which do not (see McDowell 1977, pp172-175). However, there is a problem with this concept, third-parties are necessarily themselves first-parties, and as such *their* thoughts, including those about others expressions, are equally as fallible as those which they are adjudicating. Ergo, we are led to the question: What provides these-third parties with such a privileged position, such an Archimedean point, that enables *them* to accurately determine significant expressions from insignificant ones? The answer to this question cannot simply be their belief that they are making accurate ascriptions, as that is the very thing under suspicion. As such, there seems there can be no 'accurately informed third-party', no Archimedean point from which utterances can be determined to be meaningful. If this is the case, however, there now seems no way to know with any certainty whether anything we express is actually meaningful (or actually means what we suppose). That is, we can expand the uncertainty we have the meaning of a few of our Russellian terms, to *all* of our Russellian terms.[34] For example, it is now a serious possibility that, at some future date, it may be ascertained that the supposed referent of 'Queen Elizabeth II' may actually be nothing more than the result of some group hallucination or dream and no such person actually exists. Consequently, it is only at that point of discovery that we find that all our talk of the current Queen is, and has always been, meaningless. Therefore, some serious level of scepticism of meaning must ensue from the abandonment of narrow content and the adoption of pure referentialism, a level of scepticism that deserves investigation.

Within Philosophy of Language the two most prominent arguments for scepticism of meaning are found in the work of Kripke (1982) and Quine (1988). Although these authors derive their scepticism from different angles (the former from the problem of induction in rule following, the latter through the indeterminacy of translation), both are in agreement that no known fact can justify, or determine, our meaning one thing rather than another by any specific symbol. They then strengthen this purely epistemic claim, converting it into a metaphysical one, through the argument that even if idealised epistemic access to all the possible objective facts where given, there would still be no information that could determine our meaning one thing rather than another. Miller (2007, p145) identifies this

---

[34] This move is obviously founded on the validity of the Cartesian method: That if we know parts of a system are false, we must then doubt the whole.

form of scepticism as being what he terms constitutive scepticism. That is, "scepticism about the existence of a certain sort of fact", i.e., scepticism about meaning's very existence. However, this is not where our argument concerning correspondence theories has concluded. For, we did not, and cannot, make the move from the epistemic claim over the knowledge of meaning to a metaphysical one about meaning's existence. We concluded that there are no *known* facts that can provide certainty over a term's meaning, but still asserts that such facts do exist (i.e., the referents themselves). Under referentialism, if idealised epistemic access to the facts about the world were given, it would instantly become known which expressions contain semantic value and which do not. Consequently, correspondence theories of meaning stop short of constitutive scepticism, but still lead to, what Miller (*Ibid.*) terms, epistemological scepticism, "which concedes that the sort of fact in question exists, [but…] questions our right to claim knowledge of that sort of fact".[35] As such, to hold externalism is to allow for the possibility that individuals can be mistaken about their own thoughts, i.e., that they can be wrong about the thoughts they believe they possess. That is, as McDowell (1984, p288) has it, "a sort of *de re* sense may determine a *de re* sense[…] or it may determine nothing. And in the latter case[…] there can only be a gap – an absence". Or as Evans (1982, p44) articulates this point:

> [T]here does not seem to me to be anything incoherent in the idea that it may be, for a subject, exactly as though he were thinking about a physical object (say) which he can see, and yet that, precisely because there is no physical object he is seeing, he may fail to have a thought of the kind he supposes himself to have. It is not part of this proposal that his mind is wholly vacant[…] The claim is simply that there is a kind of thought we sometimes have[…] when, in virtue of the absence of any appropriate object, there is no such thought to be had[…] it does not appear to me that [this] proposal can be faulted on grounds of coherence alone.[36]

To summarise the preceding: A *de re* conception of content requires that, for many expressions to be meaningful, the object to which they refer exist, and if those objects are not veridical, then any atomic sentence in which those expressions are a constituent expresses no thought, and is without meaning. (Alternately, if the object to which a term refers is different to that which is supposed, then any proposition that contains the relevant expression will express a different thought than that which is assumed.) The speaker, mistakenly believing it to be otherwise, is simply uttering a false second-order belief about the supposed thought. It is only an accurately informed third-party, therefore, who, knowing the true conditions of the world, can differentiate between the meaningful and the

---

[35] In this regard, this argument can be seen as being somewhat akin to Putnam's permutation argument, which does not say that the expression "the cat is on the mat" has no meaning at the metalinguistic level, just that we have no knowledge of what it means (see Putnam 1981).

[36] See also McGinn (1989, p28).

meaningless expressions. But, as I have argued above, there can be no accurately informed third-party in reality; any ascription one purports to make is equally under suspicion of merely being a false second-order belief as the utterances they judge. Therefore, as there are no known facts which can determine whether our expressions are meaningful or not, accepting a *de re* conception of content requires accepting an epistemological scepticism of meaning.

Metaphysical assumptions

Thirdly, correspondence, *eo ipso*, necessitates certain metaphysical assumptions. Such as:

1) **Eternalism:** If it is both assumed that future and past tensed propositions are meaningful and that propositions only contain semantic content if they refer to concrete states of affairs found in the world, then it must be concluded that even historical and future events, and not just present ones, have some form of metaphysically concrete existence. As such, for tensed propositions to maintain their semantic content as supposed, correspondence theories of meaning entails an *eternalist* view of time (i.e. that the past and the future have an existence equal to the present) and is inherently opposed to both *presentism* (i.e. that only the present has any form of concrete existence) and, because of future tenses, the *growing universe theory* (i.e. that only the present and the past have concrete existence). Furthermore, this position, *eo ipso*, leads to determinism, as future tensed propositions such as "Tomorrow I will question the control I have over my own actions" now entails a predetermined future state of affairs to which to correspond, true or false to be meaningful. Therefore, this seems to imply that to adopt a correspondence theory of meaning and the view that future tensed propositions as meaningful, is to take the view that our actions are predetermined and we have no freedom of the will. Nevertheless, it may be remarked, as is done by Aristotle in *On Interpretation* (§9), that:

   > Everything must either be or not be, whether in the present or in the future, but it is not always possible to distinguish and state determinately which of these alternatives must necessarily come about[...] One of the two propositions in such instances [the original and its negation] must be true and the other false, but we cannot say determinately that this or that is false, but must leave the alternative undecided.

   That is to say, that although a proposition now requires a future to denote, it is not necessary that the truth or falsehood of that future be known. However, this does not solve the problem, for although it is not necessary for the truth or falsehood to be known for a future-tensed proposition to be meaningful, it is still necessary for the truth or falsehood of that future to exist and, consequently, that that future still be set, i.e. externalist meaning necessitates that the future be metaphysically certain even if it does not require that it be epistemologically so.

2) **Realism of properties and relations:** Moreover, for a correspondence theory of meaning to accept that propositions such as "The yellow ball is bigger than the blue cube" are meaningful, properties (e.g. "yellow") and relations (e.g. "bigger than") need to be external to interlocutors (i.e., need to denote a fact or state of affairs in the world) and therefore have some form of ontological existence. Consequently, to hold a correspondence theory of meaning also entails an expansion of a theorist's ontology, so as to accommodate relations and properties (see McGinn 1989, pp40-41).[37, 38]

3) **Counterfactuals and false propositions:** Finally, there is a question of how Russellian singular terms and thoughts can still refer when they are *false* and if, *eo ipso*, they do not correspond to any concrete state of affairs? Such false propositions and counterfactuals are a question we have already encountered and detailed in a previous chapter (see the discussion of Peircean negation), however it should be highlighted that such a question is here being taken to also concern related issues with signs for such things as negation (e.g. "I am not sat at a desk"), disjunctives (e.g. "either I am sat at a desk or I am not"), modal statements (e.g. "I could be sat at another desk") and empty names (e.g. "Sherlock Homes was sat at a desk"). These myriad difficulties are all here being included under the umbrella of 'counterfactuals and false propositions' as all seem to centre on the issue of reference to a states of affairs, or a fact, that does not pertain. Therefore, so it appears at first blush, if a solution is presented for counterfactual and false propositions (e.g., if a concrete or abstract form of referent is identified) then a large portion of the difficulties raised by these related issues will also be abated, if not removed. However, it should be made apparent that this conflation of forms of statement is a simplification of the individual arguments involved and glosses over much nuance found within and between them.

---

[37] This is making use of the distinction between internal and external properties and relations, where:
  Internal relations/properties: What provides an object's properties or relates object a to object b, in any way, is a property, or are properties, held by the objects themselves, i.e. any relationship/property is ultimately reducible to the concerned objects themselves. That is, what it is that relates one thing to another etc. is a fundamental, internal property (or properties) present within the objects related (e.g. a is a lighter shade of blue than b). Therefore, as a relationship/property is essential to relata/objects, and following the identity of indiscernables, if a relationship/property no longer existed, the objects involved would no longer be the same.
  External relations:/properties: Any form of relation/property is not caused through the internal properties of the objects involved, but rather is imposed upon them from an external position (e.g. c is to the left of d according to p, and only because of p). Therefore, if a relationship/property where to change, the item(s) involved would be unaltered and remain exactly the same; they would maintain their identity through the change.
[38] However, not all correspondence theories may take this position. For example, see certain metaphysical readings of Wittgenstein's *Tractaus Logico-Philosophicus* (e.g., Anscombe 2001), which replace a realism concerning relations and properties with a commitment to a saying/showing distinction.

Twin Earth Cases

Those who hold a correspondence theory and state that meaning is external are, *eo ipso*, led to the conclusion that if two terms, even those that share the same sign, denote different referents, then they must have a different meaning. That this is the case can be seen in an argument from Putnam (1975, pp222ff.), who gives the following thought experiment: Imagine that along with our world there is another, twin, Earth that is identical in every way with our own, except for one difference, water is not composed from H_2O but a unique chemical structure that we will, for the purposes of exposition, label as XYZ. Now, given this *gedankenexperiment*, it seems intuitive that, in 2014 and with a common knowledge of the hidden structure of the two Earths' respective waters (H_2O and XYZ), that our (as a resident of Earth) use of 'water' does not refer to, does not mean, that substance on Twin Earth that has the same surface structure and is locally called 'water' but which is composed from XYZ.[39] Given this, we can conclude that:

> there is no problem about the extension of the term 'water'. The word simply has two different meanings (as we say): in the sense in which it is used on Twin Earth, the sense of water$_{TE}$, what *we* call 'water' simply isn't water; while in the sense in which it is used on Earth, the sense of water$_{TE}$, what the Twin Earthians call 'water' simply isn't water. (*Ibid.* p224)

As such, if it is taken that meaning is determined by referent, that meaning 'ain't in the head', it must be concluded that if two terms have different referents, they possess different meanings, even if the signs for the terms are the same.

Although this all seems intuitive and is a fact to which many would freely assent, there are day-to-day cases where such an analysis seemingly leads us to unintuitive consequences. For example, my knowledge of trees is very limited, I could not tell a beech from an elm; I possess only the most superficial knowledge concerning the structure of beeches and elms, which leads me to consistently confuse the one for the other. However, even though I do not possess the conscious knowledge to single out a beech from an elm, it seems, or so some argue, intuitive to say that my propositions concerning beech trees are not synonymous with those concerning elms. Furthermore, we would not like to say that only an arborist's, or any other individual otherwise expert in trees', use of 'beech' has a different meaning from their use of 'elm'. Or, as Putnam (*Ibid.*, p226) avers:

> Suppose you are like me and cannot tell an elm from a beech tree. We still say that an extension of 'elm' in my idiolect is the same as the extension of 'elm' in anyone else's, viz., the set of all elm trees, and that the set of all beech trees is the extension of 'beech' in *both* of our idiolects. Thus 'elm' in my idiolect has

---

[39] However, the reliance on intuitions here may be disputed by some. For example, Devitt (2014, pp270-271) rejects any use of intuitions as evidence and instead recommends we should instead gather empirical evidence concerning actual language use. That is, rather than merely assert intuition as evidence, we should engage in some experimental philosophy if our argument is to remain valid (see *Ibid.*, p288). However, compare this position on the value of intuitions with that of Gendler (2010) or Papineau (2014, pp186-189).

a different extension from 'beech' in your idiolect (as it should). Is it really credible that this difference in extension is brought about by some difference in our *concepts*? My *concept* of an elm tree is exactly the same as my concept of a beech tree (I blush to confess). (See also *Ibid.* 1988, pp24ff.)

However, it seems that certain correspondence theories go against these intuitions and seem to lead to "meaning incommensurability" (Papineau 2014, p171). For example, where congruence is used for reference fixation, if the logical form of our mental representation 'beech' is identical to the logical form of our representation for 'elm' then the two terms are identical in our idiolect, i.e. the two representations *must* denote the same state of affairs (e.g. the set of all elm and beech trees, or even the set of all trees) and mean the same thing. Consequently, by holding that terms with different referents have different meanings we are led to the conclusion that within any discussion I undertake with an expert concerning beech and elms (or any other discussion between experts and the laity), we will not mean the same things by our terms and communication, as generally conceived, cannot be said to be taking place.[40]

Moreover, such an externalism as is produced from the above reasoning may lead such theorists as Johnson-Laird, Fodor etc. into still further difficulties. For Putnam (*Ibid.*) continues his Twin Earth argument quoted above as follows:

> Let us roll the time back to about 1750. At that time chemistry was not developed on either Earth or Twin Earth. The typical Earthian speaker of English did not know water consisted of hydrogen and oxygen, and the typical Twin Earthian speaker of English did not know 'water' consisted of *XYZ*. (*Ibid.* p224)

Consequently, in 1750 residents of both Earth and Twin Earth have only a knowledge of the, identical, surface structure of their respective 'water', not their hidden microstructure. This then raises the question: Does the use of 'water' by Earthians in 1750, and previously, refer to (i.e., mean) just Earth's H_2O, as in our use in 2014, or both H_2O and XYZ? Putnam (*Ibid.*) concludes that the correct answer to this question is firmly the former, that "the extension of the term 'water' was just as much H_2O on Earth in 1750 as in [2014]; and the extension of the term 'water' was as much *XYZ* on Twin Earth in 1750 as in [2014]" and, therefore, the terms' meaning haven't changed. To state otherwise, argues Putnam (*Ibid.* pp235-238), is to conclude that the meaning of our terms are affected by, and subject to, scientific discoveries and, therefore, that, for example, Archimedes' notion of 'gold' (or most other natural kind terms) is different to ours. However, this leads us to the conclusion that two people with the same mental state (or mental representation) may not mean the same thing by that mental

---

[40] Additionally, a similar case is found for ambiguous or vague propositions, for these are also everyday locutions to which we would intuitively, so it is argued, attach a meaning but whose referent may not be static between interlocutors. For example, and following the same line of reasoning as presented in the main text, the purposely vague and ambiguous statements of a politician, say, must have some meaning even though there is not one state of affairs they denote. As such, it seems that to rely on a solely *de re* concept of meaning also goes contra to our intuitive ascriptions of meaning to such terms. For further examples of this kind, see Kripke (1971, pp118-128).

state(/mental representation), which is obviously problematic for externalists such as Fodor and Johnson-Laird et al. For, such theorists state that if $person_1$ possesses the same mental representation (mental state) as $person_2$, their mental representations, *eo ipso*, share a referent. Consequently, if $person_1$ and $person_2$ have the same mental representation they *must* refer to the same state of affairs and mean the same thing. Furthermore, *mutatis mutandis*, this means that if Archimedes and Putnam have different mental representations, both of which they express using the sign 'gold', those representations must refer to a different state of affairs and mean different things. As such, to hold certain forms of the correspondence theory of meaning necessitates further holding the, to some, unintuitive notion that scientific discoveries into the hidden nature of referents change the meaning of those state of affairs' corresponding mental representations.

Substitutivity cases

As well as Twin Earth cases, where one symbol shares two referents, there are related, obverse cases, where two symbols share one referent, which also give light to certain implicit assumptions for an externalist semantics to operate consistently. The importance of such obverse cases was first highlighted by Gottlob Frege (1948, p209) during his analysis of meaning, wherein he affirms:

> $a=a$ and $a=b$ are obviously statements of differing cognitive value; $a=a$ holds a priori and, according to Kant, is to be labelled analytic, while statements of the form $a=b$ often contain very valuable extensions of our knowledge and cannot always be established a priori[…]If the sign 'a' is distinguished from the sign 'b' only as object (here, by means of its shape), not as sign (i.e. not by the manner in which it designates something), the cognitive value of $a=a$ becomes essentially equal to that of $a=b$, provided $a=b$ is true.

Meaning, that there are certain cases where two names can have differing cognitive meanings whilst, without our knowledge, denoting the same state of affairs. For example, Charles Dodgson and Lewis Carroll have the same referent (they are the same person), but to inform someone that (a) 'Charles Dodgson is Lewis Carroll' is not the same as telling them that (b) 'Lewis Carroll is Lewis Carroll', as (a) has the possibility of being *informative* (if you did not know that Lewis Carroll was the pen name of Charles Dodgson, for example), while (b) is merely a tedious tautology.[41] Nevertheless, If the meaning of a proposition were exhausted by its referents (as *de re* thinkers suppose) then both (a) and (b) should play an equivalent role within language and (b) would be as informative as (a). However, if they are not, as it in fact appears, then it seems reference cannot be enough for the provision of semantic content. Ergo, so it is argued, reference alone is not a sufficient explanation for the provision of meaning. This is a line of reasoning exemplified by the Hesperus/Phosphorus problem (wherein Hesperus refers to Venus in its mode of appearance in the evening and Phosphorus denotes Venus in

---

[41] Even though both 'Charles Dodgson' and 'Lewis Carroll' consist of two separate words, that does not limit the fact that they are the one name, point to the one referent and function as the one elementary component within a proposition. This is also true in many other, not so obvious, instances. For, "The designation of a single object can also consist of several words or other signs. For brevity, let every such designation be called a proper name" (Frege 1948, p210).

the morning) and are known as substitutivity cases for their stress on cases that break the principle of substitution (i.e. the intuitive notion that co-referring expressions can be substituted for one another without changing the meaning of the proposition in which the substitution is rendered). These substitutivity/Frege cases led Frege (*Ibid.*) to abandon a purely *de re* account of language and include sense (i.e. *Sinn* or 'the usual mode of presentation') into his philosophy. Such cases are equally as problematic for correspondence theorists as they were for Frege, for if they are correct, they demonstrate that our intuitive ascriptions of meaning and externalist theoretical accounts of semantic content are not aligned. Or, as Fodor (2010, p51) has it, "the moral [of substitutivity/Frege cases] is that reference can't be the same thing as content". Consequently, correspondence theorists must tacitly take the assumption that the informativeness of these sorts of cases are explainable within their hypotheses (as is done by Frege's inclusion of sense), or that the *apparent* informativeness of such Frege cases are both mistaken and explainable.

The implementation problem

However, Putnam's Twin Earth and Fregean substitutivity cases raise one further question other than those concerning the relationship between reference and meaning, the mutability of meaning and informativeness. It should be remembered, from our literature review, that a key theory underpinning RTM was that intelligence is a computational process, i.e., that the mind is a result of the ordered manipulation of symbols. These processes are defined by Fodor (1994, p8) as follows:

> Computational processes are ones defined over syntactically structured objects; viewed in extension, computations are mappings from symbols to symbols; viewed in intension, they are mappings from symbols under syntactic description to symbols under syntactic description.

However, the Twin and substitutivity cases highlight a conflict such a theory has with a referentialist account of meaning like correspondence. For, computation relies upon relations between symbols while correspondence relies upon relations between symbols and the world, raising the question 'How can internal relations (computation) produce external relations (reference)?' This conflict boils down to question of how a referentialist semantics, how a mind/world relationship, can be implemented in a syntactic theory and, as such, has been labelled as 'the implementation problem' (*Ibid.*, p24). However, it may be observed that this is an issue addressed by our earlier discussion of the need for a reliable mode of reference fixation. Nevertheless, the importance of this implementation problem within our discussion at this point is that, as Fodor (*Ibid.*, p24) observes, Twin and substitutivity cases may make any such account moot, for "Frege cases and Twin cases show that content *could* come unstuck from computation". Meaning, that even if there is a mechanism that reliably fixes reference and seemingly solves the implementation problem, such a mechanism will not be able to account for cases where the posited connection between mental representations and the world break down. That

is to say, Twin Earth and Frege cases "lead to predictive/explanatory failures of broad content psychological theories" (*Ibid.*, p39). For, as we have seen, such cases show that the internal computations posited by RTM (i.e. mental representations) and external referent do not seem to always be aligned as expected; Twin Earth and Frege cases seem to show that, contra to any reliable mechanism of reference fixation, one mental symbol may be aligned with two different parts of the world (Twin cases) or, conversely, that two separate mental symbols can be aligned with one state of affairs (substitutivity cases) and the content of such signs are not as predicted. Consequently, those who hold a referentialist semantics while simultaneously adhering to a computationalist account of mind (such as Fodor, Johnson-Laird et al), must have a response to the implementation problem along with the problems Twin Earth and Frege cases pose.

Slingshot arguments

Finally, along with the raft of commitments listed above, slingshot arguments should also be included in our catalogue of assumptions entailed by correspondence theories of meaning. These being "argument[s] (or more precisely, a family of arguments) that [are] designed to provide a formally strict proof of the claim that all true sentences designate (denote, refer to) one and the same thing, as well as all false sentences do" (Shramko and Wansing 2009, p430) and need to be included here for "slingshot arguments, originally are not tailored against the analysis of token-level causes and effects as facts, but against philosophical recourse to facts in general." (Baumgartner 2010, p112). That this form of argument should be analysed in these pages can be seen in the paradigmatic version of the slingshot argument found in Church (1943), which, like the other instantiations of the slingshot (e.g. Davidson 1969), moves step-by-step from one sentence through other propositions that have the same truth value, to a completely different proposition than the original, but which shares the same truth values. For example (an example taken from Church 1956), given the proposition:

1) Sir Walter Scott is the author of Waverley.

We can use the principle of substitutivity (i.e. that logically equivalent terms can be substituted without loss of meaning), to generate the synonymous proposition:

2) Sir Walter Scott is the man who wrote the twenty-nine Waverley novels altogether.

From which, using redistribution (i.e. that the clauses in certain propositions can be redistribution without loss of meaning), we get the synonymous proposition:

3) The number, such that Sir Walter Scott is the man that wrote that many Waverley novels altogether, is twenty-nine.

However, if we are concerned with truth values, this is logically equivalent to:

4) The number of counties in Utah is twenty-nine.

Which can be used to generate the logically, truth value, equivalent:

5) Twenty-nine is the number of counties in Utah.

From which we can produce:

6) Salt lake is a county in Utah

7) …

And so on and so forth. Leading to the conclusion that:

> Hence we have a means of showing any two true sentences to be synonymous. By a similar method any two false sentences can be shown to be synonymous. (Church 1943, p300)

That is, that all true/false propositions refer to the same state of affairs and the *reduction ad absurdum* that within a correspondence theory of meaning all true/false propositions have the same semantic content.

Summary

As such, this investigation has determined that introducing a correspondence theory of meaning into a philosophy is not an insubstantial addition, for such a theory necessitates accepting the following assumptions:

1) A commitment to externalist semantics and a rejection of internalist hypotheses and arguments

2) An ontological realism concerning states of affairs/facts.

3) That a reliable and consistent mode of reference fixation is possible. This is an assumption that itself entails a number of sub-assumptions:

   a. If a congruence or structural reference is held, it is also necessary to explain away the apparent limitations of genre or pictorial form where language/mental representations are concerned.

   b. Also, the one-way nature of reference needs to be explained within a congruent system, which appears to employ a two-way relation of structural equivalence.

   c. Finally, if the bearer of the (congruent or correlational) correspondence theory is also a physicalist, they also require that Brentano's question/problem can be answered and reference can be explained in a scientifically reducible, i.e., empirical, manner.

4) An epistemic scepticism of meaning.

5) A number of metaphysical entailments:

   a. An eternalist stance on the metaphysical status of time (leading to determinism).

   b. That counterfactuals can refer to something and therefore remain meaningful even while singular terms are Russellian. (N.B. This also includes those propositions that entail negation, disjunctives, empty names and modality.)

c. That certain properties and relations have some form of ontological reality.

6) That if two terms have different referents, they must have different meanings, no matter their surface level similarity and *vice versa*.

7) That meaning is malleable.

8) That the informativeness of Frege cases can be explained.

9) That the challenges posed to any solution of the implementation problem by Twin and Frege cases are avoidable.

10) That slingshot arguments can be answered.[42]

§iv Correspondence theories of meaning and cognitive science

Now that we have fully investigated correspondence theories of meaning, it should be apparent that their adherence necessitates a large number of substantial assumptions. This should clarify our earlier assertion that a concern with the adoption of such a theory exists and its acceptance should not be given before such a referentialist move's logical consistency has been essayed. As such, it would seem to be prudent for those involved within cognitive science and the RTM who have adopted this referentialist semantics to engage in a full critical analysis of their position and fully justify its adoption. Nevertheless, and although some critical analysis has been performed on certain instantiations of the representational paradigm (see Putnam 1988; McGinn 1989; Edwards 2014), there exists certain areas of the literature where this state of affairs is not currently found, with some relying only on those arguments we have seen (i.e., the CRA and SGP) to justify this substantial update to their theory of mind (see Brooks 1990; Harnard 1990b; Johnson-Laird 1980; 2002; 2013; Waskan 2006; Rapaport 2011; Schweizer 2012). That is to say, considering the significance and substance of a move toward correspondence, there currently exists a paucity of critical analysis concerning the success of its addition into the RTM within certain areas of the literature.

However, as the issue currently stands, this accusation is purely a groundless, unsupported assertion. Consequently, this thesis will progress by undertaking a small scale study to attempt to validate this claim.

---

[42] It should also be clarified that a correspondence theory of meaning also presupposes the legitimacy of the concept of ground. However, this legitimacy is a contemporaneous discussion and is beyond the scope of this thesis. Nevertheless, it should be established that referentialists, including those discussed in these pages, are committed to the validity of the concept and the "intelligibility, unity, the instantiation [and] the epistemic accessibility" (Correia and Schnieder 2012, p30) of their own theories is dependent upon that of ground's. For further insight into this contemporaneous debate, see also Kit (2012) and Audi (2012).

Study: The presence of critical analysis of correspondence theories of meaning and the RTM found within the literature

Aim of study

The purpose of this investigation will be to support our contention that there currently exists a dearth of critical analysis concerning the success, or otherwise, of the addition of a correspondence theory of meaning into the hypotheses of certain theorists currently operating within the RTM paradigm (N.B. 'Critical analysis' is herein being defined as 'serious engagement with the arguments of such referentialists and/or the implications of such arguments'). That is, this study will attempt to corroborate our assertion that certain theorists (e.g., Brooks 1990; Harnard 1990b; Johnson-Laird 1980; 2002; 2013; Waskan 2006; Rapaport 2011; Schweizer 2012 etc.) do not sufficiently analyse the implications the introduction of a correspondence theory of meaning entails and, further, that this deficiency is not rectified in the secondary literature.

Method of study

To discover the accuracy of this claim and to determine the level of critical analysis present in certain debates around the RTM, this study will identify all those works which cite the relevant pieces of literature and investigate the title and, where available, the abstract and individual references to determine the level of analysis within that paper concerning the point of interest. These citations will be identified via three different sources:

1) Web of Knowledge (http://wok.mimas.ac.uk/)

2) Scopus (http://www.scopus.com)

and, where necessary (i.e. where the first two, academic, sources fail),

3) Google scholar (http://scholar.google.co.uk/)

Multiple sources are being used so as to maximise the number of citations identified whilst simultaneously counteracting any bias present in any one source. Once the citations have been obtained, they will constitute the source data for a quantitative investigation into the presence of critical analysis found within them, with the number of works from the source data that engage in critical analysis being counted so as to quantify its presence. There will be considered a deficient number of publications analysing the theory if the literature is found to contain an unusually small or unusually univocal number of articles ('unusual' here being defined in relation to similar debates found elsewhere). As such, this study will endeavour to meet its aim and quantitatively establish the accuracy, or otherwise, of our earlier claim (i.e. that there currently exist a dearth of critical commentary concerning addition of a correspondence theory of meaning into the RTM).

Sample and generalizability

Of course, the full scope of the relevant literature (e.g. all the works of Johnson-Laird, Rapaport, Schweizer, Waskan etc.) cannot be canvassed in a small study of this nature. As such, this preliminary study (and, subsequently, this thesis as a whole) will limit its attention to just the hypotheses of Philip Johnson-Laird and no other. (This limitation to the works of Johnson-Laird, over any of the other referentialist representational theory of mind, has been selected as he is both a vocal (producing scores of publications arguing for its adoption) and long-term advocate of this move (arguing for its adoption as early as 1980).[43]) Again, not all of Johnson-Laird's works can be essayed, for many of his publications focus on matters beyond just his Peircean correspondence theory. Therefore, investigating the citation data for all his works would generate much surplus data that will bloat our study, consume much of the available resources and potentially obscure informative results. Consequently, so as to minimise surplus data while maximising the number of relevant citations found, only articles and monographs by Johnson-Laird that are specifically focused on his correspondence theory of meaning, or are central to its theoretical development, will be included within our sample. After reviewing Johnson-Lairds body of work, it was assessed that the publications that meet this criteria and that, therefore, will be included in this investigation, are:

1) Johnson-Laird, P., 1980, "Mental models in cognitive science" in *Cognitive science 4,* pp71-115
2) Johnson-Laird, P., 2002, "Peirce, logic diagrams, and the elementary operations of reasoning" in *Thinking and Reasoning*, Vol. 8, No. 1, pp69-95.
3) Barres P.E., Johnson-Laird P.N. 2003, "On imagining what is true (and what is false)" in *Thinking and Reasoning*, 9 (1), pp. 1-42.
4) Khemlani S., Orenes I., Johnson-Laird P.N., 2012, "Negation: A theory of its meaning, representation, and use" in *Journal of Cognitive Psychology*, 24 (5), pp. 541-559.

Therefore, and to summarise, this study will aim to identify the quantity of critical literature concerning Johnson-Laird's Peircean correspondence theory of meaning, by using the citations of the most relevant of his works as data to quantitatively determine the number of critical works currently published. As such, any findings this investigation produces will be generalisable to the work of Philip Johnson-Laird only and not to any other theorist employing similar strategies (e.g., Harnard, Fodor, McGinn, Rapaport, Schweizer, Waskan etc.).[44]

---

[43] Additionally, and as we will see, Johnson-Laird's particular position is one which has received a noticeable lack of critical engagement with its theory of meaning.

[44] For a critical discussion of referentialism within the language of thought, see Putnam (1988) and Edwards (2014). Furthermore, as Edwards (*Ibid.,* p363) construes the LoT hypothesis as "the claim that thoughts[…] have a compositional structure", his arguments would be equally applicable to analogical models of a compositional nature (e.g., Fodor 2010, pp173-175). However, this does preclude Johnson-Laird's mental icons. Furthermore, see McGinn (1989) for a defence of one form of isomorphic representations.

Findings

As of Wednesday 4th June 2014, Philip Johnson-Laird's four articles "Mental models in cognitive science", "Peirce, logic diagrams, and the elementary operations of reasoning", "On imagining what is true (and what is false)" and "Negation: A theory of its meaning, representation, and use" have been cited by a combined 265 unique publications (see Appendix, Table 2).[45] From these 265 titles, we were able to access 171 distinct abstracts (see Appendix, Table 3) and 56 citation quotes (see Appendix, Table 4). A study of this available information, indicated that, for one reason or another (i.e. their title, abstract or citation suggested that they may engage in some critical analysis of Johnson-Laird's Peircean correspondence theory of meaning), the following 17 publications were found to be of potential relevance to our research interests:

1) Baggio, G.; van Lambalgen, M.; Hagoort, P., 2012, "Language, linguistics and cognition" in R. Kempson, N. Asher and T. Fernando (eds.) *Handbook of the philosophy of science: Philosophy of linguistics*, Oxford: Elsevier, pp325-355.

2) Cadinu, M.R., Kiesner, J., 2000, "Children's development of a theory of mind" in *European Journal of Psychology of Education*, Vol. 15, No. 2, pp93-111.

3) Du, P.; Liu, D.; Zhang, L.; Hitchma, G.; Lin, C., 2014, "The processing of contradictory and non-contradictory negative sentences" in *Journal of Cognitive Psychology*, Vol.26, No.4, pp461-472.

4) Elqayam, S., 2005, "Mental models, model-theoretic semantics, and the psychosomatic conception of truth" in *Philosophia Scientiae*, Vol. 9, No. 2, pp259-278.

5) Ford, M., 1995, "Two modes of mental representation and problem solution in syllogistic reasoning" in *Cognition*, Vol. 54, No. 1, pp1-71.

6) Goodwin, J.W.; Hein, U., 1982, "Artificial-intelligence and the study of language" in *Journal of Pragmatics*, vol. 6, no. 3-4, pp241-280.

7) Macbeth, G. et al, "Models for the negation of conjunctions and disjunctions" in Europe's journal of psychology, Vol.10, No.1, pp135-149.

8) Marmolejo-Ramos, F., 2007, "New advances in the scientific study of text comprehension" in *Universitas Psychologica*, Vol. 6, No. 2, pp331-343.

9) Orenes, I.; Beltran, D.; Santamaria, C., 2014, "How negation is understood: Evidence from the visual world paradigm" in *Journal of memory and language*, Vol.74, pp36-45.

10) Oversteegen, E; Schilperoord, J., 2014, "Can pictures say no or not? Negation and denial in the visual mode" in *Journal of pragmatics*, Vol.67, pp89-106.

---

[45] These 265 unique publications excludes those written by Johnson-Laird himself. This exclusion was made, as these publications by Johnson-Laird have already been canvassed during our literature review and were already found to be lacking; a finding that initially inspired this study.

11) Pietarinen, A.-V., 2010, "Challenges and opportunities for existential graphs" in B.M. Paavola P.-V. Pietarinen and H. Rydenfelt (eds.) *Ideas in action: Proceedings of the* Applying Peirce *conference*, Helsinki: Nordic pragmatism network, pp288-303.

12) Rapp, D.N.; Kurby, C.A., 2008, "The 'Ins' and 'Outs' of Learning: Internal Representations and External Visualizations" in J.K. Gilbert, M. Reiner and M. Nakhleh (eds.) *Visualization: theory and practice in science education,* Springer, pp29-52.

13) Schmidt, J.R.; Thompson, VA, 2008, ""At least one" problem with "some" formal reasoning paradigms" in *Memory & Cognition*, Vol. 31, No. 1, pp217-229.

14) Schroyens, W., 2010, "Mistaking the instance for the rule: A critical analysis of the truth-table evaluation paradigm" in *Quarterly Journal of Experimental Psychology*, Vol. 63, No. 2, pp246-259.

15) Sowa, J.F., 2011, "Cognitive architectures for conceptual structures" in S. Andrews, S. Polovina, R. Hill and B. Akhgar (eds.) *Proc. 19th International Conference on Conceptual Structures*, Heidelberg: Springer, pp. 35-49.

16) Verdonik, D., 2010, "Between understanding and misunderstanding" in *Journal of Pragmatics*, Vol. 42, No. 5, pp1364-1379.

17) Vosgerau, G., 2006, "The perceptual nature of mental models" in C. Held, M. Knauff and G. Vosgerau (eds.) *Mental Models & the Mind: Current developments in Cognitive Psychology, neuroscience, and philosophy of mind*, Oxford: Elsevier, pp255-276.

However, a closer scrutiny of the content of these 17 titles revealed that only four of these articles are pertinent to the question that concerns this study.[46] Furthermore, only one of these publications, Vosgerau's "The perceptual nature of mental models", engaged critically with the *arguments* that concerns this study. This relevance can be seen in the following quotes:

1) "If a mental model directly corresponds to the modeled situation, the relations in the model have to correspond directly to the modeled relations as well. This leads to the even more central requirement that the relations between elements of a mental model have to be natural as well[… For] example, it is quite clear that the relation `to the right of' is represented not by an arbitrary symbol or another "abstract notation" but in the natural way by itself[...] On the contrary, Johnson-Laird introduces several abstract notations. Indeed, the notations he applies vary across his writings. For example, he introduces a symbol for negation, which is clearly a highly sophisticated notation". (Vosgerau 2006, p258)

---

[46] The other titles were found to be irrelevant for a number of reasons. For example, the apparent relevance of certain publications was found to be misleading (e.g. Baggio et al 2012; Cadinu and Kiesner 2000; Elqayam 2005; Goodwin and Hein 1982; Marmolejo-Ramos 2000; Pietarinen 2010; Rapp and Kurby 2008; Schmidt and Thompson 2008; Schroyens 2010; Soea 2011; Verdonik 2010), while the critical analysis of others was found to concern hypotheses beyond the scope of this study (e.g., Ford 1995, p69; Oversteegen and Schilperoord 2014).

2) "Negation is a sophisticated logical notion, and hence every theory of reasoning that introduces the notion of negation "should offer some account of how such an apparatus is acquired" (Johnson-Laird 1983, 66). Johnson-Laird does not offer such an account and therefore does not meet his criteria for theories of reasoning". (*Ibid.* p270)

3) "Taken together, introducing symbols for negation into mental models contradicts both the constraint of structure preservation and the constraint of naturalness". (*Ibid.* p272)[47]

The final three pertinent articles (Du et al's "The processing of contradictory and non-contradictory negative sentences"; Macbeth et al's "Models for the negation of conjunctions and disjunctions"; and Orenes, Beltran and Santamaria's "How negation is understood: Evidence from the visual world paradigm") are very recent studies that engage with the *implications* of Johnson-Laird's arguments. All three papers present empirical evidence consistent with Johnson-Laird's hypotheses. This can be seen in the following quotes:

1) "Regarding representation, our results support the ideas that (1) the negative operator plays a role in the mental representation, and consequently a symbolic representation of negation is possible, and (2) it is not necessary to use a two-step process to represent and understand negation". (Orenes, Beltran and Santamaria 2014, p36)

2) "Experimental comparisons and a complimentary descriptive study yielded evidence consistent with the theory-driven predictions derived from the Mental Models Theory". (Macbeth et al 2014, p135)

3) "types of negation, rather than time delays, had a significant influence on the processing of negative sentences". (Du et al 2014, p461)[48]

Discussion

From this analysis, it has been found that only four articles, or 1.5% of the literature canvassed, engaged in the critical analysis of the success, or otherwise, of the addition of a Peircean correspondence theory of meaning into Johnson-Laird's theory of mind in any way. Furthermore, it must be highlighted that 1) three of these studies are empirical investigations whose findings are consistent with one aspect of Johnson-Laird's updated theory of mind and 2) all four articles did not

---

[47] Vosgerau's paper, in turn and as of 04/06/2014, only has nine citations, which are either authored by Vosgerau or not published in English.
[48] These papers, in turn and as of 04/06/2014, have not yet been referenced.

engage with the whole of Johnson-Laird's theory and its implications, focusing instead merely on one of the suite of hypotheses that make up his philosophy of mind in isolation (i.e., his introduction of mental symbols for such things as negation). This is a finding that must surely meet our earlier definition as an unusually small and an unusually univocal number of publications. As such, this study has attained its aim and confirmed our assertion that there currently exists a paucity of publications that critically analyse the addition of a correspondence theory of meaning into the hypotheses of Johnson-Laird. Although this small-scale study should in no way be considered an exhaustive analysis, its findings should, due to its method and focus on core articles, be indicative of the general make-up of the wider literature surrounding Johnson-Laird's adoption of a Peircean correspondence theory of meaning. (Additionally, this is a finding that has been supported by other, non-structured, investigations into the literature performed by this thesis' author.)

As such, and in conclusion, it can be seen that there is, according to the preceding study, a lack of focused critical investigation into the success, or otherwise, of Johnson-Laird's inclusion of a Peircean correspondence theory of meaning into his philosophy of mind – a conclusion that confirms the assertion that initiated this study. Furthermore, it must be concluded that any further study that does so, would constitute a contribution to the existent published knowledge in that field.

§v Thesis aim

In light of the above investigation and discussion, this thesis will attempt to contribute to the published knowledge and task itself with filling this gap in the literature. Consequently, this research will undertake an investigation the aim of which will be as follows:

Aim: To assess the validity of the hypothesis that the introduction of a Peircean correspondence theory of meaning into Johnson-Laird's mental model theory of mind saves the latter from the critiques of its commentators and to test the success of the inclusion of a Peircean referentialist semantics into such a representational theory of mind.

Now that this aim has been established, it is necessary to determine how it will be met, i.e. to define an appropriate methodology and method capable of attaining such a goal. Consequently, the proceeding chapter will be dedicated to just such a definition.

Methodology

Chapter Three: Methodology

Introductory remarks

During the process of the preceding literature review, it was established that, according to the psychologist Philip Johnson-Laird (and other referentialists, such as Fodor 2010, Rapaport 2011 and Schweizer 2012), for cognitive science and the RTM to successfully account for meaning and remain a valid hypothesis, it must introduce a correspondence theory of meaning into its philosophy. Specifically, Johnson-Laird argues that for the RTM to successfully and completely explain cognition, then the mental representations that are central to the theory must be composed from atomic elements that correspond with physical objects in the real world and, further, that these atomic elements must then be combined in the exact same relationship as the state of affairs that is being represented. However, the preceding literature review also identified that there exists a dearth of detailed study into the success or failure of the introduction of such a Peircean referentialist semantics and its attended assumptions into Johnson-Laird's RTM. An observation that generated the following research aim:

> **Aim:** To assess the validity of the hypothesis that the introduction of a Peircean correspondence theory of meaning into Johnson-Laird's mental model theory of mind saves the latter from the critiques of its commentators and to test the success of the inclusion of a Peircean referentialist semantics into such a representational theory of mind.

Consequently, due to the novelty of a thorough study in to the addition of a Peircean correspondence theory of meaning into Johnson-Laird's RTM, this investigation will be experimental in nature: It will focus its attention on testing for the success/failure of the semantic updates made by Philip Johnson-Laird. Consequently, the purpose of this chapter is to clearly put forward how this will be validly and reliably achieved. This goal will be realised by identifying, reviewing and justifying a successful research methodology (post-positivism) and strategy (critical philosophical analysis), which will generate and answer the relevant research questions needed to appropriately achieve our research aim. Accordingly, the proceeding analysis will keep to the following structure:

i)   A review of the research methodology the proposed study will adhere to and a clarification of the philosophical (i.e. ontological and epistemological) assumptions that will underlay the framing of the research and the lens through which any results will be perceived.

ii)   The identification of the specific research method and design that will be applied in this instance.

iii)   A clarification of the specific research design this method will implement to achieve the aim of our study.

iv)   An analysis of the validity of this research design.

v)   A defence of the appropriateness of this research design.

vi)   A clarification of the entire group from which the study's sample will be drawn and the entire group to which its findings will be generalisable (i.e. the work this study will draw upon and the theses to which it will be applicable).

§1 Research methodology

Before we proceed, it is first necessary to clarify that every investigation and field of research is constrained by a number of philosophic assumptions; assumptions that provide the framework for any conducted study and determines the lens through which any data is interpreted. These philosophic suppositions are comprised of commitments that are of both an ontological and an epistemological nature, and operate in unison to constitute a research paradigm or methodology. For, as Kuhn (1970, pp4-5) observes, no group:

> could practice its trade without some set of received beliefs[...] Effective research scarcely begins before a[…] community thinks it has acquired firm answers to questions like the following: What are the fundamental entities of which the universe is composed? How do these interact with each other and with sense? What questions may legitimately be asked about such entities and what techniques employed in seeking solutions?

That is to say, researchers must unavoidably view a research problem through a certain prism; they necessarily have axiomatic – sometimes unquestioned – philosophical judgements on the nature of the world (i.e. ontological suppositions) and on how knowledge of that world is obtained (i.e. epistemological suppositions). To clarify and codify these axiomatic suppositions: Ontological research assumptions are ones which concern "what one works with" (Sharrock and Read 2002, p166), while epistemic ones "concern[...] knowledge and how knowledge is formed" (Williamson 2000, p313).

Nevertheless, an important point that merits special attention at this juncture is that, as Kuhn (1970) and Maxwell (2013, p36) highlight, ontological and epistemological traditions and paradigms not only inform an individual's research practice, but are also "shared by researchers working in a specific field or tradition". Therefore, anyone working within a standardised (or "coherent" (Kuhn 1970, p10)) field of research – such as cognitive science – needs to hold similar paradigmatic or methodological views if they are to produce cumulative findings (that is, "findings which build upon each other within frameworks which are more or less settled" (Hughes and Sharrock 1997, p5)). For, if results are produced using methodologies based on divergent ontological and epistemic foundations, then they will be fundamentally unable to engage with one another; they can only ever perceive their findings through divergent lenses and, as such, operate at cross purposes (Sharrock and Read 2002, p167).

However, although a research tradition's methodological assumptions can be both philosophically and logically dubious, the point of this research is not to dispute Johnson-Laird's hypotheses at this level; the aim of this study is to produce results that tackle his theories at face value, not at this lower, paradigmatic, methodological or philosophical position. That is to say, as the objective of this study is to produce data and findings that are acceptable to those working within the RTM (i.e. according to their own lights), the study this research will undertake to validate or falsify a referentialist semantics will operate from within Johnson-Laird's and cognitive science's own paradigm. Consequently, the epistemological and ontological philosophical assumptions of Johnson-Laird and the referentialists (i.e. their paradigm and research methodology) will be, for the purposes of this research, accepted and adhered to throughout. As such, before this chapter can proceed, it is necessary to identify the paradigm/methodology to which Philip Johnson-Laird specifically, and cognitive science generally, adheres, as well as a clear statement of the ontological and epistemological assumptions this commitment entails. Finally, how these assumptions will prejudice or constrain the assessment of method design appropriateness and validity, as well as the interpretation of any data, must also be made explicit.

Methodology, cognitive science and Philip Johnson-Laird

Although it would be largely uncontroversial to identify the field of cognitive science and the work of Johnson-Laird as located within the post-positivist paradigm, evidence for such an attribution, if needed, can readily be found in the literature. For example, post-positivism (as defined by this study) is a methodology in which researchers hold a number of assumptions:

    i) A realist ontology

and

ii) A critical epistemology that recommends that theories can only be accepted if: a) It fits into the nexus of a larger network of beliefs without causing large scale disruption (i.e. on Quinean grounds); or if b) it has passed the test of a varied methodological approach, to triangulate between falsifiable results, without resulting in falsification (i.e. on Popperian grounds).[1]

These foundational commitments for an adherence to a (realist) post-positivist methodology can easily be identified as operating behind the work of Johnson-Laird and many other cognitive scientists. For example, it is common for cognitive scientists to express a realist position on the world and the objects under their study, with those such as Harré (2002, p2) stating:

> The program for cognitive science presented here will be realist[…] pass[ing] beyond what can be perceived by the senses, into the deeper realms of material reality.

Additionally, all of the referentialists' myriad correspondence theories of meaning are predicated on there being an independent reality 'out there', to which mental representations can refer – a realist ontological position that is made explicit when Johnson-Laird (1993, p387) states that he "assumes that there is a physical world in space and time". Additionally, those working within cognitive science commonly assume the validity of their findings is measured by Quinean and Popperian standards. For example, Waskan (2006, p57) states of an hypothesis:

> Our endorsement of the theory[…] is thus not warranted on Popperian grounds (i.e., by the fact that the theory has passed a series of severe tests), but is warranted for reasons that are for more Quinean (i.e., the theory lies at the nexus of a larger network of beliefs, and abandoning it up would cause a large-scale disruption to the coherence and simplicity of this network).

Moreover, cognitive science has inherited its views on the world and knowledge acquisition from modern scientific practices and, further, sees itself as operating within the purview of these practices; tacitly assuming that any findings produced will be cumulative, or coheres, with that of these 'traditional' sciences (see Pylyshyn 1984; Harré 2002; Friedenberg and Silverman 2006; Waskan 2006; Bermúdez 2010). Consequently, cognitive science must be committed to the same methodology and post-positivist paradigm (i.e. a realist ontology and a Quinean/Popperian epistemology) adopted by the traditional sciences since their abandonment of positivism.[2]

Taking these findings into account, then, for this research to be applicable to the field it intends (i.e., cognitive science generally and the work of the Johnson-Laird specifically) it must employ a method

---

[1] However, the focus on Quine and Popper here gives only a highly restricted view of the post-positivist paradigm. Others, such as Kuhn (1970), Lakatos (1984) and Feyerabend (2011) would reject this definition. However, throughout this research, 'post-positivism' will be taken to be a realist position as defined by Quine (1960) and Popper (1968).

[2] Again, this claim glosses over much of the debate surrounding 'the method' of science and is a simplification of the true state of affairs.

that is conducive with the post-positivist paradigm and acceptable to its methodology, i.e. it must be conducive to a realist ontology and a Popperian/Quinean epistemology. What needs to be made explicit, however, is what a commitment to these positions entails and the affect that they have on the assessment of design appropriateness.

Methodological entailments

Although a clear statement of a realist ontological commitment should be rather intuitive (i.e. a belief in a real world 'out there' that is both independent from an observer and are the causes of the relevant sensations within an observer's consciousness), a well-defined account of the post-positivist Popperian/Quinean epistemology and its influence is required before we advance. Consequently, and to clarify: The epistemology of post-positivism arose from criticisms, most notably by Quine (1953; 1960; 1987; 1999; 2008) and Popper (1957; 1968; 1974; 1994), levelled against the positivist paradigm prevalent in the natural sciences in the first half of the last century; criticisms that rejected the classical positivist epistemological position of knowledge as justified true belief, while replacing them with others (see Popper and Miller 1983; Quine 1987).[3, 4] What needs to be clarified, however, is that post-positivists' critique of positivism's epistemology comes from two separate directions and leads to a sophisticated, critical conception of knowledge. As such, post-positivism relies on a complex epistemology that draws on both Quinean and Popperian theories.[5]

Quinean epistemology

During Quine's critique of the positivist epistemology, he argues that such a view of knowledge as justified true belief falls victim to a vagueness of boundary (akin to the Sorites paradox) due to its reliance on certainty. For, although this view of understanding necessitates absolute certainty over one's knowledge, the concept of certainty (like the concept 'heap') comes in degrees: We can be more certain of some facts than we are of others, and to limit what is considered knowledge as to be those ideas that are absolute would "do violence to both the usage and utility of the word [knowledge]" (Quine 1987, p110). Furthermore, Quine strengthens his case here with an argument he terms the 'doctrine of empirical under-determination'. This is the argument that:

> Scientists invent hypotheses that talk of things beyond the reach of observation. The hypotheses are related to observation only by a kind of one-way implication; namely, the events we observe are what a

---

[3] The view of knowledge as justified true belief is also – pejoratively – referred to by Popper and Miller (1983, pp105-107) as the 'common sense' or 'bucket' theory of knowledge.

[4] See also Gettier (1963) and the extended literature concerning 'Gettier cases' for similar arguments against the concept of knowledge as justified true belief.

[5] The divergence between these two post-positivist epistemologies arose from their author's rejection, or acceptance, of reductionism, i.e. "the supposition that each statement, taken in isolation from its fellows, can admit of confirmation or infirmation at all" (Quine 1953, p41) – that is, Quine's second dogma of empiricism. Popper's epistemology accepts reduction and, as such, relies on the falsification of the statements of science individually (see Popper 1968, p57ff.). Alternatively, Quine (1953, p41) rejects reductionism and offers the "counter suggestion[…] that our statements about the external world face the tribunal of sense experience not individually but only as a corporate body".

> belief in the hypothesis would have led us to expect. These observable consequences of the hypotheses do not, conversely, imply the hypotheses. Surely there are alternative hypothetical substructures that would surface in the same observable ways (Quine 2008, p228).

That is to say, as long as the argument exists that the suite of hypotheses that make up the body of science could, theoretically, be replaced by a logically distinct, coherent suite of competing hypotheses, which still match the available data, then the sciences' claim to indisputable, justified knowledge is undermined. Interestingly, this analysis of Quine's leads him to a position reminiscent of that of the French physicist, historian and philosopher of science Pierre Duhem (1861-1916), who put forward the thesis, expressed by Ariew (1984, p313) "that there are in principle an indefinite number of theories that fit the observed facts more or less adequately".

Following this analysis, Quine (1987, p110) concludes that for "scientific or philosophical purposes the best we can do is give up the notion of knowledge as a bad job" and, as such, replaces the traditional positivist epistemology with one he terms holism. This is the view that, as we cannot be certain of any individual hypothesis, what we are concerned with is the suite or nexus of hypotheses we hold and the accordance between them. What now determines the acceptance of an hypothesis as knowledge, therefore, is its acceptance into this nexus of existing beliefs without disturbance. Quine (*Ibid.*, p228) articulates this epistemology as follows:

> It is holism that has rightly been called the Duhem thesis and also, rather generously, the Duhem-Quine thesis. It says that scientific statements are not separately vulnerable to adverse observations, because it is only jointly as a theory that they imply their observable consequences.

This new view of knowledge can be further clarified by an analogy that comes from Otto Neurath:

> Neurath has likened science to a boat which, if we are to rebuild it, we must rebuild plank by plank while staying afloat in it[…] If we improve our understanding of ordinary talk of physical things, it will not be by reducing that talk of physical things, it will be by clarifying connections, causal or otherwise, between ordinary talk of physical things and various further matters which in turn we grasp with help of ordinary talk of physical things. (Quine 1960, p3)

Before this exposition of Quine's holism concludes, however, we first need to acknowledge how nuanced the theory actually is. For, there exists a development of Quine's position over the course of his writings: In his earliest articulations of holism (e.g. in "Two dogmas of empiricism") Quine expresses an extreme, or strong, variation of holism, stating that no hypothesis can be accepted in isolation from *every* other (i.e. certainty only comes when an hypothesis fits with *all* of a culture's existing knowledge). In later writings Quine was to moderate this extreme view (e.g. in *Theories and things*), positing a weaker version of holism, wherein certainty of an hypothesis rests in its acceptance

with *many* other hypotheses or with modules, or 'chunks', of knowledge (i.e. "It is an uninteresting legalism[…] to think of our scientific system of the world as involved e*n bloc* in every prediction. More modest chunks suffice, and so may be ascribed their independent empirical meaning, nearly enough, since some vagueness in meaning must be allowed for in any event" (Quine 1999, p71)).

Popperian epistemology

In contrast to Quine, Popper's rejection of the positivist's conception of knowledge rests on his argument that, although knowledge is true (i.e., it corresponds to the independent facts as they are), it can be neither justified (as this leads to an infinite regress of justification – what justifies an individual's justification? (Popper 1974, pp21-24)) nor based on belief (as knowledge can be stored in such things as books, which are incapable of belief (*Ibid.*, p24)). After abandoning this view of knowledge, Popper (1968, p18) draws on Hume's problem of induction to determine a new 'criterion of demarcation' for the empirical sciences (i.e., criteria for delimiting science from pseudo-science), concluding that knowledge of the world is not only difficult to attain but must always remain uncertain, i.e. that we can only ever be certain that knowledge is wrong, not that it is right. Popper argues, therefore, that research is not in the business of confirming hypotheses, but of refuting, or falsifying, them (Popper 1957; 1968; 1974; 1994; Popper and Miller 1983).[6] This new criterion of demarcation is a standard Popper (1968, p18) terms 'falsifiability' and defines thus:

> I shall not require of a scientific system that it shall be capable of being singled out, once and for all, in a positive sense; but I shall require that its logical form shall be such that it can be singled out, by means of empirical tests, in a negative sense: it must be possible for an empirical scientific system to be refuted by experience. Consequently, only methods capable of being repeated and of generating reproducible data will be deemed valid.

Accordingly, armed with this standard of falsifiability, Popper (*Ibid.*, p20) posits that only the 'fittest' of theories, i.e. those which have survived "the fiercest struggle for survival", should be accepted as knowledge (or, more accurately, as no theory can be confirmed, still remains as a "provisional conjecture" (*Ibid.*, p264)) and, therefore, recommends the application of as multifaceted and a variegated a test as possible to any thesis (see Popper and Miller 1983, p15). That is to say, to remain as valid, as a 'provisional conjecture', under a Popperian epistemology, a theory must be rigorously tested by not just the classic quantitative methods but also, where appropriate (i.e., where results are repeatable and falsifiable (*Ibid.*, p150)), by other approaches such as qualitative ones.

---

[6] To clarify, this is a *naïve methodological falsificationist* reading of Popper, rather than the *sophisticated methodological falsificationism* some, such as Lakatos (1984, p31), sometimes ascribe to him. The reading found here is 'naïve' in that it holds theories can in fact be falsified, that is:

> For the naïve falsificationist any theory which can be interpreted as experimentally falsifiable, is 'acceptable' or 'scientific'. For the sophisticated falsificationist a theory is 'acceptable' or 'scientific' only if it has corroborated excess empirical content over its predecessor (or rival), that is, only if it leads to the discovery of novel facts. (*Ibid.*, pp31-32)

As such, this thesis' use of 'falsification', 'falsify' and 'Popperian epistemology' will refer to naïve methodological falsificationism and will assume that theories can be disproved in their own right.

Methodological constraints

Now that the above exposition has led us to an appreciation of what an adherence to the post-positivist paradigm of Johnson-Laird and cognitive science entails (i.e. the adoption of a realist view of the entities under our scrutiny and an adherence to Popper's criterion of demarcation and Quine's holism), what needs to be established is how these positions constrain our method design and determine method appropriateness.

Therefore, it first needs to be made explicit that, due to the constraints of a realist ontology, any entities scrutinised by this research must be deemed to be objectively real, i.e. must not be considered wholly dependent on any one observer or group of observers. Secondly, when reviewing the validity of a potential research method design under the lights of a Popperian epistemology, it is fundamental to ensure, if it is to be considered valid, that the method meets that epistemology's methodological rules and criterion of demarcation, i.e. is repeatable and produces results that are themselves falsifiable or testable. Finally, from the perspective of a Quinean epistemology, for a prospective method to be valid within cognitive science and the work of Johnson-Laird it must operate within the same conceptual scheme, the same suite or nexus of hypotheses; i.e. it must be 'naturalised'. However, for a method to be naturalised in this manner, it must adhere to two theses: 1) The no first philosophy thesis and 2) the continuum thesis (see Colyvan 2001; Roland 2014). That is to say, for a method to be applicable in this instance it must not hold a "body of truths which are prior to and firmer than those which science has to offer" (Siegel 1984, p671) and must be continuous with cognitive science, not separate to and from it, i.e. its relationship with the existing work must be such that it "share[s] various methods and principles, and that the problems that they each pursue differ only in degree of generality" (Daly 2010, p187).

To summarise this discussion, then, to produce findings that are cumulative to and with those of Johnson-Laird and cognitive science, the method this study implements must concern itself with non-relative, objective entities, be both repeatable and naturalised and produce findings susceptible to falsification.

§2 Research method

Taking the above methodological discussion into account, the standard repeatable, naturalised research approaches to apply in such instances are either quantitative or 'traditional' qualitative

methods.[7] These methods will be, following Straus and Corbin (2008, p11), defined and delineated by this research as follows:

> **Quantitative methods:** Any type of research that produces findings arrived at by statistical procedures or other means of quantification
>
> **Qualitative methods:** "[A]ny type of research that produces findings not arrived at by statistical procedures or other means of quantification".

In practice this, generally, means the implementation of such methods as experimentation, observation and surveys (quantitative methods) or interviews, ethnography, case studies and phenomenology (qualitative methods) (see Davies 2007; Gray 2009; Denscombe 2010). Nevertheless, although this may be the case, it needs to be highlighted that method appropriateness and method selection are not dependent only upon the paradigm and methodology by which one is constrained, it is also determined by the individual characteristics of the specific study and entities of interest (see Kumar 1996; Vogt et al 2012; Maxwell 2013) – in this instance that means semantic content and the correspondence between mental representations and an ontologically real world. However, once this extra facet is taken into account, it is found that these 'traditional' methodological approaches will not be valid in this instance. For the purposes of transparency, this is a conclusion that merits justification.

Suitability of quantitative methods

As mentioned, when considering the suitability of quantitative methods for studying the entities that concern this research, issues arise. This is because such entities and concepts as meaning and a mental representation's correspondence with the world cannot be directly empirically tested; all that can be measured is behaviour which has been inferred from the entities and concepts we are positing. As such, we can never be certain if the mental entities we are putting forward are, in fact, the actual causes of the actions we are studying, or whether some other fundamental and unknown source or variable is the origin of the measured behaviour. The primary difficulty identified here can be stated as follows: In our case, there is no way we can be certain that the independent and dependent variables we isolate within any experiment are in fact the variables we suppose them to be.

---

[7] Although quantitative and qualitative methods are here being said to be distinct, clearly defined categorisations of research types, it needs to be highlighted that the two may not necessarily be mutually exclusive (see Patton 2002). In recent years many theorists have eschewed the 'paradigm wars' and have instead expounded the validity of approaches that combine both method types within the one piece of research (see Slater 1988; Straus; Patton 2002; Corbin 2008). This is a research strategy known as mixed methods research and is an approach whose inception and application is, essentially, motivated by the concept of triangulation. This is the idea that, as it can never be known for certain that any individual method or method type is infallible, it is prudent to exploit a number of differing approaches so as to zero in on those results that crop up again and again and are, therefore, likely to be accurate, i.e., to identify those results that are likely to be uninfluenced by any one method.

This is deserving of clarification: Although it is perfectly valid for psychology to measure mental phenomena by their physical manifestations, in our case this cannot be made certain. An example experiment should better delineate this issue – if we were to state one formally:

Method: Independent measures experimental design.

Dependent Variable: Correlation between a mental model and a corresponding state of affairs.

Independent Variable: The meaning of the mental model.

Data: Count of normative responses concerning participants' meaning of their mental models (i.e., number of meanings that do and do not match the state of affairs).

The above – very rough – description for an experiment would potentially generate highly informative data and possibly even meet our research aim were it able to be scrupulously put into practice. However, the variables in question, and which it necessitates, are, unfortunately, ones that concern entities that are currently too far removed from our current scientific and technical knowledge to be numerically measurable in both a confident and unerring fashion. That is, an experiment of the type in the above proposal cannot be valid while we simply do not know enough about the operation of reference, of the correspondence between mental elements and the world, for it to be unfailingly recreated. Therefore, while we are unable to ensure that we have experimentally established a direct link between a participant's particular mental model and a particular state of affairs and, ergo, that our experiment is well founded, we will not be able to state that our results pertain to our research aim. In other words, if we do not know how the correspondence we are concerned with physically operates (a question that exceeds the concerns of this research), we cannot be certain that we can, or have, synthesised it. This question (how the link between our minds and the world operates) is one for future psychological or scientific research to attempt, and only once it has been answered can we conduct any quantitative experiment in the knowledge that it is legitimate. This issue still stands even if we follow James (1982, p151) and limit the purview of psychology and our research to only "the correlation of mental states with brain-states"; i.e. neuroscience. For, in this instance, the relation between mental states and part of the world could feasibly be measured via some form of scan (providing an accurate measure for one of our variables is granted), but how meaning could be measured via this process is something that still needs to be answered.

Consequently, and in summary, where quantitative methods are concerned, this is our problem: We can only investigate what concerns this study via their physical effects, yet if we do not know enough about these effects (as is the case in this instance) we cannot be certain that what we are measuring is accurate, that our experiment is valid and that the data we gathering bears any relation to the aim with which we are concerned. As such, a quantitative method design will not be appropriate in this instance.

Suitability of qualitative methods

In the case of qualitative methods, it is important to note that the combination of cognitive science's, and, therefore, this research's, post-positivist epistemology and the type of entities here investigated, results in the conclusion that many 'traditional' qualitative methods (e.g. ethnography, interviews, case studies etc.), would, in this instance, not be able to produce valid results. Any study that relied upon qualitative anecdotal evidence from interviews or case studies, would involve the implementation of methods that are neither repeatable nor falsifiable, which is, therefore, invalid according to a Popperian epistemology (Popper and Miller 1983, p150). Furthermore, naturalising qualitative data concerning individuals' experience of meaning within the nexus of the referentialists' hypotheses may be problematic, if not impossible. For example, Johnson-Laird (1983, p6) has declared that "If the long promised Newtonian revolution in the study of cognition is to occur, then qualitative explanations will have to be abandoned". Moreover, Johnson-Laird (*Ibid.*, p2) has also stated that the processes of cognition are inaccessible via reflection or introspection and, therefore, cannot be unerringly described by interview, case study etc. Consequently, as this thesis sees no way of aligning 'traditional' qualitative data concerning reference and semantic content into the suite of Johnson-Laird's and cognitive science's hypotheses, 'traditional' qualitative methods will also have to be considered an invalid approach to this research on Quinean, as well as Popperian, grounds. As such, any form of 'traditional' qualitative method would be neither valid nor appropriate in this instance.[8]

Suitability of alternative methods

Although the above is the case, a potential 'non-traditional' qualitative method does present itself as being both a valid and fruitful alternative: That of the philosophic method of critical analysis. Although this method of philosophy is not considered a 'traditional' qualitative method, that should not discount it as a possible approach in this instance. For, critical analysis both coheres with our earlier definition of qualitative research (i.e. it is a "type of research that produces findings not arrived at by statistical procedures or other means of quantification" (Straus and Corbin 2008, p11)) and is considered an acceptable method under both Quinean and Popperian epistemologies (this latter point is one that will be examined in detail later in this chapter). Furthermore, the application of critical philosophic analysis stands as a potentially appropriate method in this instance as it has the capability

---

[8] As alluded to in a previous note, the exclusion of both quantitative and qualitative methods still leaves the possibility that a mixed methods approach could feasibly allow us to identify valid and reliable data via triangulation. However, it should be noted that by combining methods designed to produce potentially incompatible data (i.e. numbers and words), there is no guarantee that what is produced from mixed methods research is in any way harmonious. Indeed, as Patton (2002, pp465-466) avers, when dealing with mixed methods, a researcher "should not expect that the findings[…] will automatically come together to produce some nicely integrated whole". Additionally, the constraints of available resources (time, equipment etc.) mean that the application of both a quantitative method – if one were feasible – and a qualitative method – if one were valid – would not be possible in this instance, for sufficient resources only exist for the undertaking of the one method. Ergo, only the application of the one method, the most potentially fruitful, could be undertaken given the circumstances. Consequently, even if a valid yet unproductive quantitative experiment were possible to support invalid though fruitful qualitative findings, the application of such a mixed method would not be practicable.

to determine both the logical accuracy of a correspondence theory of meaning as well as determining the effect of the introduction of such a theory into Johnson-Laird's mental model theory as a whole (i.e. determine the hypothesis' internal and external logical consistency). Consequently, it appears that the successful application of such a method could be capable of achieving our aim by establishing the logical consistency of Johnson-Laird's thesis. As such, it would be prudent at this juncture to proceed with a detailed enquiry into this putative method of critical philosophical analysis.[9]

Analysis and synthesis

An important point to highlight whilst introducing the philosophical method of critical analysis is the fact that it is an approach more accurately labelled as 'analysis and synthesis'. Furthermore, it needs to be noted that there are actually multiple conceptions of analysis that have variously been practiced throughout the history of philosophy; beginning with the ancient Greeks and continuing up through to the modern period (where, despite its name, 'Analytic philosophy' is not the sole reserve for the method of analysis today). These analytic approaches range from the 'regressive' analysis of the ancient Greeks, such as Plato and Socrates (where an individual works backwards from what is being sought (i.e. the nature of a concept) so as to better understand it, e.g. we can regress backward from 'knowledge' to identify it as 'justified true belief'), to the 'transformative' analysis seen in the works of such modern philosophers as Frege (1952) (where a problem, statement or thesis, such as language, is translated into a more transparent format (i.e. predicative logic) so as to make its operation more perspicuous). However, though these differing conceptions of analysis exist, it should be clarified that they are not competitive, not mutually exclusive, approaches and can be applied in unison so as to achieve a united goal. Although this is the case, this section will focus upon the more general approach of analysis, an approach that can be seen as 'decompositional' and has been favoured by thinkers such as Russell (2010, pp147-148). Consequently, this section will be of a more inclusive nature and will exclude much of the details concerning the nuanced positions that exist in the subset of analytical methods.

With these initial points clarified, we can begin to develop an understanding of our putative method of analysis and synthesis by first looking at these terms' etymologies: The derivation of 'analysis' comes from the ancient Greek '*Analusis*', meaning 'an unravelling' or 'a loosening', while 'synthesis' comes from the Greek '*syntithenai*', or 'put together, combine'. As such, when applied to a problem,

---

[9] We defend this focus on logical consistency as its presence would show that a suite of hypotheses/facts/beliefs etc. can *all* be true, at the same time, in *some* possible world (although not necessarily ours), i.e., it would identify that a set of hypotheses etc. do not contradict each other (regardless of whether they are currently true or not in this world). This is important as the corollary of this is that if a suite of hypotheses etc. are *not* logically consistent, those hypotheses etc. must contain some falsity in *all* possible worlds (including this one). Therefore, although establishing consistency of a theory is insufficient to determine its accuracy, the identification of any dissonance gives an *a priori* method of showing that a theory contains falsity.

the aim of analysis and synthesis is to first unravel the issue, to decompose it into its component parts, before recombining these components back into a whole. Although this process may seem self-defeating, merely disassembling and reassembling a problem, the power of such an approach can be seen in the following quote (originally found in the 1662 work by Antoine Arnauld, *Port-Royal Logic* (or, *Logique de Port-Royal*), but here quoted from Finocchiaro (2005, pp255-256)):

> The art of arranging a series of thoughts properly, either for discovering the truth when we do not know it, or for proving to others what we already know, can generally be called method.

> Hence there are two kinds of method, one for discovering the truth, which is known as *analysis*, or the *method of resolution*, and which can also be called the *method of discovery*. The other is for making the truth understood by others once it is found. This is known as *synthesis*, or the *method of composition*, and can also be called the *method of instruction*.

As such, analysis is used as a method of discovery; logically unravelling a problem so as to clarify the facts that compose it and making them, the problem itself and the whole's logical operation more perspicuous. Whereas synthesis is a method of instruction, a way to concurrently demonstrate the logical operation of the problem whilst also functioning as a proof of the complex hypothesis, i.e. as an explanation, a demonstration of the 'why', of the original statement. Furthermore, this project of analysis and synthesis is one that also tests whilst it discovers and demonstrates. For, the method of analysis will automatically uncover the logical consistency of an hypothesis' operation once its workings are unravelled and synthesis demonstrates, proves, whether and why that hypothesis' operation is, or is not, logical.

What needs to be clarified at this juncture, however, is that critical philosophic analysis, or the method of analysis and synthesis, is, like any other research method, not guaranteed to produce accurate data. For, as Russell (1918, p498) observes:

> When you pass from the vague to the precise by the method of analysis and reflection that I am speaking of, you always run a certain risk of error. If I start with the statement that there are so and so many people in this room, and then set to work to make that statement precise, I shall run a great many risks and it will be extremely likely that any precise statement I make will be something not true at all.

As such, philosophers have developed a rigorous and systematic set of analytic 'tools', such as rules of inference (e.g. *modus ponens*, biconditionals, disjunctives etc.), logical fallacies (e.g. ad hominem, appeal to authority, straw man etc.) and even thought (or, *gedenken-*) experiments to enable hypotheses and arguments to be analysed and synthesised in an orderly, perspicuous and, most importantly, a repeatable manner (see Morton 2004; Baggini and Fosl 2010; Daly 2010; Papineau

2012; Haug 2014).[10] This systematic use of philosophic tools throughout the process of analysis and synthesis is most clearly seen in transformational analysis, where propositions are translated into formal logic and then *formally* analysed and synthesised. However, the same process is followed where normal (i.e. non-transformed) discourse is concerned. Ergo, akin to scientists employ Bunsen burners, microscopes, petri dishes etc. to execute the method of experiment in an orderly, repeatable fashion, so philosophers employ rules of inference, logical fallacies and thought experiments to execute the method of analysis and synthesis.

§3 Research design

Now that we have established which method will be applied to meet our research aim, it is necessary to elucidate the particular research strategy this study intends to follow, i.e. to clarify the specific design in which the method of analysis and synthesis will be implemented. As such, to assess the validity of the hypothesis that concerns this research (that the introduction of a Peircean correspondence theory of meaning into the RTM, saves the latter from the critiques of its commentators), Johnson-Laird's arguments and theories will be analysed so as to discover if they can answer in the affirmative to the following research questions:

1) Does Johnson-Laird's updated account of meaning circumvent the semantic objections of the RTM's commentators (i.e. the CRA, the SGP, holism and the homunculus and symbolic fallacies)?

2) Is Peirce's iconic correspondence theory of meaning internally and externally consistent?

3) Is the introduction of a correspondence theory of meaning, generally, into Johnson-Laird's broader philosophy of cognition logically consistent?

These particular three questions have been identified as they exploit critical philosophical analysis' ability to derive and examine an hypothesis' logic in such a manner that we will be able to interrogate Johnson-Laird's arguments from all relevant angles and meet our aim. The specific rational for each question is as follows:

- Question 1) addresses the success of Johnson-Laird's argument that a correspondence theory of meaning can avoid the critiques of certain commentators (see Johnson-Laird 1993, p34). It is necessary to ensure that Johnson-Laird's hypotheses are able to answer in the affirmative to this question as these arguments were a motivating factor in his rejection of syntactic semantics (see *Ibid.* 1988, p115; 1993, p348). Therefore, if these challenges remain successful

---

[10] Thought experiments are, following Daly (2010), being classified as a philosophical 'tool' and not a method here, as *gedankenexperiments* are being taken as a form of conceptual analysis. That is, as a tool to be employed in the service of a method.

even after Johnson-Laird's amendments, then his referentialist semantics will be labelled as being redundant.

- Question 2) concerns itself with the accuracy of a Peircean correspondence theory of meaning in its own right as well as the eligibility of its introduction into Johnson-Laird's pre-existing suite of hypotheses. That is, it will establish if the theses that comprise Peirce's semantics logically cohere in their own right, as well as whether there is any dissonance between the constituent theses of Johnson-Laird's updated semantics and those of his wider philosophy of mind. As such, question 2) determines whether Johnson-Laird's strategy is an appropriate inclusion into any account of cognition generally, and Johnson-Laird's specifically.

- Question 3) attempts to uncover whether the assumptions implicitly operating within any correspondence theory of meaning (i.e., those we uncovered during our literature review and that are necessary for any referentialist hypothesis' successful operation) are logically consistent with the existing theoretical positions held by Johnson-Laird. That is, it establishes if there is any dissonance between Johnson-Laird's RTM and correspondence theories broadly.

Consequently, and in summary, the method design this study will implement will be to apply critical philosophic analysis (i.e. decompositional analysis and synthesis) to Johnson-Laird's hypotheses, thereby uncovering their logical accuracy and success; the rules of inference, logical fallacies and thought experiments will be applied to these hypotheses so as to analyse them, breaking them down into their constituent logical parts and thereby perspicuously demonstrating their workings. This process will be implemented so as to establish the logical operation of Johnson-Laird's updated semantics: Establishing both the rational process of his referentialism in its own right, as well as that of the full suite of beliefs that constitute Johnson-Laird's contemporary philosophy of cognition. Once this process has been achieved, and the logical operation of Johnson-Laird's arguments have been made perspicuous, the success of these arguments will be judged on three metrics:

1) The relation they bear to the CRA, the SGP, holism and the homunculus and symbolic fallacies.

2) The logical coherence of their specific correspondence theory of meaning in its own right, i.e. on the internal consistency of Johnson-Laird's semantics, and the logical coherence

between their iconism and the full suite of Johnson-Laird's account of cognition, i.e. on the external consistency of Johnson-Laird's semantics.

3) On the logical consistency with Johnson-Laird's existing philosophy of mind with the positions necessary for a correspondence theory of meaning generally.[11]

Therefore, if analysis and synthesis shows that Johnson-Laird's semantics is still susceptible to such criticisms as those advanced by Searle (1980) and Lewis (1970), or if it is found that the hypotheses that make up Johnson-Laird's account of semantic content is either internally or externally inconsistent, then the theory has been falsified. If, however, Johnson-Laird's proposal is found to stand up to the CRA, the SGP, holism and the homunculus and symbolic fallacies, as well as being both internally and externally consistent, and is demonstrably so be synthesis of the proposals' analysed parts, the theory will be said, following Popper (1968, p264) to stand as a "provisional conjecture".

Now that this research design has been proffered, it is a prerequisite of this chapter to demonstrate its validity *in this instance*, i.e., to establish that the application of philosophical analysis to the work of Johnson-Laird meets the standards of the post-positivist paradigm. (N.B., as we were originally led to the acceptance of post-positivism from Johnson-Laird, there will unavoidably be some repetition from §1 of this chapter. However, it is incumbent upon any investigation to clarify that its method accords with its methodology, so this reiteration is unavoidable.)

§4 Design validity

Before commencing this discussion proper, it is interesting to note that, historically, analysis and synthesis has proved itself as both a valid and fruitful research approach when applied to cognitive science research. For, critical philosophic analysis has consistently played a central role in cognitive science since the field's inception in the latter half of the 20th century. That is, throughout cognitive science's lifetime, philosophers have validly worked alongside psychologists, linguists, computer scientists and others to develop and refine hypotheses central to the field. While doing so, philosophers have made significant contributions in a variety of areas: From work on the theory of functionalism during the discipline's nascent stages (see Putnam 1980), through to tackling today's cognitive issues (see Fodor 2010). Furthermore, as Brook (2009), Dennett (2009) and Thagard (2009)

---

[11] The success, or otherwise, of Johnson-Laird's theory could also be judged via an examination of the internal consistency of his existing mental model theory of mind. However, this analysis is beyond the scope of this research, where only Johnson-Laird's arguments concerning the accuracy and necessity of a correspondence theory of meaning is being considered. As such, this research will be taking these hypotheses, like it did with those that make up Johnson-Laird's methodology, as given.

observe, this use of philosophy within cognitive science rests on sound methodological foundations. For example, as Brook (2009, p221) argues:

> The results of [philosophic] activities are theories and models offered by philosophers similar to theories and models offered by others in cognitive science[…] The main difference is that philosophers tend to go after bigger and sometimes more abstract objects than researchers with other backgrounds[… E]xcept for level of generality and abstraction, there is nothing distinctive to this methodology [of philosophy].

Moreover, and as Papineau (2014, p177) observes, the philosophical method has also historically played a valid role in the 'traditional' sciences in the form of thought experiments: Archimedes' analytical questioning of buoyancy, Galileo's thought on falling bodies and motion and Einstein's analysis of the behaviour of light are just a few of the examples where scientists have applied analysis and synthesis to problems, unpacking the consistency of concepts and thereby producing valid and acceptable findings (see also Gendler 2010).

Nevertheless, it is incumbent upon this investigation to demonstrate *why* such an approach is, according to post-positivist lights, both ontologically and epistemologically valid in this instance, rather than merely stating historical precedence. That is, and as we have already seen during our discussion concerning the post-positivist paradigm, for the philosophic method of analysis and synthesis to remain valid in this application, our study must be shown to meet the standards of the Popperian and Quinean epistemologies and must be shown to view the entities under its scrutiny as being objectively real.

Ontological validity

It should be remembered, from our introduction to the RTM, that a foundational tenet of Johnson-Laird's approach is Marr's tri-level hypothesis (see Marr 1982; Pylyshyn 1984; 1999). This is a theory that states that the one process of cognition can be described at three distinct levels of abstraction. That is, it posits an ontological monism but a conceptual ternary and argues that the processes of the mind can be detailed at either: 1) The computational, 2) the representational or 3) the physical level. As such, although Johnson-Laird, and therefore this research, is concerned with such entities as propositional attitudes and mental representations (i.e. with entities at the first and second level of abstraction), it should be remembered that although these entities may appear to lack ontological substance on first inspection, they are, in reality, conceptual abstractions from Marr's third level, and are, as such, subsequently founded upon a physical substance – i.e. the brain or the whole person (see Waskan 2006, pp36-76). That Johnson-Laird adheres to Marr's tri-level hypothesis is implicit in his adherence to a form of RTM, but it is a commitment that has also made explicit. For example, he says that:

> [I]f you want to understand *how* the mind works then you had better first ask *what* it is doing. This distinction has become familiar in cognitive science as one that Marr (1982) drew. (Johnson-Laird and Byrne 1991, p17)

And:

> To solve any such problem [as computation], we must bear in mind a lesson from computation: we need at least three different levels of explanation. (Johnson-Laird 1993, p58. See also *Ibid.*, p390).

Moreover, there are a number of further instances in which a commitment to an ontologically real physical world 'out there' can be seen in Johnson-Laird's writings. For such instances we can look to his approval of Craik's 'relation structure' and 'physical models' (Johnson-Laird, 1980, p73; 1983, p9), as well as his use of terms like 'the world', 'concrete objects' and 'entities' as examples of things modelled by his mental representations (*Ibid.*, 1980, p89; 1983, pp.410-415; 2002, p80; 2013, p3; Johnson-Laird, Legrenzi and Legrenzi 1972).

Consequently, this study, following cognitive science generally, will take an ontologically realist view upon the entities under its scrutiny (i.e. semantic content), in that it takes them to be ultimately derived from the physical structure and operation of the brain. Consequently, and to clarify, this study will be adhering to a realist ontology commensurate with that of the post-positivist paradigm.

Nevertheless, now that we have established this study's ontological suitability, it is incumbent upon us to now investigate whether this method is cumulative to and with the post-positivist epistemology and, therefore, is falsifiable and repeatable (in accordance with Popper) and naturalised (in accordance with Quine).

Epistemological validity

As we have seen, to be valid under the lights of a Quinean epistemology, our method must adhere to the 'no first philosophy' and the continuum theses and operate within the bounds of cognitive science and work within that field's existing nexus of hypotheses, i.e. it must be naturalised and constrained by the currently standing theses and empirical findings of the field. That critical philosophical analysis is, in this instance, a naturalised process is best seen in relation to the information mined throughout our literature review, as well as the discussions found in this chapter, all of which will constitute the foundation and framework that determines and delimits what "body of truths" is to be analysed and synthesised (Siegel 1984, p671).[12] Consequently, this study attempts to position itself within both the

---

[12] To clarify, this thesis is therefore taking a naturalist metaphilosophical stance and is stating that philosophy is continuous with the sciences. Further, this thesis is taking a 'constructive naturalist' position, where:

> constructive materialists" take the sciences to be trying to discover facts about the world and take philosophy to also be involved in this endeavour[…] Faced with the task of placing prima facie non-natural phenomena – such as mentality, mathematics, or mentality – in the natural world, these naturalists think that we should start with our best understanding of these phenomena themselves and determine if

suite of hypotheses held by cognitive science and that field's methodology, so as to share various methods and principles and to pursue problems that differ only in degree of generality (see Daly 2010, p187).[13]

However, and as we have also seen, for a method to be valid under the methodology of cognitive science, it must be commensurate to not only a Quinean epistemology but also a Popperian one. To establish method validity under this latter epistemology, it is required to demonstrate that the design is repeatable and its results falsifiable. As such, it should be remembered from our previous discussion of philosophic analysis and synthesis that the method follows a standard, repeatable progression to attain its conclusions. This is a standardised progression most clearly apparent in the practice of transformational analysis (e.g. Frege 1952) and is founded on the rigorous and systematic application of analytic 'tools', such as rules of inference (e.g. *modus ponens*, bi-conditionals, disjunctives etc.), logical fallacies (e.g. ad hominem, appeal to authority, straw man etc.) and thought experiments to enable hypotheses and arguments to be analysed and synthesised in an orderly, perspicuous and, most importantly, repeatable manner. This orderly process, which makes our proposed research design reliable (i.e. determines to what extent other researchers will arrive at similar results if they studied the same case using exactly the same procedures) also provides the method with the ability to be falsifiable, making it subject to Popper's criterion of demarcation and, therefore, valid (see Popper 1968, p18). For, the derivation of results via any rigorous, systematic and repeatable procedure, including critical philosophic analysis, can be reapplied by other researchers and, in that way, tested for accuracy and precision. This is a fact acknowledged by Popper himself, when it is stated:

> There is only one way to make sure of the validity of a chain of logical reasoning. This is to put it in the form in which it is most testable: we break it up into many small steps, each easy to check by anybody

---

As such, this investigation is innately opposed to any non-naturalist (i.e., any denial that philosophy is continuous with science) or deflationary (i.e., holds that philosophical questions should be explained away) metaphilosophical positions and arguments (e.g., Siegel 1984; McDowell 2009; Williamson 2014a; Williamson 2014b; McGinn 2014). However, akin to our position on the post-positivist paradigm, this thesis will not be engaging with these metaphilosophical debates and the accuracy of constructive naturalism will be merely taken as given by this investigation.

[13] Nevertheless, there may be some conflict here. For, within the necessary list of principles, which must be held to adhere to Quinean naturalism identified by Roland (2014), semantic, as well as theoretical, holism is included (i.e. that "only sufficiently rich collections of statements, as opposed to individual statements, have empirical content" (*Ibid.* pp45)). However, this investigation and its correspondence based semantics is opposed to this position on content. As such, if it is taken that Quinean naturalism necessarily entails theoretical *and* semantic holism, the naturalist metaphilosophical position of the paper is challenged. Additionally, this thesis' metaphilosophical position will be that only theoretical, and not semantic, holism is a necessary position for a Quinean naturalism. For further clarification of this point, see §1 of this thesis' Findings chapter.

who has learnt the mathematical or logical technique of transforming sentences.[14] (Popper and Miller 1983, p155)

Furthermore, that the application of such a method to establish inconsistency is an epistemologically valid method when attempting to refute a theory, can be seen in the importance Popper (1968, p72) places on a system being consistent, and his recommendation that self-contradictory hypotheses should be rejected:

> The requirement of consistency plays a special role among the various requirements which a theoretical system, or an axiomatic system, must satisfy. It can be regarded as the first of the requirements to be satisfied by every theoretical system, be it empirical or non-empirical[...] it is not enough to mention the obvious fact that a self-contradictory system must be rejected because it is 'false'[...] the requirement of consistency will be appreciated if one realizes that a self-contradictory system is uninformative. It is so because any conclusion we please can be derived from it. Thus no statement is singled out, either as incompatible or as derivable, since all are derivable. A consistent system, on the other hand, divides the set of all possible statements into two: those which it contradicts and those with which it is compatible. (Among the latter are the conclusions which can be derived from it.) This is why consistency is the most general requirement for a system, whether empirical or non-empirical, if it is to be of any use at all. (See also *Ibid.*, p264)

Additionally, certain comments made by Johnson-Laird imply a tacit acceptance of the validity of critical philosophical analysis. For he states that:

> Reasoned argument alone is seldom responsible for a permanent change in behaviour, even in the methodological habits of scientists. Example is more powerful than precept[.] (Johnson-Laird 1983, pxiii)

Demonstrating that reasoned argument, or philosophic analysis, is an acceptable approach to research within Johnson-Laird's methodology – even if it is not sufficient on its own to alter the behaviour of scientists. This, along with Johnson-Laird's application and approving discussion of philosophical works that employ the method of analysis and synthesis (e.g. Johnson-Laird 1988; 1993; 2002; 2006), demonstrates an implicit acceptance of the validity of our proposed method.

Finally, it should be highlighted that the method of analysis and synthesis is employed by both Quine (1969) and Popper (1968) to establish their respective epistemologies (as both deriving their conclusions through the application of critical philosophic analysis to the concept of knowledge). Therefore, if the field of cognitive science, or Johnson-Laird, were to deny this method as a valid

---

[14] However, it should be noted that, as Lakatos (1984, 164) identifies, the rejection by Popper (1968) of induction also includes the validity of "inductive logic". As such, this research will apply only deductive logic when subjecting Johnson-Laird's hypotheses to analysis and synthesis.

approach toward generating true hypotheses, their epistemology would be self-refuting. That is to say, as both Popper and Quine use philosophic analysis to produce their epistemologies, any field or individual which employs them must accept the validity of analysis and synthesis.

Consequently, any results produced by this study are considered to concern ontologically real entities, is naturalised within the field of cognitive science, is repeatable and subject to possible falsification by other researchers. Consequently, the method of analysis and synthesis is a valid approach when attempting to meet our research aim.

Now that philosophic analysis' validity has been demonstrated, its appropriateness now needs to be established. That is, it needs to be clarified why analysis and synthesis is the *correct* method to employ to meet our research aim.

§5 Design appropriateness

Firstly, the productivity of analysis and synthesis can be seen in an observation made by Brook (2009, p217), who discerns the considerable contributions the method has already made to the field of cognitive science, when it is stated:

> Philosophers have been a part of cognitive science since the activity was merely a twinkle in the eyes of a small but hardy group of pioneers in the 1960s. Hilary Putnam and Jerry Fodor come immediately to mind. In the 1960s, those two did much to articulate the view that came to be known as functionalism[...] Functionalism is something like the official philosophy of mind of cognitive science and has been from the beginning.

Moreover, when establishing whether this putative research strategy is the correct method to employ in meeting our research aim, it is informative to again look to that method's capability of establishing the logical success, or failure, of hypotheses by exposing and determining their logical validity. That is, as critical philosophical analysis will be able to uncover the claims and corollary assumptions that make up Johnson-Laird's referentialist semantics, as well as the interrelations between the two, the proposed method is capable of making the internal and external logical consistency of the hypothesis more perspicuous and, therefore, determining success. For, as we have seen above, critical philosophic analysis is able to ascertain the internal and external consistency of a theory, and as, under a Popperian epistemology, "[w]e regard incompatibility as falsification of the theory" (Popper 1968, p264), this is an appropriate method to follow when attempting to disprove an hypothesis.[15] As such,

---

[15] However, there is a caveat here, for "compatibility alone must not make us attribute to the theory a positive degree of corroboration: the mere fact that a theory has not yet been falsified can obviously not be regarded as sufficient" (Popper 1968, p264). Ergo, implementation of this method will not be able to confirm Johnson-Laird's hypothesis, only refute it. However, this is an inevitable conclusion of a Popperian

in regards to the appropriateness of this method design in this instance, it is interesting to note the historical successes of the method of philosophy within cognitive science and observe its similar potential in this instance.

Furthermore, it is important to note that Johnson-Laird's adoption of a correspondence theory of meaning was motivated by the need to save the RTM from the philosophical arguments of Searle (1980) and Lewis (1972); thereby saving the theory, and much of cognitive science, by providing it with a successful account of semantics that sidesteps the arguments of these philosophers (see Johnson-Laird 1988, p115; 1993, p34). As this study has tasked itself with testing the success of Johnson-Laird's move in this regard, it seems only appropriate that the same method employed by Searle and Lewis – i.e. the philosophical method of analysis and synthesis – be once more employed. This is a line of reasoning that Thagard (2009, p238) follows when demarcating the role of philosophy in cognitive science, when he contends that:

> Philosophy can be useful to cognitive science in providing defences against philosophical arguments challenging the core assumptions of cognitive science concerning representation and computation. In this way, philosophy can provide self-defence methods for cognitive scientists against philosophers critical of the whole field.

Finally, as we have already seen, the 'traditional' post-positivist methods (i.e. quantitative and qualitative approaches) can neither validly nor successfully answer the question with which this research concerns itself. As such, this state of affairs would suggest, via inference to the best explanation, that the critical philosophic analysis would be the most appropriate approach when investigating the introduction of a correspondence theory of meaning into mental model theory.

Consequently, it appears that for a number of reasons, the method of analysis and synthesis, if implemented correctly, is appropriate, i.e. is capable of generating data germane to the research aim of this research.

§6 Generalisability

Before this chapter reaches its conclusion, it is first necessary to define clearly the scope of its findings. As such, it is incumbent to highlight that although a number of referentialist semantics have been advanced within the RTM (e.g. Johnson-Laird 2002; Fodor 2010; Schweizer 2012 etc.), the nuance that

---

epistemology – in which facts can only be falsified, not verified – and is an unavoidable consequence for any study operating within the post-positivist paradigm.

exists between these positions means this study will not be able to generalise any of its findings to them all. As this research concerns itself only with the hypotheses and arguments of Johnson-Laird, it will be only to that work that any findings can be applied. Consequently, if this study finds the arguments of Johnson-Laird to be either consistent, or inconsistent, it will not be possible to generalise this result to any other hypothesis that posits a correspondence theory of meaning within the RTM.[16]

Conclusion

To thoroughly answer how this study intends to meet its research aim in the best possible manner, this chapter has posed and answered the following questions:

Q1. What will be done to answer the research aim?

A1. Critical philosophical analysis, i.e. decompositional analysis and synthesis, will be applied to the logic of the arguments and theories of Philip Johnson-Laird to establish the logical consistency of the inclusion of a Peircean structural correspondence theory of meaning, and its corollary assumptions, into the RTM.

Q2. How will this method be implemented?

A2. To identify the logical consistency of Johnson-Laird's arguments via critical philosophical analysis, his referentialist hypothesis will be analysed in regard to the following research questions:

i) Does Johnson-Laird's updated account of meaning circumvent the semantic objections of the RTM's commentators (i.e. the CRA, the SGP, holism and the homunculus and symbolic fallacies)?

ii) Is Peirce's structural correspondence theory of meaning both internally and externally consistent?

iii) Is the introduction of a correspondence theory of meaning into Johnson-Laird's broader philosophy of cognition logically consistent?

If Johnson-Laird's theories cannot answer in the affirmative to all three of the above questions, then they are not to be considered logically consistent and, following Popper and Quine, thereby falsified.

Q3. What justifies the implementation of this research design?

---

[16] This focus on the work of Johnson-Laird alone raises another issue with the validity of a quantitative research design from our discussion earlier in the chapter: A quantitative experiment would tell us only that semantic content operates via reference and, therefore, that *some* form of correspondence theory of meaning is accurate, it tells us nothing about Johnson-Laird's particular Peircean correspondence theory nor his theory of mental representation as a whole. Therefore, if we want to investigate the success of Johnson-Laird's arguments specifically, as we intend, quantitative methods will not give us the information we require.

A3. As this study will be attempting to establish the success, or otherwise, of Johnson-Laird's arguments under their own lights (i.e. is determined to produce results that tackle Johnson-Laird's updated semantics at face value, not at the lower, paradigmatic, philosophical level), it will be accepting their methodological framework (i.e. their epistemological and ontological assumptions) as given. Consequently, method design 1) validity and 2) appropriateness will, in this instance, be judged according to the standards of the post-positivist paradigm. Therefore:

1) As analysis and synthesis is deemed to be valid under the post-positivist paradigm, it, and this research, is therefore capable of producing results cumulative with cognitive science and Johnson-Laird's work. That the philosophical method is conducive to the post-positivist paradigm in this manner is seen as:

    a. It is a method with an historical precedence as a valid method in cognitive science research (see Putnam 1980; Brook 2009; Dennett 2009; Thagard 2009; Papineau 2014).

    b. It will be concerning itself with entities (mental representations and propositional attitudes) deemed to be, ultimately, objectively real (see Johnson-Laird 1993, p58; Waskan 2006, pp36-76).

    c. It is naturalised within the existing suite of hypotheses held by Johnson-Laird and adheres to both the continuum and 'no first philosophy' theses and is, therefore, continuous with cognitive science, i.e. it produces findings that adhere to a Quinean epistemology (see Quine 1960, p3; Colyvan 2001).

    d. It is a method that produces results that are both falsifiable and repeatable, i.e. analysis and synthesis adheres to a Popperian epistemology (see Popper 1968, p264; Popper and Miller 1983, p155; Finocchiaro 2005, pp255-256).

    e. Comments made by Johnson-Laird (1983, pxiii) seem to implicitly accept analysis and synthesis as a valid method.

    f. If post-positivism did not allow for analysis and synthesis, then its foundational theories would have to be abandoned. For, the arguments that established post-positivism (see Quine 1953; 1960; 1987; 1999; 2008; Popper 1957; 1968; 1974; 1994) employ critical philosophic analysis. Therefore, if the methodology were to reject analysis and synthesis as a valid research approach, the position would be self-refuting.

2) It was also found that analysis and synthesis would be the best method to apply in this instance. This was found to be the case as:

    a. Analysis and synthesis has previously produced results that have made a number of contributions to the field of cognitive science and, therefore, has the capability of doing so again (see Putnam 1980; Brook 2009; Fodor 2010).

b. The method of analysis and synthesis is capable of interrogating an hypothesis to uncover its composite theses and establish those theses' compatibility (Finocchiaro 2005, pp255-256). As such, and as "[w]e regard incompatibility as falsification of [a] theory" (Popper 1968, p264), the method can successfully achieve our research aim (see also *Ibid.*, p72).

c. The work that motivated Johnson-Laird's move to referentialist semantics (i.e. the CRA, the SGP, holism and the symbolic and homunculus fallacies) derive from critical philosophical analysis. Therefore, it appears appropriate to apply the same method when investigating if these arguments have been subverted and referentialism is a successful account of meaning (see Thagard 2009, p238).

d. The 'traditional' post-positivist approaches (i.e. quantitative and qualitative methods, such as experimentation, observation and surveys (quantitative methods) or interviews, ethnography, case studies and phenomenology (qualitative methods)) are unsuitable given the subject matter under consideration here (i.e. meaning and cognition). Therefore, inference to the best explanation leads us to the conclusion that critical philosophical analysis would be the best method to apply under the circumstances.

Findings

Chapter Four:
Findings

Introductory remarks

Following on from our Method chapter, we shall now attempt to establish the validity, or otherwise, of introducing a Peircean referentialist semantics into Johnson-Laird's analogical RTM. To recap, we shall try to achieve this aim via the use of critical philosophical, i.e. decompositional, analysis and synthesis to establish the logical consistency of this theory. Once this process of deconstruction (analysis) and reconstruction (synthesis) has been completed, any logical conflict present in Johnson-Laird's thesis should be more readily identifiable. If this process shows the arguments of Johnson-Laird to be logically inconsistent, then his referentialist theory will be said to not accord with his philosophy of mind as a whole and, consequently, his account of semantics will be claimed to have been falsified and its rejection under both the Quinean and the Popperian epistemologies will be recommended. If, however, no dissonance is found by this investigation, it will be said that the hypothesis remains valid and will, therefore, be labelled a "provisional conjecture" (Popper 1968, p264). From our Method chapter, it was determined that the best way to complete this proposed critical analysis and synthesis was by subjecting Johnson-Laird's theory to the following three research questions:

i) Does Johnson-Laird's updated account of meaning circumvent the semantic objections of the RTM's commentators (i.e., the CRA, the SGP, holism and the homunculus and symbolic fallacies)?

ii) Is Peirce's structural correspondence theory of meaning both internally and externally consistent?

iii) Is the introduction of a correspondence theory of meaning into Johnson-Laird's broader philosophy of cognition logically consistent?

Consequently, this chapter will follow this division and will be structured into three parts, each focused on critically analysing the works of Philip Johnson-Laird in relation to each research question.

§1 Does Johnson-Laird's updated account of meaning circumvent the semantic objections of the RTM's commentators?

The first of our research questions was chosen as, as outlined within our literature review, a major motivating factor in Johnson-Laird's adoption of a correspondence theory of meaning were the semantic arguments encapsulated by the CRA and the SGB (see, for example, Johnson-Laird 1988, p107; 1993, p34). Consequently, if such a referentialist account of meaning fails to circumvent these semantic arguments, as Johnson-Laird postulates (see also Harnard 1990b; Fodor 2010; Rapaport 2011; Schweizer 2012), then this philosophy of mind will have failed at the first hurdle and, as such, must not be considered a valid hypothesis. To determine an answer to this question, this section will:

1) Identify a strategy that may circumvent these arguments.

2) Determine if Johnson-Laird's referentialist semantics adheres to this strategy.

3) Measure Johnson-Laird's account of meaning against a standard set by Searle (1980, p417).

Language entry and exit rules

From the discussion of the CRA and the SGP found within our literature review, it should be remembered that the main import of those arguments was that language is not self-supporting, i.e. that relating language only to itself cannot provide a successful account of meaning. Thus, it can be inferred that what is needed to avoid such criticisms are rules for supporting language, for connecting it to some other, non-linguistic, phenomena and that (in the language of Harnard (1990b, p339)) are able to get us off of the language "roundabout". Such proposed procedures as these are labelled by Putnam (1981, p11) as "language entry" and "language exit" rules, rules Putnam (1988, p47) defines as follows:

> Language [has…] "language-entry rules" (think of these as rules saying that when the speaker has certain experiences, he is to put certain sentences in the "belief box)[…] and "language-exit rules" (rules saying that when the speaker has certain sentences in the belief and desire boxes, he is to perform certain bodily movements, or saying certain words, etc.).[1]

This means that, without the procedures and sensory input needed to communicate the real, i.e. non-linguistic, world to the mental world of an individual, and *vice versa*, language is unsupported and must collapse in on itself: Without language entry and language exit rules, linguistic signs can only remain empty. Therefore, for such arguments as the CRA and SGP to be circumvented, these language entry and language exit rules must be present within a theory and it is exactly their absence that led RTM into difficulties (see also Dennett 1980, p429, who similarly identifies the lack of such rules operating in the CRA).

---

[1] See also Sellers (1963, p314) and Schweizer (2012, p195).

Peircean iconism and language entry and exit rules

With Putnam's rules in mind, it seems that, for Johnson-Laird's referentialist strategy to counter the criticisms of the CRA and the SGP effectively, it must be shown that the introduction of Peircean icons into his philosophy of mind successfully produces language entry and language exit rules. That iconic mental representations achieve this goal can be seen from the earlier explication of Johnson-Laird's updated mental models found in our literature review. From that discussion it was discovered that mental models are now derived from the translation of sensory, non-linguistic, stimuli. That is to say, "that linguistic representations of the meaning of propositions are used to construct mental models of the situation under description" (Johnson-Laird 2002, p80).[2] Furthermore, once a model has been derived from an individual's interaction with the world, "this representation of the world is used by [an individual] as a guide to action" (*Ibid.* 1993, p275).[3] Consequently, it can be seen that, within Johnson-Laird's Peircean mental model theory of cognition, there exist effective procedures that simulate language entry rules (i.e. that are capable of converting perceptual experiences of the world into mental entities within a mental 'belief box') and language exit rules (i.e. for converting mental entities within a mental 'belief and desire box' into action). As such, it follows from this discussion that Johnson-Laird is able to circumvent the semantic arguments from certain commentators. That this latter point is accurate can be seen in the fact that mental models, under Peircean iconism, avoid: (i) The symbolic fallacy; (ii) the homunculus fallacy; and (iii) semantic holism. This is the case because:

(i) Mental models now derive their semantic content from the world, not from other mental signs. Consequently, meaning is no longer merely a process of replacing one set of symbols with another and, therefore, does not commit the symbolic fallacy. That is, for semantics, Johnson-Laird uses icons to successfully introduce language entry and language exit rules into his philosophy and to get off of the 'language roundabout'.

(ii) As meaning is now said to derive from the world, it is, in the language of Harnard (1990b, p338), "extrinsic" to mental representations and "intrinsic" to a state of affairs. As such, for Johnson-Laird, semantic content is now no longer parasitic upon conscious agents, semantic content has instead been grounded and, as such, requires no interpretation by any homunculus.[4]

(iii) The fact that a mental model is now related to a physical state of affairs and derives its meaning from that part of the world, indicates that the semantic content of Johnson-Laird's

---

[2] See also Johnson-Laird (1980, p100; 1988, p15; 1993, p275; 2011, p479.)

[3] See also Johnson-Laird (1977, pp206-207; 2005, p185).

[4] Interestingly, it seems that the positing of an homunculus into syntactical accounts of meaning, from objections such as the CRA and SGP, is an intuitive method of accounting for language entry and exit rules within a system that is missing them.

mental models are no longer dependent upon their relations to other mental representations. As such, the language rules Johnson-Laird inherits from Peirce's iconic theory of signs abolish the need for the interdependence of mental representations and, therefore, he avoids the difficulties of semantic holism.[5] This last argument can be better seen with reference to the definition of semantic holism from Roland (2014. pp45), i.e.:

> Typically only sufficiently rich collections of statements, as opposed to individual statements, have empirical content.

However, with iconism, it can be seen that this is not the position in which we find ourselves. That is, with a congruent form of correspondence, a statement can have empirical content on its own, divorced from any collection of other statements. Where iconism is concerned, if the mental representation of a statement shares the relation structure of a state of affairs, the empirical content of that mental representation *is* that state of affairs, regardless of the connections between the world and any other mental representation. Consequently, semantic content under correspondence by congruence is atomistic: Only the relation between a mental representation and the world determines content and the suite of other representations within which the mental representation is positioned plays no role.[6]

Nevertheless, although the above analysis seems to vindicate the referentialist approach of Johnson-Laird (and others), there is a deficiency with this solution. Searle states that the issue with the RTM is that, "Instantiating a computer program is never by itself a sufficient condition of intentionality" (Searle 1980, p417) and Searle-in-the-room, within the CRA, fails to achieve understanding as "he doesn't know that the story refers to restaurants and hamburgers, etc." (Searle *Ibid.*, p419). This is an issue Johnson-Laird attempts to resolve via the stipulation of mental signs having reference to the world. However, the question of how this mental reference *itself* refers may be raised. This objection is best seen in a point made by Putnam (1977, pp486-487) concerning the causal account of reference fixation:

> how 'causes' can uniquely refer is as much of a puzzle as how 'cat' can [...] The problem, in a way, is traceable back to Ockham. Ockham introduced the idea that concepts are (mental) particulars. If concepts are particulars ("signs"), then any concept we may have of the relation between a sign and its object is another sign. But it is unintelligible, from my point of view, how the sort of relation the

---

[5] It is interesting to speculate whether all successful forms of language entry and exit rules, *eo ipso*, circumvent semantic holism, or just the referentialist type employed by Johnson-Laird. However, the interdependence between the meaning of mental entities that is the hallmark of semantic holism, does, at first blush, appear to be a further symptom of language trying to support itself. Suggesting that successful language entry and language exit rules, *eo ipso*, avoid the homunculus and symbolic fallacies *and* semantic holism.

[6] Again, some may argue that this leads us into certain metaphilosophical problems. For, certain readings of Quine, e.g., Roland (2014, pp56-57), identify that semantic holism is a necessary hypothesis for Quinean naturalism. Consequently, if such a stance is taken, this investigation may be found to contain an inconsistent methodology. However, it should be repeated that this thesis is taking the position that only theoretical, not semantic, holism is necessary under Quinean naturalism.

metaphysical realist envisages as holding between a sign and its object can be singled out either by holding up the sign itself, thus

```
┌─────────────────────┐
│                     │
│         COW         │
│                     │
└─────────────────────┘
```

-- or by holding up yet another sign, thus

```
┌─────────────────────┐
│                     │
│        REFERS       │
│                     │
└─────────────────────┘
```

-- or perhaps --

```
┌─────────────────────┐
│                     │
│        CAUSES       │
│                     │
└─────────────────────┘
```

That is to say, if Johnson-Laird is solving the CRA by merely positing 'reference' and 'refers' or 'correspondence' and 'corresponds' as some sort of mental particular that is relating a mental sign to its referent, then we are either led into some variation of Bradley's regress (cf. Bradley 1916), or 'refers' and 'corresponds' are as much syntactical signs as mental representations themselves and the CRA holds. Or, as Putnam (*Ibid.*, p488) concludes:

> The point is that[...] you can't treat understanding a sentence (in general) as knowing its truth conditions; because it then becomes unintelligible what that knowledge in turn consists in.

However, this analysis fails when applied to the hypotheses of Johnson-Laird, for within that account of reference, the question 'what does the knowledge of knowing an object's truth conditions consist in?' is answered. Since Johnson-Laird follows Peirce and employs a congruent correspondence theory of meaning, knowing an object's truth conditions consists in an awareness of that state of affair's logical form or, using the language of Craik (1967), its 'relation-structure'. Consequently, reference occurs through the accurate representation of that logical form or relation-structure. Therefore, Johnson-Laird's positing of 'reference', 'refers', 'correspondence' and 'corresponds' is not akin to Ockham's mental particulars and Putnam's objection misses the mark.[7] Furthermore, this approach also avoids Putnam's other 'problem of reference', i.e.:

> reference may be metaphysically singled out without being totally *determinate* (the metaphysically singled-out *R* may allow for a plurality of admissible interpretations)[.] (Putnam 1981, p48)

---

[7] Interestingly, Putnam's analysis is only appropriate for correspondence theories of meaning that are correlational (such as is found within Fodor 2010 and Rapaport 2007; 2011), a type of correspondence Putnam's argument seems to presuppose.

That is, it could be argued that reference may be stipulated by non-determinate definitions, making it unknown to what our mental representations actually refer, i.e. reference may be indeterminate. For example, we may have an interpretation of 'A cat is on a mat' such that:

  (a) A cat* is on a mat*

The definition of the property of being a cat* (respectively, a mat*) is given by cases, the three cases being:

  (a) Some cat is on some mat; and some cherry is on some tree.

  (b) Some cat is on some mat, and no cherry is on any tree.

  (c) Neither of the foregoing

Here is the definition of the two properties:

DEFINITION OF 'CAT*'

x is a cat* if and only if case (a) holds and x is a cherry; or case (b) holds and x ix a cat; or case (c) holds and x ix a cherry.

DEFINITION OF 'MAT'*

x is a mat* if and only if case (a) holds and x is a tree; or case (b) holds and x ix a mat; or case (c) holds and x ix a quark. (*Ibid.*, p34)

In other words, as it is not determinate whether a particular speech act is an instance of 'A cat is on a mat' or 'A cat* is on a mat*', we cannot be certain that our talk of cats and mats, and mental representations of these states of affairs, is about cats and mats or about cat*s and mat*s (i.e. cherries and trees). Moreover, as Putnam (*Ibid.*, pp217-218) shows, this account of reference leads to indeterminacy of whether 'cats' and 'mats' refer to anything in the universe. However, this argument again presupposes a correlational and not a structural theory of correspondence, for it is only under this view that such definitions can be stipulated. Under the iconic (i.e., congruent) view, it is logical form that determines reference, so such logical connectives as 'or' cannot be included and the grounds for Putnam's argument falls away when considering Johnson-Laird's hypothesis.

Consequently, if we follow the above analysis, we must conclude that the thrust of the CRA is circumvented by Johnson-laird's referentialist move. Moreover, we are capable of using this analysis to diagnose why it is circumvented: From following the arguments of Johnson-Laird, it can be seen that the CRA's objections to the RTM are not successful as they are not an accurate representation of how language comprehension and production is said to operate.[8] As Johnson-Laird (among others) shows, Searle's thought experiment is lacking a way to structurally ground the Chinese symbols manipulated within his room, i.e. the Chinese Searle-in-the-room operates corresponds only with the symbols in his book, not with the world. Nevertheless, as Searle (1980, p417) observes, "one way to

---

[8] This is a standard rebuttal to Searle's argument and can be found in the likes of Dennett (1980), Block (2002) and Schweizer (2012).

test any theory of the mind is to ask oneself what it would be like if my mind actually worked on the principles that the theory says all minds work on". Therefore, before we can conclude that Johnson-Laird's RTM avoids such semantic criticisms with any confidence, it seems incumbent upon us to take up Searle's challenge here and apply his test to Johnson-Laird's account of cognition and generate our own Chinese Room-style thought experiment, a thought experiment that is produced through our doctoring of the original CRA (cf. *Ibid.*, pp417-418).

CRA 2.0

Suppose that Searle is locked in a room and given a large batch of Chinese writing. Suppose furthermore (as is indeed the case) that he knows no Chinese, either written or spoken, and that he is not even confident that he could recognise Chinese writing as Chinese writing distinct from, say, Japanese writing or meaningless squiggles. To Searle, Chinese writing is just so many meaningless squiggles. Now suppose further that after this first batch of Chinese writing, Searle is given a second batch of Chinese script together with a set of rules for translating both batches into a recognisable (to Searle) representation of a worldly state of affairs, as well as instructions to correlate the second batch's representation with the first batch's.[9] The rules are in English, and Searle understands these rules as well as any other native speaker of English. Now suppose also that he is given a third batch of Chinese symbols together with some instructions, again in English, that enable him to translate these signs into a further recognisable representation and correlate elements of this third batch's representation with those of the first two batches'. Also, these rules instruct Searle how to translate the representations back into certain Chinese symbols. Unknown to Searle, the people who are giving him all of these symbols call the first batch "a script", they call the second batch a "story" and they call the third batch "questions". Furthermore, they call the symbols he gives them back in response to the third batch "answers to the questions" and the set of rules in English that they give him, they call "the program". Now, just to complicate the story a little, imagine that these people also give him stories in English, which Searle understands without any translation, and they then ask him questions in English about these stories, and he gives them back answers in English. Suppose also that after a while he gets so good at following the instructions for translating the Chinese symbols and manipulating the representations, and the programmers get so good at writing the programs, that from the external point of view – that is, from the point of view of somebody outside the room in which Searle is locked – his answers to the questions are absolutely indistinguishable from those of native Chinese speakers. Nobody just looking at his answers can tell that he did not speak a word of Chinese before entering the room. Let us also suppose that his answers to the English questions are,

---

[9] For this experiment's purposes, the representations' exact pictorial form does not matter, just so long as the pictorial form accurately portrays the logical form of the batches to Searle-in-the-room (i.e. are in a manner he can interpret from his existing experience of the world (e.g. a picture or 3D model, say)), then any pictorial form will suffice.

as they no doubt would be, indistinguishable from those of other native English speakers, for the simple reason that Searle is a native English speaker. From the external point of view – from the point of view of someone reading Searle's answers – the answers to the Chinese questions and the English questions are equally good. But in the Chinese case, unlike the English case, he produces the answers by externally/physically manipulating representations whose logical form he can identify and that have been formally translated from Chinese symbols. As far as the Chinese is concerned, Searle simply behaves like a successfully grounded computer; he performs computational operations on grounded elements. For the purposes of the Chinese, Searle is simply an instantiation of the computer program.

Now, the claims made by strong AI, the RTM and Johnson-Laird are that such a grounded computer understands the stories and that the program in some sense explains human understanding. But we are now in a position to examine these claims in light of our thought experiment:

1. As regards the first claim, it seems to us quite obvious in the example that Searle-in-the-room only understands the Chinese stories once they have been translated into a representation whose logical form he can discern and that refers to a state of affairs with which Searle-in-the-room is familiar. From these representations, and the updates made upon them, Searle-in-the-room is made aware of the content of the stories, how the procedures from the script are updating their content and the outcomes of the investigation commands from the questions. Consequently, it can be said that Searle-in-the-room, in some manner, is understanding the Chinese he is given and which he is producing. For the same reasons, if a story interpreting computer were grounded in the same manner, it would also understand the stories, whether in Chinese, English, or whatever, since in the Chinese case the computer is Searle-in-the-room, and in cases where the computer is not Searle-in-the-room, the computer still has access to the grounded representations.

2. As regards the second claim, that the program explains human understanding, we can see that the computer and its program provide sufficient conditions for understanding, since the computer and the program are functioning, and there is understanding.

However, there is an issue here, for, in the above thought experiment, Searle-in-the-room's ability to ground the representations he translates and produces are predicated upon his prior experience with the world, i.e. his own mental models produced prior to his entry into the room. Therefore, it is not an accurate account of cognition as is found in humans and as it would be implemented in a machine to achieve strong AI. If we were to imagine instead an individual to be born within the room, then any

representations produced by the program will not be grounded to that person and the conclusions found within Searle's original article will be recreated. For, as Dennett (1980, p429) argues:

> it has long been a familiar theme within AI circles that [some] programs – I call them bedridden programs since their only modes of perception and action are linguistic – tackle at best a severe truncation of the interesting task of modeling real understanding. Such programs exhibit no "language-entry" and "language-exit" transitions, to use Wilfrid Sellars's terms, and have no capacity for non--linguistic perception or bodily action. The shortcomings of such models have been widely recognized for years in AI […] A computer whose only input and output was verbal would always be blind to the meaning of what was written.

As such, what is required here is a further adaptation of this thought experiment, so as to capture the language and exit rules whose importance were discussed earlier, thereby preventing 'bedridden' AI and providing the appropriate input necessary for those born in the room to autonomously ground the symbols present in the room's books. That is, to enable an individual born into the room to create, in the language of McGinn (1989, p155), "*original*" content. Consequently, the updated thought experiment will look something as follows (this thought experiment is again based on an adulteration of Searle's own words (cf. *Ibid.*, p420)).

CRA 2.1

Suppose we wrote a different kind of program from the above 'bed-ridden' program. Suppose an individual were to be born inside a robot, and this person would not just take in formal symbols as input and give out formal symbols as output, but would rather actually operate the robot in such a way that the robot does something very much like perceiving, walking, moving about, hammering nails, eating, drinking – anything you like. The robot would, for example, have a television camera attached to it that enabled it to 'see', it would have arms and legs that enabled it to 'act', and all of this would be controlled by its human 'brain'. Such a robot would, akin to the above computer, have genuine understanding and other mental states in that it could autonomously generate and ground its own representations.[10] The input from the robot's sensors (whatever its format) would constitute the base data that the rule book will translate into representations. However, an update needs to be made here, in the beginning all that would be present in the robot would be a largely empty rule book that simulates the innate capacity of the brain, and which translates sensory input into some standardised representational format, retranslate representations into commands for behaviour and that simulate the language learning processes of an infant (how these complex language learning processes operate, beyond it being some form of reference fixation, are far beyond the scope of this

---

[10] An interesting point of discussion is the effect the nature of the input (i.e. binary, script, electrical impulses etc.) has on the pictorial form of the representations produced and the role this nature has on determining the qualitative worldview of the person-in-the-robot. This raises the question: What is it like to be a robot? (cf. Nagel 1974).

argument and this thesis, therefore that it can be accurately simulated must be charitably assumed for the sake of the argument). Now, given that the person-born-in-the-robot is somehow able to understand and employ the rule book, it can be used so as to produce representations from the sensory data as well as to correlate those representations with the Chinese symbols from the language input (using the simulated language learning process rules from the rule book), thereby autonomously populating the rulebook with the information provided to Searle in the above thought experiment. Once this has been completed, the state of affairs found between Searle-in-the-room and person-in-the-robot are equivalent and the conclusions found from the earlier example follow.[11]

Summary

The above philosophical analysis has shown that the root issue operating behind the CRA and the SGP was the concept of language 'entry and exit rules'. That is, the semantic arguments levied against the RTM by its detractors demonstrated that semantic content without intentionality, or bedridden AI, (i.e., any system that only relates symbols with other symbols) fails to provide these entry and exit rules and are therefore inextricably led into a series of difficulties (e.g. the symbolic and homunculus fallacies). Furthermore, it was found that the language entry and language exit rules employed by Johnson-Laird's use of Peirce were of such a manner (i.e. a one-to-one relationship between mental representation and worldly state of affairs and sets of state of affairs) that they circumvent semantic holism and the issues that it engenders. As such, the above analysis and syntheses demonstrated that via an iconic, i.e. congruent, correspondence theory of meaning, these language entry and exit rules are now accounted for within Johnson-Laird's philosophy of mind and these semantic arguments are successfully avoided. As such, Johnson-Laird's analysis appears to be correct; his theory of cognition remains a valid supposition when scrutinised from this perspective; and the answer to the research question that concerns this section is a 'yes'.

§2 Is Peirce's structural correspondence theory of meaning both internally and externally consistent?

To achieve our stated goal, this chapter will continue by applying the 'tools' of philosophy to Johnson-Laird's Peircean correspondence theory of meaning and attempt to demonstrate the internal and external consistency of such a theory. To that end, this section will investigate the logic of the iconic theory in its own right as well as the eligibility of its introduction into Johnson-Laird's pre-existing suite

---

[11] Nevertheless, there is one final note here. In the above example, person-born-in-the-robot is still a homunculus operating the switches and simulating the processor of the computer, not the computer itself. As such, a combination of this version of the robot reply and that known as the systems reply – i.e., a CRA 2.2 – needs to be created, so as to provide an accurate picture of the true state of affairs found in fully embodied human language learning, understanding and production and which would be instantiated in a machine to produce strong AI.

of hypotheses. Consequently, we shall proceed by attempting to establish whether there exists any dissonance between the theses that comprise Peirce's semantics, as well as exploring whether there are any inconsistencies between any of the hypotheses that constitute Peirce's referentialist semantics and those of Johnson-Laird's wider account of cognition. As such, this section will be divided into two sections: One that looks at the external consistency of Peircean iconicity in regard to Johnson-Laird's philosophy of mind and another that surveys its internal consistency, independent from the work of Johnson-Laird.

External consistency

The first point to make when engaged in the task of establishing the external validity of the inclusion of a Peircean correspondence theory of meaning within Johnson-Laird's mental model theory, is that the foundational theses of analogue representations within the RTM are unaffected, i.e. Johnson-Laird's adherence to the RTM coheres equally with iconic Peircean mental representations as with analogous Craikean ones. This is the case as the operation of iconic representations still relies on a finite set of representations that are processed by a finite number of computational operations to produce such things as cognition, perception, reasoning, language, action etc. That is to say, a Peircean correspondence theory of meaning is consistent with the following central hypotheses of Johnson-Laird's account of cognition:

> **Folk psychology:** Since iconic models allow that propositional attitudes, such as beliefs and desires, are both ontologically real and are sufficient explanations for the causes of action, their inclusion creates no tension in Johnson-Laird's hypotheses.

> **Computation:** Since iconic models are as capable of being processed by computational operations as any other analogical representation, they cohere with a computational explanation of cognition.

> **Functionalism:** Since iconic models can be instantiated by any number of physical systems and materials that are capable of functioning in such a manner (e.g. from paintings, sculptures and patterns of neuron firings), they are consistent with the functionalist position.

> **Marr's tri-level hypothesis:** As iconic representations occupy as valid a position at the second – representational – level of Marr's hypothesis as any other analogous representation, the inclusion of a Peircean correspondence theory creates no tension.

Nevertheless, although the above is the case and the inclusion of a Peircean correspondence theory of meaning is consistent with the foundations of Johnson-Laird's RTM, there is one area that the inclusion of such a theory may create tension: Johnson-Laird's adherence to the strong AI hypothesis (see Johnson-Laird 1983; 1993). For, if the strong AI hypothesis is taken to be defined as the position

that computational programs are sufficient for (machine) intelligence (as is expressed in Searle 2002, p51, and seems to be expressed in Searle 1980, p417), then Peircean mental models cannot be considered as a form of such an AI as it dictates that iconic representations of worldly states of affairs are also necessary for mentality (i.e. he states that only a computational program *plus* the iconic grounding of signs is sufficient for (machine) intelligence) and there may be some tension in Johnson-Laird's position on AI. However, for tension to be created here, Johnson-Laird will have had to not only have made an explicit commitment to this reading of strong AI, but also to have explicitly stated his continued commitment to it, neither of which he has done. Furthermore, this reading of strong AI is one that has been rejected within cognitive science and many working in AI since the concept's inception (see Fodor 1980, p431), bringing into question Johnson-Laird ever being committed to such a position. Finally, if the strong AI hypothesis is defined as the position that programs alone are necessary for (machine) intelligence (as is taken up by Rey 2002, p202, and which seems to be expressed in Searle 1980, p418) then Johnson-Laird's position on AI can still be defined as a strong one and no possibility of a conflict arises.

However, there do appear to be a number of genuine inconsistencies between the theories of cognition put forward by Johnson-Laird and by Peirce. For example, Johnson-Laird (1980; 2008; 2010) is unequivocal in his view that mental images do not constitute the base level of cognition and reasoning, whereas Peirce argues that this is, in fact, the case (see §3.363; §3.556; §3.560; §3.613; §5.148; §5.162) – a view of Peirce's not touched upon by Johnson-Laird. This dissonance is problematic as a tension is now present between the theories of the two authors, and, with it, confusion over what is and what is not taken from Peirce's works – state of affairs that, at best, renders Johnson-Laird's own adopted theory nebulous and ill defined. Additionally, and as Vosgerau (2006) observes, Peirce's icons rely on a "logical algebra[…] contain[ing] signs [including…] indices[…] used for expressing negation" (cf. Peirce §§3.385-3.386), an hypothesis Johnson-Laird (2002, p85) explicitly accepts, averring that "Like graphs, mental models therefore use a symbol to designate negation" (see also Barres and Johnson-Laird 2003; Johnson-Laird 2006; Johnson-Laird and Khemlani 2013). However, and as seen in our literature review, Johnson-Laird (1980, p125) has previously defended an important property of mental models that:

> models have an important structural property deriving from a constraint on the set of possible mental models: a *natural* mental model of discourse has a structure that corresponds directly to the structure of the state of affairs that the discourse describe.

This naturalism constraint, labelled by Vosgerau (2006, p255) as "structural preservation and naturalness", has historically limited the scope of mental models to include only those things found in the world, generally, and those things in the state of affairs a model denotes, specifically. As such, it

can easily be seen that the adoption of Peirce's mental signs and indices for such things as negation (and "other abstract concepts" (Johnson-Laird and Khemlani 2013, p11)) create conflict with Johnson-Lairds naturalism constraint.[12]

Internal consistency

As Peirce's iconic theory of signs is comprised from many individual hypotheses, all of which interact with many of its others in numerous, nuanced ways, this subsection will not attempt to list all these interactions and demonstrate their consistency or otherwise. For our purposes, only those areas where inconsistency is determined to have been found will be explored. As such, we will continue with a presentation of a number of posited internal inconsistencies in Peircean iconism, along with a defence of this criticism.

Interpretant

The first internal inconsistency of a Peircean semantics has to do with its application of a concept Peirce labels 'Interpretant'. As we have seen (from our literature review), Peirce (§§274-284) identified icons as a mode of signs, but what has not been clarified is that, according to Peirce (*Ibid*.), something can only be considered a sign if it (the symbol, that which Peirce terms a 'First') is in such a relationship to a worldly state of affairs (referred to as a 'Second') that it causes an interpreter (known as a 'Third' or 'Interpretant') to view the symbol as they do the state of affairs.[13] This is what is meant when Peirce (*Ibid.*) states:

> A sign[…] is a First which stands in such a genuine triadic relation to a Second, called its Object, as to be capable of determining a Third, called its Interpretant, to assume the same triadic relation to its Object in which it stands itself to the same Object.

This reliance upon an Interpretant within a Peircean correspondence theory is where issues start to arise for Johnson-Laird, issues that remain no matter your reading of 'Interpretant'.[14] Classically there are two readings of Peirce in this regard (see Eco, 1976): 1) A mental interpretation; and 2) a non-mental reading. Unfortunately for Johnson-Laird, both positions result in difficulties:

---

[12] In fact, some, such as McGinn (1989, pp197-198), argue that for mental models to validly solve Brentano's problem, they need to adhere to the naturalism constraint. Therefore, Johnson-Laird's introduction of mental signs into his mental models may not only be externally inconsistent, it may also be internally inconsistent in that it leads him into difficulties explaining intentionality.

[13] This dependence upon a First, Second and Third is true in all cases of Peirce's analysis of signs. As previously mentioned, Peirce (1906, pp496-497) has a number of different analysis of signs and their instances, but in all cases "three things are concerned in the functioning of a Sign; the Sign itself, its Object, and its Interpretant".

[14] It should be remembered that the main source for our understanding of Peirce comes from his unpublished notebooks, which do not provide a consistent, unified theory. This is a state of affairs that results in the ambiguity over the term 'Interpretant' alluded to here and that has led Eco (1976, p1457) to aver "it is rather difficult to find two separate passages on a same topic in which he [Peirce] does not contradict and re-propose what he has said previously[…] A Peircean student is more entitled to say "Peirce said X on the day Y" than "Peirce said that…"".

(i) If you hold a necessarily mental reading of 'Interpretant', such as can be found in §1.339 and that is taken up by the likes of Burks (1949), where an Interpretant must be capable of consciousness and the possession of ideas, then Peirce's (and, by implication, Johnson-Laird's) icons can only be considered as such in the presence of a conscious Interpretant, i.e. to an homunculus. To clarify, under this reading of Peirce, he (and Johnson-Laird) is explaining the meaning of language by converting it into mental icons, but these mental icons can then only be meaningful in the presence of another mind, a mental Interpretant, leading to the homunculus fallacy and, therefore, an infinite regress.

(ii) If you take a non-mental view of Interpretant, i.e. one that relies only on a 'Quasi-Mind' (see Peirce 1906, p523), such as can be found in §2.228 and that is taken up by the likes of Mounce (1997), wherein an Interpretant may be another sign, then problems also arise.[15] For, this is an account of meaning that is based solely on replacing one symbol for another and, therefore, introduces the symbolic fallacy (a fallacy that was, it should be remembered, a motivating factor in Johnson-Laird's adoption of iconic mental models (see Johnson-Laird 1988)). Additionally, this leads to a further external inconsistency for Johnson-Laird when he imports Peirce's theory, for, as Mounce (*Ibid.*, p27) forcefully argues, this non-mental Interpretant leads to a public view of language (as a "private language is precisely one in which there is no first sign"), a view of language that is fundamentally opposed to the RTM's necessarily private view.[16]

Consequently, Johnson-Laird's employment of Peirce's theory of signs and icons necessitates an employment of Peirce's corollary concept of an Interpretant, but this leads him into certain problematic conclusions, some of which Johnson-Laird explicitly states are fallacious and need to be avoided.

---

[15] A 'quasi-mind' is clarified by Peirce (1906, p523) thus:

Thought is not necessarily connected with a brain. It appears in the work of bees, of crystals, and throughout the purely physical world; and one can no more deny that it is really there, than that the colors, the shapes, etc. of objects are really there. Consistently adhere to that unwarrantable denial, and you will be driven to some form of idealistic nominalism akin to Fichte's. Not only is thought in the organic world, but it develops there. But as there cannot be a General without Instances embodying it, so there cannot be thought without Signs. We must here give "Sign" a very wide sense, no doubt, but not too wide a sense to come within our definition. Admitting that connected Signs must have a Quasi-mind, it may further be declared that there can be no isolated sign. Moreover, signs require at least two Quasi-minds; a *Quasi-utterer* and a *Quasi-interpreter*; and although these two are at one (i.e. *are* one mind) in the sign itself, they must nevertheless be distinct. In the Sign they are, so to say, *welded*. Accordingly, it is not merely a fact of human Psychology, but a necessity of Logic, that every logical evolution of thought should be dialogic. You may say that all this is loose talk; and I admit that, as it stands, it has a large infusion of arbitrariness. It might be filled out with argument so as to remove the greater part of this fault; but in the first place, such an expansion would require a volume, --- and an uninviting one; and in the second place, what I have been saying is only to be applied to a slight determination of our system of diagrammatization, which it will only slightly affect; so that, should it be incorrect, the utmost *certain* effect will be a danger that our system *may* not represent every variety of non-human thought.

[16] Recall that within any form of RTM, semantics is necessarily founded upon private correlations between mental elements and language; mental elements that are only accessible, and known to and by, their authors (see Fodor 1975, pp68-69).

Negation

Another inconsistency within Johnson-Laird's Peircean correspondence theory of meaning is the latter's argument that certain concepts can only be signified through mental mathematical indices, i.e. through mental signs (see §§3.385-3.386). This is a view that has led Johnson-Laird (2002, pp84-85) to aver that:

> mental models[…] use a symbol to designate negation.[17]

There are a number of issues with this move (further to its flouting of the naturalism constraint). Firstly, if it is said that meaning is only derived when a sign stands in an iconic relationship with a state of affairs, the following question is raised: How can a mental symbol be meaningful and yet not be grounded, i.e., even though it bears no connection to any part of the world? Furthermore, such ungrounded, un-iconic symbols seem to introduce both the homunculus and symbolic fallacies. This is the case as the inclusion of a purely mental 'sign' only explains one set of symbols by replacing them with another, i.e. replaces the words 'not the case' with a mental '¬'. This can be seen when comparing comments made by Peirce with Johnson-Laird's own: At §4.127, Peirce asserts that "the conception of a "meaning", which is, in its primary acceptation, the translation of a sign into another system of signs", which is a view that aligns remarkably well with the accusation, levelled by Johnson-Laird (1993, pp333-334), that certain theories of meaning "can tell you that two words are related, or that one sentence is a paraphrase of another, but they are as circular as dictionaries" (see also *Ibid.* 1988). Alternatively, this argument can be interpreted as a failure of Peircean iconism to provide language entry and language exit rules for the mental symbol of negation (and certain 'other abstract concepts'). That is to say, Peirce's use of purely mental symbols leads him (and, by extension, Johnson-Laird) into the – now familiar – symbolic and homunculus fallacies, as it reintroduces the very properties Johnson-Laird diagnosed as the cause of the RTM's difficulties concerning semantic content.

Finally, the use of this account of negation to explain the meaning of counterfactuals (as is done by Barres and Johnson-Laird 2003, p2. see also Johnson-Laird 2002; Khemlani, Orenes & Johnson-Laird 2012) seems insufficient. For in such counter-factual cases as 'there is a car in this room', negation of the true state of affairs (i.e. this room without a car) does not fully capture the meaning of the proposition, as the resulting model will be identical to the, equally false, proposition, 'there is a dinosaur in this room', or many other, similar expressions. This leads, by *reductio ad absurdum*, to

---

[17] It needs to be clarified that this tactic of the inclusion of mental symbols to iconic representations is also applied in the case of "other abstract concepts" (Johnson-Laird and Khemlani 2013, p11), not just negation as discussed here (e.g., "The [model] theory allows that models can be tagged with numerals denoting their probabilities" (*Ibid*, p23)). In the case of these other 'abstract concepts' the following arguments, here concerning negation, are equally applicable.

many sentences which seemingly possess differing meanings being, in fact, the same proposition. Moreover, the negative, yet true, statement, 'there is no car in this room', can only be modelled by the mental representation of the affirmative, yet false, proposition of there being a car in the room with the attachment of the symbol for negation (see *Ibid.*). Therefore, it seems Johnson-Laird's account of negation is dependent on the meaningful representation of counterfactuals and his account of the meaningful representation of counterfactuals seems dependent on his account of negation. As such, his account of counterfactuals and negation is circular. However, counterfactuals and negation cannot be left to do each other's wash without inevitable difficulties arising (e.g., accusations of begging the question). As such, the hypothesis that the meaning of counterfactuals can be explained via negation fails and we are no longer left with any explicit account for the semantics of false propositions.[18, 19]

Canonical decomposition

Finally, during our discussion of Peircean mental models, we saw that a major boon to its adoption was that it seemed to explain such things as the productivity and the systematicity of language. However, as Fodor (2010, p17) avers:

> The content of a thought is entirely determined by its structure together with the content of its constituent concepts.

That is, productivity and systematicity are only provided if representations are shown to have constituents, to be compositional in nature. Nevertheless, although this compositionality is crucial for accounting for the above characteristics of language and, therefore, the continued success of Johnson-Laird's referential theory of mental models (or any other – i.e. discursive and non-discursive – RTM), Fodor (*Ibid.*, pp173-175) identifies a potential problem with Johnson-Laird's theory in this instance:

> Icons don't have *canonical* decompositions into parts*; all* parts of an icon are *ipso facto* constituents. Take a picture of a person, cut it into parts whichever way you like; still, each picture part pictures a person part[…] which is to say, however they are sliced, there's no distinction between their canonical parts and their mere parts[… But this is not] true of a discursive representations. Only a specifiable subset of the parts of a discursive symbol (namely, its canonical parts) are syntactic or semantic constituents[…] Because they decompose into syntactically and semantically heterogeneous constituents, discursive representations can have logical forms (maybe all discursive representations do that can express truths).

---

[18] However, it should be recalled that a number of recent empirical studies (e.g., Du et al 2014; Macbeth et al 2014; Orenes, Beltran and Santamaria 2014) goes contrary to these findings, concluding instead that "a symbolic representation of negation is possible" (Orenes, Beltran and Santamaria 2014, p36). However, the present thesis highlights the fact that these findings are only *consistent* with Johnson-Laird's account of negation, and are not conclusive of its veracity. As such, we are taking these empirical findings to be measures of some other phenomena, rather than that of a mental sign for negation.

[19] As previously highlighted, McGinn (1989, pp197-198) argues that for mental models to validly solve Brentano's problem, they need to adhere to the naturalism constraint. Therefore, Johnson-Laird's introduction of mental signs into his mental models may lead him into difficulties explaining intentionality.

By contrast, because they decompose into syntactically and semantically homogenous parts, iconic representations don't have logical forms.

That is to say, as models, or any other iconic representation, have no definitive composition, no decompositional 'ground floor' upon which objective observers can agree, there is no distinction in their parts that determine the contribution they make to the icon as a whole. This means that there are no individual discrete 'atoms' that make up an icon; such atoms are said to be unable to exist since each segregation of parts of an icon is equally valid. As such, because icons do not have what Fodor terms a canonical decomposition, icons, so it is argued, cannot be compositional in nature and cannot account for the productivity and systematicity of language.

Summary

To summarise the results of this investigation, the analysis and synthesis of Peirce's correspondence theory of meaning found a number of inconsistencies:

1) Johnson-Laird and Peirce differ on the role of mental images: The latter states that they are the basis of cognition while the former repudiates this. This is problematic as it demonstrates a clash between the philosophies of the two thinkers and makes those specific aspects Johnson-Laird has adopted from the iconic theory of signs, and those he has not, unknown.

2) Peirce relies on an Interpretant for his signs and symbols to be meaningful. This reliance leads him (and, by extension, Johnson-Laird) into the homunculus and symbolic fallacies, fallacies Johnson-Laird was trying to avoid by adopting such a theory of semantics in the first instance.

3) Although Johnson-Laird is vociferous in his rejection of mental signs through their violation of the symbolic fallacy and homunculus fallacies, his iconic correspondence theory of meaning inherits from Peirce mental signs for such things as negation. Therefore, this inclusion not only creates conflict within Johnson-Laird's hypotheses – specifically, it leads to a conflict with the model theory's existing naturalism constraint – it also concludes in a transgression of the very fallacies he is attempting to avoid.

4) Peircean mental models are led into an explanation of counterfactual propositions via the mental sign for negation. This explanation is both insufficient (as it leads to false counterfactuals with different cognitive meaning being analysed into the same proposition) and circular (as counterfactuals are required to analyse certain negative propositions and *vice versa*).

5) As Fodor (2010, pp173-175) demonstrates, Peircean icons cannot have canonical decompositions and, therefore, cannot be compositional in nature. This is a deficiency that results in such mental models being unable to account for the productivity and systematicity

of language, aspects of language that are, seemingly, only explainable through compositional representations.

Consequently, this analysis has found that we must respond in the negative to our second research question and state that: No, Peirce's congruent correspondence theory of meaning is both internally and externally inconsistent.

§3 Is the introduction of a correspondence theory of meaning into Johnson-Laird's broader philosophy of cognition logically consistent?

Finally, our methodology determined that the last measure against which the success of Johnson-Laird's updated semantics is to be held, is the external consistency of such an approach to meaning's implicit theoretical positions within his wider philosophy. Our literature review identified a number of such suppositions required for the consistent application of a congruent correspondence theory of meaning and, following from this, this section will investigate these assumptions in regard to Johnson-Laird's iconic mental models. This investigation will be conducted in an attempt to determine the logical consistency and completeness of such a proposed externalist theory of cognition. As such, this chapter will survey Johnson-Laird's works so as to identify whether the identified raft of assumptions are commensurate within his philosophy as a whole, or whether any inconsistency is created by their inclusion. Unfortunately, Johnson-Laird is not explicit on his position concerning a number of these listed assumptions. Consequently, in such instances where these positions are not made clear, this thesis will of necessity engage in some speculative analysis in order to deduce an answer to the question of consistency within Johnson-Laird's complete philosophy from his explicit writings.

To recap, our literature review uncovered that, in order for a correspondence theory of meaning to be consistent, it must assume the following positions:

1) A commitment to externalist semantics and a rejection of internalist hypotheses and arguments.
2) An ontological realism concerning states of affairs/facts.
3) That a reliable and consistent mode of reference fixation is possible. This is an assumption that itself entails a number of sub-assumptions:
   a. It is necessary to explain away the apparent limitations of genre or pictorial form where language or mental representations are concerned.
   b. The one-way nature of reference needs to be explained within a system that appears to employ a two-way relation of structural equivalence.

     c. Brentano's question/problem can be answered and reference can be explained in a scientifically reducible, empirical way.

4) An epistemic scepticism of meaning.

5) A number of metaphysical entailments:

     a. An eternalist stance on the metaphysical status of time (leading to determinism).

     b. That counterfactuals can refer to something, and therefore remain meaningful, even while singular terms are Russellian. (N.B. This also includes those propositions that entail negation, disjunctives, empty names and modality.)

     c. That properties and relations have some form of ontological reality.

6) That, if two terms have different referents, they must have different meanings, no matter their surface level similarity.

7) That meaning is malleable.

8) That the informativeness of Frege cases can be explained.

9) That the challenges posed to any solution of the implementation problem by Twin and Frege cases are avoidable.

10) That slingshot arguments can be answered.[20]

Taking this into consideration, this section will take each of these assumptions in turn, clarifying if they are, respectively, consistent or otherwise within the philosophy of Johnson-Laird or, where no explicit position is found, if it can be *inferred* that each assumption is consistent or otherwise within the philosophy of Johnson-Laird.

A commitment to externalist semantics and a rejection of internalist hypotheses and arguments

Johnson-Laird's externalism of meaning is a defining aspect of his contemporary RTM. In fact, Johnson-Laird (1988, p107) has argued on numerous occasions that if theories "say nothing about how words are related to the world" then those theories will fail, for, so he argues, they assume, "that meaning is merely a matter of relating one set of symbols to another" rather than being grounded in referents. Moreover, Johnson-Laird's explicit externalism of semantic content can be seen throughout his writings, with further examples being seen in *Ibid.* 1980, p89; 1983, pp182-191; pp230-231; 1988, p107; 1993, pp333-334; 2002, p80; 2006, pp21-37; Johnson-Laird and Byrne 1991, pp7-11.

---

[20] It should also be remembered that the success of Johnson-Laird's referentialist semantics also presupposes the legitimacy of the concept of ground. As such, and as the discussion surrounding ground is beyond the scope of these pages, it should be clarified that this research is assuming both the legitimacy of the concept of ground and its coherence with Johnson-Laird's wider philosophy. However, for a full analysis of this concept, see Audi (2012), Correia and Schnieder (2012) and Fine (2012).

Furthermore, Johnson-Laird not only explicitly holds an externalist semantics, he supports this position with arguments against the internalism of those such as McGilvery (1998), Chomsky (2000) and Pietroski (2005), saying:

> One decisive objection to all these [internalist] theories[…] is that they say nothing about how words relate to the world. They contain no machinery that explains how my warning about [a] table could guide your movements[… Such theories] can tell you that two words are related, or that one sentence is a paraphrase of another, but they cannot tell you anything about the state of the world. They are as circular as dictionaries; they commit the "symbolic fallacy" of assuming meaning is merely a matter of relating one set of symbols to another. But when you understand my warning, you grasp its "truth conditions": you know how the world should be if my assertion is true. As the philosopher David Lewis points out, the translation of an utterance into a representation such as [internalist semanticists do] provides no more of an account of its truth conditions than does a translation into Latin. (Johnson-Laird 1988, p107. See also *Ibid.* 1983, p231; 1993, pp331-334.)

Finally, although evidence of Johnson-Laird's adherence to externalism and antipathy toward internalism can be readily found within his work, an explicit handling of the three arguments we canvassed in our literature review, and that are commonly employed by internalists to defend their position, is not so redolent. As such, we will have to look to what *is* expressed so as to infer Johnson-Laird's potential position on: 1) Negative facts, 2) tractability and 3) poverty of the stimulus for language learning.

1) **Negative facts:** As we saw in our literature review, internalists argue that an externalist semantics is only capable of explaining the positive facts of language (i.e. the semantic content of terms), but not negative ones (e.g., the coreferentiality of some statements but not others). However, this argument seems to be avoided by Johnson-Laird's RTM position. For, such arguments from negative facts, as can found in Pietroski (2005), appear to be aimed solely at externalism regarding natural language (e.g., Pietroski (*Ibid.*, p255) is concerned only with "a theory of meaning for natural language"), but by doing so they miss the externalism of mental representations to which Johnson-Laird adheres (as, for Johnson-Laird, it is not language that possesses semantic content, it is the mental models to which natural language is translated (see Johnson-Laird 1980, pp89-91; 1983, pp248-259; 1988, p107; 1993, pp343-348)). Consequently, it appears that Johnson-Laird need not be concerned about negative facts, as he sees natural language only becoming meaningful through the mental representations into which they are translated.[21] Further, Johnson-Laird sees his mental models to be a *begriffsschrift* of sorts, with there being no possibility of hidden meanings where logical form

---

[21] Interestingly, the expression of cognitive science's position on mental models in Pietroski (2005, p270) is considerably outdated.

or relation structure is concerned (see Johnson-Laird 1980; 1993; 2002; 2006; 2013), which therefore removes the ambiguities and peculiarities found at the surface level of language and which the likes of Pietroski (2005, 266-279) use as their ammunition against externalism. In fact, it seems that cognitive scientists, such as Johnson-Laird, could appeal to the same innate biological and psychological constraints on the interpretation of natural language to explain negative facts in natural language that Pietroski does, without giving up their externalism of mental representations.

2) **Tractability:** Secondly, we saw that some internalists contest that externalism demands "a theory of everything" (McGilvery 1998, p237) as, supposedly, it demands computation over internal symbols and *all* the entities found in the world. However, this reasoning is based on a confusion over the role of computation in external RTM and completely neglects the role of reference fixation within that account. Under Johnson-Laird's theory of semantic content, computations are only needed to be performed over *iconic* mental representations and any change in iconicity resulting from these computations will, *eo ipso,* change the part of the world that the representation denotes. As such, Johnson-Laird does not "insist on relating a computational theory's domain to users of cognitive states and to things outside the head" (*Ibid.*, p232): all that is required is a computational domain over an individual's cognitive states (i.e. mental representations) and, due to the iconism of these cognitive states, this will, *eo ipso*, have the effect of computations over things outside the head. Consequently there are no tractability issues with Johnson-Laird's externalism and the implementation of such a system is viable.

3) **Poverty of the stimulus for language learning:** Finally, it has been seen that internalists, such as Chomsky (2002), use evidence showing that infants do not receive the requisite stimuli to account for the linguistic knowledge that they are able to demonstrate, i.e. some argue that there is a poverty of the stimulus for language learning. In response to this charge Johnson-Laird may be able to refer to his congruence-based correspondence theory of meaning. That is, it may be the case that Johnson-Laird relies upon, or could rely upon, the construction of appropriate reference fixations by infants to explain language learning, thereby rejecting the heart of the poverty of the stimulus argument (i.e., that a child does not receive the level of stimuli to explain their level of knowledge). For an example that this may be an available strategy for Johnson-Laird against such arguments, we can look to his approving references to the work of Bowerman (1977) when highlighting the presence of inductive learning

heuristics, as employed by infants, to fix reference. He concludes from this that:

> A baby girl of sixteen months hears the word 'snow' used to refer to snow. Over the next months, as Melissa Bowerman has observed, the infant uses the word to refer to: snow, the white tail of a horse, the white part of a toy boat, a white flannel bed pad, and a puddle of milk on the floor. She is forming the impression that 'snow' refers to things that are white or to horizontal areas of whiteness, and she will gradually refine her concept so that it tallies with the adult one. The underlying procedure is[…] inductive. (Johnson-Laird 1993, p234. see also *Ibid,* p243)

As such, it seems that the apparent presence of learning heuristics to fix reference, deduced from Bowerman's study, is a potential avenue for Johnson-Laird to reject innate knowledge and abilities, as they arguably demonstrate that *learning* is necessary for language and that there is no poverty of the stimulus (see also Hutto 2008).

Consequently, it can readily be seen that a commitment to externalist semantics and a rejection of internalist hypotheses and arguments is consistent with Johnson-Laird's wider philosophy and no dissonance is created with his move to an externalist semantics in this regard.

An ontological realism concerning states of affairs/facts

As was shown within our Method chapter, Johnson-Laird has demonstrated both an implicit and explicit commitment to an ontologically real world 'out there'. This demonstration can be seen implicitly in his adherence to Marr's tri-level hypothesis and its physical level of explanation (see Johnson-Laird and Byrne 1991, p17; Johnson-Laird 1993, p58; p390), as well as his approval of Ken Craik's 'relation structure' and 'physical models' (*Ibid.* 1980, p73; 1983, p9). Finally, Johnson-Laird has explicitly made a commitment to an ontological realism concerning facts and states of affairs through his continued reliance on logical items such as 'the world', 'concrete objects' and 'entities' as examples for things modelled by his mental representations (*Ibid*., 1980, p89; 1983, pp.410-415; 2002, p80; 2013, p3; Johnson-Laird, Legrenzi and Legrenzi 1972). As such, no inconsistency, surrounding ontological commitments, has been introduced into Johnson-Laird's work with this theoretical shift concerning language.

Reference fixation

As we have seen throughout this investigation, Johnson-Laird's recent writings demonstrate a commitment to reference fixation by congruence, or, more accurately, iconicity. This use by Johnson-Laird of congruence for reference fixation within a correspondence theory of meaning is seen in his

early writings through his use of Craik's 'relation structure' (see Johnson-laird 1980; 1983; 1993), but, since 2002, this has been updated with his adoption of Peircean iconicity, i.e. that:

> interesting property, which the great nineteenth century philosopher and logician, Charles Sanders Peirce, called "iconicity"[...] He meant that the structure of such a diagram is the same as the structure of what it represents, and so the parts of the diagram map onto the parts of the scene, and the relations among the parts of the diagram are the same as the relations among the parts of the scene. Mental models are similarly as iconic as possible. (*Ibid.*, 2008a p208. For further explicit examples of Johnson-Laird's reliance on iconicity for reference fixation, see *Ibid.* 2002, pp80ff.; 2006, pp21-37, 2010, p2; 2013, pp9-11.)

However, although the above does answer the question regarding the mode of reference fixation within our literature review, it does not appear to provide an answer to the further questions raised therein concerning reference fixation within correspondence theories of meaning. That is:

3a)   How are the apparent limitations of genre or pictorial form, where language and mental representations are concerned, explained away?

3b)   How is the one-way nature of reference explained within a system that appears to employ a two-way relation of structural equivalence (i.e., iconicity)?

3c)   How is Brentano's question/problem answered? That is, how is reference (i.e. iconicity) explained in a scientifically reducible, empirical way?

Consideration of the published writings of Johnson-Laird does not seem to produce a satisfactory answer to any of the above questions. In fact, looking to the literature only produces the following questions and queries:

3a)   If, as Marr's tri-level hypothesis states, mental representations are just an abstraction from physical nature and representations are ultimately biological electro-chemical entities, this suggests, following Daitz and – certain readings of – Wittgenstein, that the pictorial form or genre of Johnson-Laird's icons are, similarly, biological, electrical and chemical. However, this raises the question: How can biological electro-chemical icons represent anything other than electro-chemical biological properties? That is, how can biological electro-chemical signs signify the visual, aural, emotional etc. properties with which we are all conscious?

3b)   As has been established, Johnson-Laird relies upon iconicity to explain the relation between sign and signified. However, this equivalence relation of iconicity does not seem to allow for the one-way relationship seen in reference without some form of post-hoc stipulation. For example, this stipulation may perhaps be temporal in character, i.e. that

any new structure that shares the logical form of an existing state of affairs will refer to the earlier structure, and not *vice versa*, simply by fiat. However, it should be noted that although this stipulation may hold in most cases (e.g. a picture and a scene), there are non-standard situations that test this possibility (e.g. an architect's model, a draughtsman's blueprints and a finished building). As such, it seems difficult to identify any stipulation limiting the directionality of reference for icons, let alone one that is not post-hoc.

3c) Marr's tri-level hypothesis – which, again, states that the middle, representational, layer is instantiated by the physical, electro-chemical, processes of the brain or a whole person – may appear to answer Brentano's question/problem by stating that the intentionality of icons is ultimately instantiated in physical nature. Furthermore, Johnson-Laird could explain the relationship between a worldly state of affairs and a mental representation by a causal (see Fodor 1990; 2010), or even teleological (see McGinn 1989), account of content. However, these explanations, in themselves, are unsatisfactory. For Brentano's question to be answered satisfactorily, these accounts need to be cashed out and their physical instantiation identified. That is, it needs to be shown how the nervous, electrical, chemical etc. structure of the body and the neurology of the brain is capable of producing iconic mental models and their causal or teleological relationships with the world. However, how this question is answered by Johnson-Laird, or anyone else, is yet to be determined.[22]

Consequently, although the assumption that a mode of reference fixation exists is compatible with Johnson-Laird's philosophy, it appears that the detailed entailments of this fixation, i.e. that: (i) The apparent limitations of genre or pictorial form can be explained away; (ii) the apparent one-way nature of reference can be answered; and (iii) Brentano's question/problem can be resolved) - have not been accounted for. As such, Johnson-Laird's theory of cognition seems to be incomplete in this respect and an introduction of correspondence must be said to create certain problems in Johnson-Laird's philosophy as a whole.

Epistemic scepticism of meaning.

---

[22] Furthermore, there also exist arguments that functionalist accounts of mind, such as is used by Johnson-Laird, are innately incapable of averting Brentano's question. Such arguments state that functionalist theories cannot reduce synonymy, i.e. equivalence of referent between interlocutors, to functionalist explanations, as we all hold differing beliefs about the same terms and untangling this knot is "totally utopian" (Putnam 1988, p75). However, it should be noted that such arguments only follow if an holistic view of meaning is taken (see *Ibid.*, pp73-89), a view of meaning criticised in our literature review and that is, as we shall see, antithetical to Johnson-Laird's *de re* linguistics.

In our literature review, it was identified that correspondence theories of meaning lead to an epistemic scepticism of meaning. That is, as articulated by Miller (2007, p145), correspondence theories of meaning have to "concede[…] that the sort of fact in question [that fixes meaning] exists, [but…] questions our right to claim knowledge of that sort of fact". However, this is a conclusion Johnson-Laird (1983 pp193) readily accepts, stating:

> unfortunately, there is no reason to suppose that the chemical tests for [natural kind terms] are infallible[…] moreover, even if a "definite" answer has been obtained to the hidden structure of a [natural kind term] either through scientific investigation or through the application of some "common sense" test, the answer is defeasible[…] An answer is most unlikely to be forthcoming, and, I submit, could never be established definitively.

As such, Johnson-Laird (*Ibid.* p194) concludes that, "we do not know that we know [a natural kind term's] extension – perhaps we do, perhaps we don't". Consequently, the implicit epistemic scepticism of meaning entailed in the adoption of a *de re* semantics is consistent with the rest of Johnson-Laird's philosophy.

Metaphysical entailments

Certain of the metaphysical assumptions a correspondence theory entails (i.e., an eternalist stance on the metaphysical status of time and that properties and relations have some form of ontological reality) are far too complex to be analysed in detail here. As such, this thesis will merely grasp the nettle in saying that, in the main, these positions are both consistent within Johnson-Laird's philosophy as there are no explicit statements that reject or are dissonant with the majority of them (see Johnson-Laird 1983; 1993; 2002; 2006; 2013) and, consequently, that the implications of such an ontology are acceptable to Johnson-Laird. However, it should be noted that Johnson-Laird has made, in passing, some remarks that do seem to clash with at least one of these metaphysical positions. He has remarked, for example, that, where free will is concerned:

> Even a system that is not deterministic can be completely understood though its specific behaviour cannot be predicted. The compatibility of free will with computational explanation reveals the crassness of the assumption that psychology is a means for the prediction and control of individual behaviour. No science can predict the products of your imagination, or what you will do on every occasion. Cognitive science does not threaten your freedom or dignity. (Johnson-Laird 1993, p390)

Nevertheless, the eternalism of time necessitated by correspondence (a commitment, it should be remembered, needed to account for the meaning of tensed propositions) does threaten this freedom. As such, Johnson-Laird will have to abandon the above position on free will if his philosophy is to be consistent with his congruence theory of meaning. Additionally, there is, as we have seen, an issue for icons with a correspondence theory's entailment that counterfactuals refer to something to remain meaningful, with this chapter's previous section demonstrating that a Peircean sign for negation does

not answer this question successfully. In light of this, how Johnson-Laird's theory answers this question is unknown.

Consequently, although a realism concerning properties and relations and temporal eternalism are not overtly inconsistent with Johnson-Laird's other writings, clashes, or questions, concerning free will and negation do seemingly exist, or are raised with the introduction of correspondence. (. This also includes those propositions that entail negation, disjunctives, empty names and modality.)

Surface level similarity

Twin Earth cases show that if, like Johnson-Laird (2002, p84), you hold that the meaning of most, or even all, of our terms is iconic, is the correspondence of our natural language (by way of mental representations) with the world, then you must conclude that if two symbols (mental or otherwise) are the icon of – correspond to – the same state of affairs, then they have the same meaning.[23] Though this conclusion is a necessary one from the given premise, it is not yet clear if it is one that fits into Johnson-Laird's wider philosophy without discord. Although an explicit adherence to this necessary conclusion of his philosophy is not common, there are passages that suggest that Johnson-Laird adheres to it. For example, he avers:

> Many different scenes can be described by the same words, and many different words can describe the same scene. Hence, the relations between words and scenes must depend on elements at a lower level than either[…] the relations between words emerge from their relations to models of the world. (Johnson-Laird 1993, pp339-342)

In this regard, therefore, no inconsistency within Johnson-Laird's philosophy is found.

That meaning is mutable.

Although Johnson-Laird assents to the conclusions of the Twin Earth argument (i.e., that if two terms have a different referent then they have a different meaning) this may, as we have seen, lead him into further inconsistencies. For, our literature review found that Putnam (1975, pp215-271), further, purports to show with his Twin Earth argument that two people who share the same mental state (or, in the language of cognitive science, the same 'mental representation') do not necessarily mean the same thing by that state (or representation). From this, it must be concluded that any description of semantic content that is based only on mental states (or representations) cannot be sufficient. Nevertheless, it should be noted that this conclusion is only generated if we accept that:

> the extension of the term 'water' was just as much H_2O on Earth in 1750 as in [2014]; and the extension of the term 'water' was as much *XYZ* on Twin Earth in 1750 as in [2014]. (*Ibid.*, p224)

---

[23] The meaning of '*most*' of our terms, as Johnson-Laird (2002, p84) argues that such things as negation cannot be represented iconically and are therefore represented via mental symbols.

That is, Putnam's argument only stands if we accept that the meanings of our terms are immutable. However, if this assumption is denied and Johnson-Laird takes the position that semantic content can change over time, then Putnam's problematic conclusions are avoided. *Pace* Putnam, such a position asserts that both theorists of 1750 share the logical form of the surface structure of 'water' and, therefore, that the term used to name the substance is synonymous for and between the Earth and Twin-Earth scientists of 1750 (i.e. 'water' means 'clear, tasteless liquid' for both). Moreover, this position contends that, once the micro level logical form of the substance is discovered, the denoting model is updated and meaning splits between the two Earths. In turn, this leads to the position in 2014 where 'water' has a different meaning on Earth than on Twin-Earth. However, it may be contested that this necessitates the acceptance of the unintuitive notion that meaning changes depending on scientific discovery, and, therefore, that, for example, Archimedes' use of 'gold' is different to our use of the term today (cf. Papineau 2014, p171). Nevertheless, this is a conclusion Johnson-Laird (1983, pp194-195) freely accepts, stating:

> Natural kind terms can change their extension over time. For example, the word *tree* originally designated oak trees and the word *cattle* first meant property, then movable property, then livestock, and only finally bovine animals[…] The extensions of natural kind terms are not immutable.[24]

Consequently, Johnson-Laird responds to Putnam's argument that there exist situations in which two individuals with the same mental representation can mean different things by denying such situations can arise and by stating that the meaning of our terms are not static. He asserts that our mental models correspond only to those things that share their logical form and as these models are updated with further knowledge, so is this logical form and those states of affairs to which the model does and does not correspond. As such, for Johnson-Laird, the meaning of 'water' splits in Putnam's Twin-Earth scenario; experts and the laity mean different things by their use of expert terms (e.g., 'beech' and 'elm' may refer to the set of all trees for the layman); and vague terms, such as a politician's speech acts, refer to a large set of entities (i.e., vague and ambiguous terms and propositions either refer to nothing and are meaningless or mean all those states of affairs that share its logical form). Furthermore, it should be highlighted that a dynamic concept of meaning is consistent within Johnson-Laird's philosophy as a whole, which states that all that is meaningful is mental models. Therefore, if the mental model to which a natural language term is translated is updated (e.g., by scientific discovery into the hidden structure of entities), the meaning of that term *must* change, must be malleable across both time *and* individuals.

---

[24] It should be noted that this is also the response favoured by many contemporary commentators, with Chomsky (2000, p32) stating:
> If I know nothing about elms and beeches beyond the fact that they are large deciduous trees, nothing beyond this information might be represented in my mental lexicon. (See also Fodor 2010 pp16-17.)

That the informativeness of Frege cases can be explained

As we saw, another case that demonstrated nuance in the relationships between terms and referents arose through substitutivity/Frege cases. Within such cases, it was found that there are certain terms (e.g., 'Charles Dodgson' and 'Lewis Carroll' or 'Hesperus' and 'Phosphorus') that play a different role in the informativeness of propositions although they share the same referent. In response to such cases, we saw that Frege moderates his externalism of meaning by establishing both a psychological and an empirical element to meaning (i.e., by including sense as well as reference into his description of language). Interestingly, this is a response that Johnson-Laird (1983, p192) has, at one point, also adopted in response to such cases. He states:

> Something at least is to be found in the mind, and something that has to do with the meaning of words[…] I want to[…] restore the study of meaning to psychology.

In fact, at certain points, Johnson-Laird (*Ibid.*, p203) has gone as far as to state that meaning can be derived *purely* psychologically (i.e., from sense or narrow content alone), asserting that the "meanings of some words are mental constructions that are imposed upon the world in the absence of an objective correlate". As such, Johnson-Laird has once taken a position akin to that held by Jerry Fodor during the 1980s and touched upon within our literature review: That of narrow, internalist content and a moderate internalism (see Fodor 1987; 1991). Nevertheless, and akin to Fodor (1994), this moderate internalist position has subsequently been abandoned by Johnson-Laird (1994, p334), who later stated that meaning is explained by "show[ing] how linguistic expressions relate to models[…] and then [showing] how models relate to the world" alone. In this conception, content is only accounted for on a broad, *de re*, basis, with no role for any form of psychological, narrow content.[25] Consequently, for Johnson-Laird (1988, p107), any symbol (apart from those regarding abstract concepts) that is now not grounded is meaningless or leads to fallacious conclusions. However, due to this position, Johnson-Laird, like the contemporary Fodor (2010), is forced to adopt a controversial strategy in regard to answering the informativeness of Frege cases. For, if narrow content cannot be appealed to and meaning is said to be purely *de re*, such informativeness of substitutivity cannot be answered along Fregean lines (i.e. with a distinction between sense and reference, narrow and broad content) on pain of inconsistency. In fact, the informativeness of *content* of Frege cases cannot be answered by such *de re* accounts of meaning at all. Rather, only the *apparent* informativeness of content in Frege cases can be accounted for. For, if broad content or reference alone determines meaning of natural kind terms, 'Hesperus' *must* be synonymous with 'Phosphorus', or 'Charles Dodgson' with 'Lewis Carroll': No matter what our perceptions of them are, 'Hesperus is Phosphorus

---

[25] Excepting, of course, those mental symbols for such things as negation and "other abstract concepts" (Johnson-Laird and Khemlani 2013, p11), whose inconsistency were analysed in the previous section. However, even granting these mental symbols for abstract concepts, it is difficult to see what role this particular form of narrow content can play in Frege cases, cases that do not concern 'abstract concepts'.

and Lewis Carroll is Charles Dodgson' is exactly as informative as 'Hesperus is Hesperus and Lewis Carroll is Lewis Carroll' (or any other truth preserving variation of the co-referring terms). If the two propositions appear to possess different levels of informativeness of content, it is that appearance alone that merits analysis. Unfortunately, Johnson-Laird fails to provide an explicit analysis of this kind within his work. However, one potential strategy open to him can be found within Fodor (1990, p161-146; 2010, pp68-100), wherein Frege cases are said to not demonstrate informativeness of propositions in regard to content, but the informativeness of propositions in regard to *beliefs* and/or *desires*.[26] This form of an explanation proceeds by positing two or more mental representations for the one state of affairs (this can be achieved via mental icons possessing a variation of the one fact's logical form) and identifying differing sets of beliefs and desires concerning or attending those representations (see Fodor 1990, p167; 2010, p73). For example, an individual may possess two synonymous representations, one that picks out the 19th century academic and theologian Charles Dodgson and the other that is the icon of the writer of *Alice in Wonderland*, Lewis Carroll. Along with these two, semantically equal, representations, said individual also possesses (for the sake of argument) three beliefs concerning these representations:

1) 'That Lewis Carroll is Lewis Carroll' is true.

2) 'That Charles Dodgson is Charles Dodgson' is true.

3) 'That Lewis Carroll is Charles Dodgson' is false.[27]

Consequently, the informativeness found in the propositions 'Lewis Carroll is Charles Dodgson' to such a person is not an informativeness of content, but an informativeness of beliefs (and desires): The proposition updates or informs the individual's beliefs, replacing 3) with 4):

4) 'That Lewis Carroll is Charles Dodgson' is true.[28]

However, and as we saw in our discussion of LoT2 within our literature review, for this account to succeed, it is necessary for an individual to possess both 1), 2) and 3) *at the same time*, i.e., to simultaneously hold mutually inconsistent beliefs. Although this commitment to doublethink may be unpalatable to some, it is a phenomenon that Fodor (1990, pp141-144; 2010, pp75-82) freely accepts and one that Johnson-Laird must also, on pain – ironically – of inconsistency. Fortunately, this is a position Johnson-Laird (Johnson-Laird, Legrenzi & Girotto 2004) endorses, identifying a number of real

---

[26] See also McDowell (1977, p173).

[27] Obviously, this is a simplification of the number of belief states concerning the two mental representations that would actually be involved in practice: The number has been restricted here for the purposes of presentation only. In reality there would not merely three beliefs at play, but many and include, where our examples are concerned, those such as: ''Charles Dodgson is a mathematician' is true'; ''Lewis Carroll is a mathematician' is false'; ''Lewis Carroll wrote *Alice*' is true'; ''Charles Dodgson wrote *Alice*' is false', and many others. Additionally, desires as well as beliefs would be included.

[28] Again, this is a simplification of the true state of affairs. In practice, it would not just be the one belief that would be updated, would be informed, by the Fregean proposition, but the raft of beliefs attending the two propositions we saw in our previous footnote, i.e., it would produce the following set of beliefs: ''Charles Dodgson is a mathematician' is true'; ''Lewis Carroll is a mathematician' is true'; ''Lewis Carroll wrote *Alice*' is true'; ''Charles Dodgson wrote *Alice*' is true', and many others. Again, desires as well as beliefs would also be informed.

world cases where mutually inconsistent beliefs or desires are held simultaneously, including the Chernobyl engineers who, at the time of the incident, simultaneously believed:

> If the reactor is intact, then it is safe.
>
> If the reactor is safe, then no graphite is outside it.
>
> The reactor is intact, and some graphite is outside it. (*Ibid.,* p41)

As such, this is a position consistent with Johnson-Laird's stance as a whole and the challenges posed to an external semantics by substitutivity/Frege cases can be validly avoided by Johnson-Laird.

The implementation problem

There is one final issue raised by Twin and Frege cases, this time concerning the implementation problem. It should be remembered that such cases seemingly demonstrated that any solution to the hostility between the theories of computation and externalism can only do so in a way that "lead[s] to predictive/explanatory failures" (Fodor 1990, p39). That is to say, such breakdowns in the relationship between mental representations and states of affairs as highlighted by Putnam and Frege cannot be accounted for by a reliable means of reference fixation and any external RTM. Taking this into consideration, and as Fodor (*Ibid.*, p24) observes, a potential solution to such a problem is apparent:

> All that follows is this: if psychological laws are broad and computationally implemented, then there must be a mechanism whose operation *prevents Twin cases and Frege cases from arising* among creatures that the laws subsume.

That is to say, if Twin and Frege cases can be demonstrated to not occur, than a reliable form of reference fixation is free to answer the implantation problem without difficulties. Our previous discussions should demonstrate that this is exactly what Johnson-Laird is capable of doing. For, his position that meaning is mutable challenges the supposition at the root of the difficulties that Twin cases present: That I and my twin share the same intentional psychology (see Fodor *Ibid.*, p122). Furthermore, we have demonstrated that substitutivity/Frege cases may not be what they first appear – a breakdown between psychology and referent – but can, in fact, be examples of cognitive dissonance and conflicting belief/desire states. Consequently, Johnson-Laird is able to meet the assumption that the challenges posed to any solution of the implementation problem by Twin and Frege cases are avoidable by denying that true Twin and Frege cases occur.

Slingshot arguments

Finally, our literature review diagnosed that to be consistent with correspondence theories, Johnson-Laird must be opposed to slingshot arguments, i.e. "argument[s…] designed to provide a formally strict proof of the claim that all true sentences designate (denote, refer to) one and the same thing, as well as all false sentences do" (Shramko and Wansing 2009, p430). Unfortunately, Johnson-Laird never explicitly faces the challenges of slingshot arguments, so no firm answers to this assumption can be

made here. However, this may still make no difficulties for this research, as the problem of slingshot arguments can be dissolved, a dissolution Johnson-Laird can easily uphold without inconsistency. For, it may be argued that slingshot arguments are not applicable in standard meaning contexts under RTM, as semantic value is not (*pace* Frege) said to be mere truth value. For, the slingshot arguments are dependent on the preservation of logical equivalence, not semantic equivalence, and unless semantic content is not taken to be truth value, logical equivalence is not preserved. This therefore rejects the formal translation inherent in the slingshot argument, wherein propositions can be translated into truth values without loss of meaning. If this is denied (as in Russellian definite descriptions, for example), the argument fails. Furthermore, as Davidson (1969 p754) observes:

> I think there is a fairly simple explanation for our frustration: we have so far left language out of account. Statements are true or false because of the words used in making them, and it is words that have interesting, detailed, conventional connections with the world.

As such, the blanket translation of propositions into the formal system necessary for the success of slingshot arguments can be denied, as it misses the detailed nuance necessary for meaning. Consequently, referentialists, such as Johnson-Laird, may follow Baumgartner (2010, p112) in arguing:

> if the question as to the identity of facts is raised in isolation, the fact theorist can easily counter a corresponding slingshot by stipulating that facts are fine-grained entities whose identity depends on the predicates and singular terms used to state them.

Consequently, if *semantic contents are taken to be states of affairs and not truth values*, as is diagnosed under Johnson-Laird's iconicity, slingshot arguments fail and the assumption that slingshot arguments can be answered is consistent within his position.

Summary

It has been found that, in the majority of cases, the assumptions inherited with a correspondence theory of meaning are commensurate with Johnson-Laird's broader philosophy. However, this investigation also identified a number of areas where dissonance seems to have been introduced or difficulties have not been addressed. The areas where *consistency* are found are:

a) A commitment to externalist semantics.
b) A rejection of internalist hypotheses and arguments.
c) An ontological realism concerning states of affairs/facts.
d) That a reliable and consistent mode of reference fixation is possible.
e) An epistemic scepticism of meaning.
f) An eternalist stance on the metaphysical status of time.
g) That properties and relations have some form of ontological reality.
h) That two signs who possess the same referent have the same meaning.
i) That meaning is mutable.

j) That the apparent informativeness of content found in substitutivity/Frege cases can be explained.

k) That the challenges presented to the implementation problem by Twin and Frege cases can be resolved

l) That slingshot arguments can be answered.

The areas where *dissonance or incompleteness* are found are:

a) That the limitations of genre or pictorial form, where mental representations are concerned, can be accounted for.

b) That the one-way nature of reference can be explained.

c) That Brentano's question/problem can be answered and that reference can be explained in a scientifically reducible, empirical way.

d) The need for determinism.

e) That counterfactuals (and those propositions that entail negation, disjunctives, empty names and modality) can refer to something and therefore remain meaningful without inconsistency.

Consequently, it has been found that the introduction of a correspondence theory of meaning into Johnson-Laird's philosophy is not wholly consistent. As such, the answer to the last of our research questions is a 'no'.

Findings

Through the application of critical philosophical analysis to the arguments and hypotheses of Philip Johnson-Laird, a number of results have been derived. Firstly, it has been discovered that externalism appears to be the correct approach to take when countering the semantic arguments levelled against cognitive science and encapsulated by the CRA and SGP. This conclusion was reached, as it was found that grounding symbols to an ontologically real world, that is independent from any one observer, successfully introduces language entry and language exit rules into a theory of mind. This grounding relation allows symbols, such as Johnson-Laird's mental models, to get off the language 'roundabout'.

Secondly, it has been determined that Johnson-Laird's appropriation of Peirce's correspondence theory of meaning necessitates the inclusion of the problematic concepts of an Interpretant and mental signifiers (for such things as negation and other abstract concepts), and that these inclusions lead Johnson-Laird into both the homunculus and symbolic fallacies he was trying to avoid by adopting a referentialist semantics. Furthermore, these mental signifiers were found to conflict with Johnson-

Laird's own naturalism constraint, in that they introduce something not found in a state of affairs into a mental representation. It was further found that dissonance also existed between Peirce and Johnson-Laird concerning the role of mental images in cognition. As such, it was concluded that Peirce's theory of iconism is both an internally and externally inconsistent addition to Johnson-Laird's RTM.

Finally, it has been shown that a number of difficulties raised by correspondence theories have not been addressed by Johnson-Laird's hypotheses: The limitations of genre/pictorial form, the one-way nature of reference and Brentano's problem need to be tackled before the theory can be confidently accepted. Moreover, it was found that there are difficulties for Johnson-Laird's theory where free will and negation are concerned, as Johnson-Laird's rejection of determinism is antithetical to correspondence theories while his account of negation has been demonstrated to be problematic.

Consequently, In relation to the research questions of this investigation, these findings can be summarised thus:

Q1) Does Johnson-Laird's updated account of meaning circumvent the semantic objections of the RTM's commentators (i.e., the CRA, the SGP, holism and the homunculus and symbolic fallacies)?

A1) Yes, Johnson-Laird's theory is successful where the relation it bears to the CRA, the SGP, semantic holism and the homunculus and symbolic fallacies are concerned.

Q2) Is Peirce's structural correspondence theory of meaning both internally and externally consistent?

A2) No, Johnson-Laird's theory fails where both 1) the logical coherence of its specific correspondence theory of meaning, in its own right, is concerned and 2) the logical coherence between the theory's iconism and the rest of Johnson-Laird's account of cognition is concerned.

Q3) Is the introduction of a correspondence theory of meaning into Johnson-Laird's broader philosophy of cognition logically consistent?

A3) No, Johnson-Laird's theory fails where the logical consistency between his existing philosophy of mind with the positions necessary for a correspondence theory of meaning, generally, is concerned.

As such, the above demonstrates that Johnson-Laird's theory is not able to answer in the affirmative to our three research questions and, following from our Method, is therefore claimed to be falsified.

Evaluation

It is necessary to establish the validity and importance of the findings outlined above before asserting any final conclusions. That is, it is incumbent on this research to identify potential weaknesses or shortcomings of the investigation and make comment upon how these may affect the validity of the findings therein, as well as to ascertain the potential influence or interest such findings may have on the current literature or to contemporary researchers.

To this end, we must first reaffirm that these findings are only applicable to the work of Johnson-Laird and should not be generalised to any other referentialist RTM (as found, for example, in McGinn 1989; Brooks 1990; Harnard 1990b; Waskan 2006; Fodor 2010; Rapaport 2011; Schweizer 2012) without further work. As we saw in our Method chapter, this limitation is due the fact that, since only the hypotheses and arguments of Johnson-Laird were analysed, the nuance that exists between the different referentialist RTM positions has not been captured and the scope of this study's findings will not include these other positions.

It should also be clarified that, through this research's application of the philosophic method of analysis and synthesis, the operation of Johnson-Laird's referentialist philosophy of mind and language has perspicuously been demonstrated and, because of this transparency, this research has been able to identify areas where inaccuracy are introduced (i.e., signs for negation etc.). Additionally, it is of interest to observe that, under a Popperian epistemology, the establishment of logical consistency is not sufficient for the determination of a theory's accuracy, i.e., it is a method that cannot aver that it is *true* that meaning is determined by reference (for there is an infinity of logically consistent theories that do not obtain). However, logical inconsistency, for Popper, can be sufficient to *disprove* a theory and can remove it as a 'provisional conjecture' (see Popper and Miller 1983, pp155-156). Furthermore, under a Quinean epistemology, the establishment of a theory's accuracy is measured against holist principles, that is:

> that scientific statements are not separately vulnerable to adverse observations, because it is only jointly as a theory that they imply their observable consequences. (Quine 1987, p228)

As such, the establishment of a lack of cohesion between the hypotheses that comprise Johnson-Laird's theory is sufficient to disprove it under both a Popperian and a Quinean epistemology. Consequently, the findings made by this research can be said to be of importance to current literature:

They show that there is an error within Johnson-Laird's philosophy that means that it cannot be an accurate description of semantic content and must be removed as a 'provisional conjecture'. Finally, these findings are also of interest in that they support the conclusions of other investigations (e.g., Vosgerau 2006).

However, although our findings have been found to be a valid and important contribution, that, in itself, is not sufficient for them to be considered conclusive. For, as Popper (1968, p20) argues, only the 'fittest' of theories, i.e. those who have survived "the fiercest struggle for survival", can stand with any confidence. As such, for the above to remain with any certainty, they will have to first be supported by a broader test of Johnson-Laird's hypotheses than has been undertaken here. That is, for confidence in the findings of this thesis to be established, they first have to be triangulated with the results of other, methodologically divergent, investigations.

Moreover, although the application of philosophic analysis and synthesis is a repeatable, ordered process, that is no guarantee of successful application. For, as Russell (1918, p498) observes:

> When you pass from the vague to the precise by the method of analysis and reflection[…] you always run a certain risk of error. If I start with the statement that there are so and so many people in this room, and then set to work to make that statement precise, I shall run a great many risks and it will be extremely likely that any precise statement I make will be something not true at all.

As such, this research's findings not only need to be triangulated with those from methodologically divergent studies, they also need to be replicated and *confirmed* in their own right.

Therefore, although this study's findings are valid and important on both Popperian and Quinean grounds, they can in no way be said to be conclusive. To establish confidence in these findings, further research is required to: a) Confirm their (and those of Vosgerau 2006) conclusions; and b) triangulate its findings with other forms of investigation.

Conclusion

When looking to the conclusions we can draw from our findings, we first have to remind ourselves of our Thesis Aim:

> **Aim:** To assess the validity of the hypothesis that the introduction of a Peircean correspondence theory of meaning into Johnson-Laird's mental model theory of mind saves the latter from the critiques of its commentators and to test the success of the inclusion of a Peircean referentialist semantics into such a representational theory of mind.

Consequently, it can be seen that the hypothesis in question has been assessed to be invalid; that the inclusion of a Peircean iconism within Johnson-Laird's research project/programme is unsuccessful and that we can now state that Peircean iconism, is in itself, both internally and externally inconsistent and that a number of updates need to be made to Johnson-Laird's philosophy as a whole before any correspondence theory of meaning can be introduced without difficulty.

However, it was also found that the successful grounding of mental models generally is effective in preventing the homunculus and symbolic fallacies as well as meaning holism. As such, this thesis concludes that the inclusion of some form of correspondence theory of meaning into Johnson-Laird's mental model RTM, after a number of updates, could be both successful and productive, but that the Peircean approach is not so. Finally, it was also found that these findings are both important and interesting and support the conclusions from the investigations of others (e.g., Vosgerau 2006). However, it should be noted that these findings and conclusions need to be supported by further investigations before their validity can be confidently asserted.

Discussion

Chapter Five: Discussion

Introductory remarks

As we have already discussed what our findings might mean, how valuable they are and why they may be of interest at the end of the last chapter, this chapter will, following from this, reflect on these conclusions and will put forward a number of pertinent interpretations and opinions we have concerning them. That is, in contrast to the traditional thesis format, this chapter will solely try to clarify some of the implications of this research's findings and attempt to suggest avenues for related future research. To achieve this, this chapter will follow the structure from that of our Findings, with our discussion being broken into three sections, each covering pertinent topics and inferences raised by each of our research questions:

1) Does Johnson-Laird's updated account of meaning circumvent the semantic objections of the RTM's commentators (i.e., the CRA, the SGP, holism and the homunculus and symbolic fallacies)?

2) Is Peirce's structural correspondence theory of meaning both internally and externally consistent?

3) Is the introduction of a correspondence theory of meaning into Johnson-Laird's broader philosophy of cognition logically consistent?

§1 Does Johnson-Laird's updated account of meaning circumvent the semantic objections of the RTM's commentators?

The first section of our Findings chapter concluded that the introduction of a correspondence theory of meaning into Johnson-Laird's analogical RTM was a valid strategy for avoiding the issues raised by the CRA and the SGP. However, the introduction of this externalism raises some interesting points of discussion regarding:

1) The position any theory that contains externalism must implicitly hold toward the mereological fallacy.

2) What *type* of externalism Johnson-Laird has included.

The mereological fallacy

Firstly, during our Findings chapter it was observed that not only must an individual's mental signs be grounded, but also that individuals must be capable of grounding their own signs, i.e., must possess language entry and exit rules. That is, it was observed that meaningful content does not come pre-loaded. Rather, human cognitive systems possess "original" content (McGinn 1989, p155). As such, this causal referentialism propounded by Johnson-Laird – and others – raises further questions concerning 'bedridden AI', i.e. those systems which "exhibit no "language-entry" and "language-exit" transitions, to use Wilfrid Sellars' terms, and have no capacity for non-linguistic perception or bodily action" (Dennett 1980, p429). For, it now seems questionable whether systems that do not possess original content and the sensory apparatus to interact with the world non-linguistically truly possess meaning. Furthermore, this implies that for a man-made system to mimic human understanding (i.e., for Strong AI, extra-linguistic interactions with the world are needed to be implemented. Ergo, it seems that adopting externalism implicitly demands holding that intelligence is necessarily embodied.

Furthermore, this – implicit – move toward embodied intelligence by Johnson-Laird requires an, as yet, unspoken acceptance and reaction to the mereological fallacy. For, this call for referentialism must conclude that the mind is not an enclosed system. Rather, externalism claims that for meaning and consciousness to be present in a system, that system must be *enmeshed* with its environment. That is, externalists state that the environment is *necessary* for the content of mental states. Consequently, adherents of a referentialist RTM, such as Johnson-Laird, are *eo ipso* opposed to the hypothesis that the mind is solely encapsulated by the brain and therefore inherently accept the mereological fallacy (see Bennett and Hacker 2003; Damasio 2005; Byrge *et al* 2014; Pfeifer *et al* 2014). For example, McGinn (1989 pp15-19; 100-107; 117) argues that his externalism is innately opposed to substantialism and any mind/brain identity theory as:

1) Under substantialism, "the brain does lie literally in the head [rather than being enmeshed with its environment]. So it is a consequence of externalism that this simple kind of [substantialist] materialism is false" (*Ibid.* p15).

2) The referentialist mind does not have any intrinsic properties, only relational ones and so cannot be a substance (*Ibid.* pp100-103), while the brain, quite clearly, is a substance.

Consequently, any externalist must follow the logical inference and conclude that the mind cannot equal the brain. Moreover, this chain of reasoning leads McGinn (1989, p103) to explicitly accept that a consistent externalism avoids the mereological fallacy, stating: "what seems immediately clear is that the physical basis [of the mind under externalism] cannot be located wholly within the head".

Weak and strong externalism

Additionally, we saw in our literature review that correspondence theories of meaning presuppose a reliable mode of reference fixation. In this regard, Johnson-Laird's use of Peirce's theory of congruent reference fixation enables him to answer this question: He answers the question 'How is reference fixed' with iconicity (i.e., with logical form or relation structure). However, an issue needs to be clarified with this hypothesis, an issue that concerns what McGinn (1989, pp7-8) identifies as the distinction between strong and weak externalism. A distinction McGinn (*Ibid.*) defines as:

> Weak externalism[…] is the thesis that a given mental state requires the *existence* of some item belonging to the nonmental world, and that its identity turns on that item. Strong externalism[…] is the thesis that a given mental state requires the existence *in the environment of the subject* of some item belonging to the nonmental world, and that its identity turns on that item. The latter thesis is stronger than the former because the former does *not* entail that the subject of the mental state in question should be environmentally related to the extramental item in question – as it might be, causally. Crudely put, strong externalism ties mental states to a particular *part* of the world – the part the subject himself inhabits – while weak externalism ties mental states to the world *at large*, whether or not the required nonmental entity is where the subject himself is located. (McGinn 1989, pp7-8)

This distinction is pertinent, as Johnson-Laird has not clarified the criteria that delimits to what an iconic mental model refers and is said to mean. That is, he has not answered the question: Can icons only correspond to those states of affairs with which an individual is causally related (i.e. does he hold strong externalism) or can they be linked to any state of affairs in the universe regardless of its relation to the individual (i.e. does he hold weak externalism)? That this distinction is important and needs to be clarified by Johnson-Laird's theory can be seen once we analyse both types of externalism and see that each raise questions that need to be answered before a theory that contains them is to remain valid.

For, to hold strong externalism is to hold that some form of causal connection needs to be established between items in an individual's environment and that individual's mental models to institute content. That is, under strong externalism, although iconism is necessary for reference it is not sufficient, iconism *and* an appropriate causal relation are. However, such an appropriate causal relation is currently absent from Johnson-Laird's philosophy and needs to be introduced if he is to validly hold strong externalism. Nevertheless, establishing such an appropriate causal relationship may be problematic. For example, if this appropriate cause is identified with reliable perceptual experiences of states of affairs, as McGinn (1989, pp70-80) argues to be intuitively the case for strong externalism, then it seems we are led into some form of circularity. For, how can we individuate a state of affairs unless we already know and understand it? Or, as McGinn (*Ibid.*) argues:

> there is a circularity lurking within the heart of [causal theories of perceptual content…] The circularity comes into focus when we ask what *sort* of causal relation it is that confers content. It cannot, of course, be any old causal relation – or else any property instantiated along the incoming causal chain would come to be represented by the experience. We need to single out the right kind of distal object, viz. the kind of object that is perceived once the content has been acquired. But now it seems unavoidable that the causal relation we need to isolate be identified as the perception relation itself. The theory then is that experiences come to have their content by virtue of being *perceptions* of the property in question[…] an experience can only count as a perception *of* an *F* if it already has the content *F*.

That is, if the appropriate causal relation between an individual and a state of affairs is taken to be that individual's reliable perceptual experience of that state of affairs, then for the capacity to fix content, an individual first needs to know to what that content refers (i.e., to understand a property etc., a person must already understand that property etc.). As such, the intuitive notion of perceptual experience is an unsuccessful candidate for an appropriate causal relation and, before strong externalism can be adhered to without inconsistency, Johnson-Laird needs to first identify a valid and appropriate causal relation for reference fixation.[1]

However, weak externalism also seems to be problematic; it also seems to lead to unintuitive or inconsistent consequences. For example, compare A) a dream of a specific state of affairs, a particular sculpture say, with B) hearing a report or description of a real sculpture that just happens to possess the identical logical form as that from the dream. In both these cases a mental representation with an identical logical form, or relation structure, is generated; a mental representation that is the icon of an ontologically real fact. Now, if iconicity alone is said to be sufficient for reference fixation (i.e., if no further criteria for reference is established), this means that the mental model produced from A) is equally as meaningful as B) and both A) and B) are about the ontologically real fact. This is a situation that intuitively seems to be false, it seems instinctive that we do not want to hold a theory that concludes that made up fantasies, which just coincidentally share the logical form of real states of affairs, are actually meaningful locutions about that state of affairs, even if the individual that possesses the model denies they are (cf. an ant's drawing of Churchill in Putnam 1981, p1).[2]

Therefore, it seems that if Johnson-Laird accepts strong externalism, his iconicity needs to be developed: As the logical form alone is only necessary for reference fixation (i.e., not sufficient) and something further is needed (e.g., some valid and non-circular appropriate cause). However, this leads

---

[1] McGinn (1989, pp1-119) also offers further arguments against the notion of strong externalism other than those surrounding appropriate causal relationships. Additionally, for further issues concerning appropriate causal relationships in strong externalism, see the literature surrounding Kripke's *Naming and necessity*

[2] Some, such as Devitt (2014, pp270-271), may reject the use of intuition here. However, others argue for the validity for such an approach (e.g., Gendler 2010; Papineau 2014, pp186-189).

us into a number of difficulties, the least of which being to establish a valid definition of 'appropriate cause'.[3] Nevertheless, if Johnson-Laird rejects strong externalism and adopts a weak form of referentialism, he is tasked with explaining the unintuitive notion that certain signs (e.g., those made by individuals with no knowledge of states of affairs) are meaningful.

Summary

The first section of our Findings chapter found that the introduction of externalism into Johnson-Laird's philosophy of mind was successful in its aims (i.e., in averting the CRA and SGP). However, this update rises a couple of areas for discussion:

1) Our Findings chapter found that certain versions of the RTM are able to avoid the issues raised by the CRA and SGP by introducing language entry and language exit rules. However, it was also identified that the implementation of such rules requires extra-linguistic interaction with a system's environment (see Dennett 1980). This necessity entails that unembodied 'cognitive' systems, or bedridden AI, are not capable of grounding their symbols and, therefore, of possessing true semantic content. Consequently, to hold such a position concludes in accepting that only embodied systems (i.e., those with extra-linguistic perceptual systems) are capable of true cognition. Furthermore, the complex and embodied intercommunication with the world such extra-linguistic interaction necessitates implies that cognition must be spread over both these systems and the world (see McGinn 1989, pp100-107). Consequently, it seems that those who argue for a referentialist RTM (e.g., McGinn 1989; Brooks 1990; Harnard 1990b; Johnson-Laird 2002; Waskan 2006; Fodor 2010; Rapaport 2011; Schweizer 2012) implicitly accept the arguments from Bennett and Hacker (2003) concerning the mereological fallacy and have made steps to ensure that such an argument can no longer be used as ammunition against their philosophy of mind.[4]

2) Although it is clear that Johnson-Laird has introduced a form of referentialism into his philosophy of mind to avoid the issues raised by the CRA and the SGP, he has not yet clarified which type of externalism has been introduced. That is, he has not identified whether his iconism encompasses a strong or a weak externalism. This deficit is important, for if it is clarified that he holds the former, then he is tasked with clarifying a valid and consistent appropriate cause that establishes reference along with iconism. Alternatively, if Johnson-Laird adheres to a weak externalism, he must explain the unintuitive notion of icons referring to and meaning unknown and unintended states of affairs.

---

[3] Some of these difficulties are discussed by McGinn (1989, pp70-89).

[4] A rigorous analysis of the success of these steps is outside the purview of this research.

§2 Is Peirce's structural correspondence theory of meaning both internally and externally consistent?

From the second section of our Findings chapter, it was discovered that Johnson-Laird's appropriation of Peirce's correspondence theory of meaning not only inevitably leads him into the homunculus and symbolic fallacies, but also produces other, fatal, inconsistencies and deficiencies. As such, although the first section of our Findings chapter found that analogue representations must explain semantic content via some form of referentialism or correspondence theory, it is now not clear what *form* of correspondence it should take (i.e., not which *type* of externalism, not whether the externalism is of a strong or a weak nature), except that it cannot be along strict Peircean lines.[5] As such, it seems that an investigation into referentialist theories of meaning that can be validly and consistently applied to Johnson-Laird's analogical models, and an attempt to sketch a solution to the deficit, which the removal of Peirce from Johnson-Laird's philosophy creates, may be of some interest within these pages as well as for future research. Subsequently, this section will engage in a discussion of how the deficit created by the removal of Peirce may be rectified.

Alternative correspondence theories of meaning and a criteria of success

For Johnson-Laird's explanation of cognition and semantics to remain valid, how the analogous mental models at the heart of his theory connect with reality needs to be revised. That is, his correspondence theory of meaning needs to be updated. For this to be achieved, a replacement referentialist theory of semantics that aligns with Johnson-Laird's analogous mental models and that can be suitably grafted onto his existent theory without inconsistency and/or deficiency needs to be found. Many potential replacements can readily be identified within the literature (e.g. Russell 1918; McGinn 1989; Waskan 2006; Fodor 2010), however this discussion will not attempt to find or construct the most apt referentialist semantics for Johnson-Laird's theory of mental models. Rather, this discussion will limit its ambitions to constructing a criteria for success against which such theories can usefully be measured. To construct such a measure, we have looked to the findings from our investigations into correspondence theories of meaning and the hypotheses of Johnson-Laird to divine a raft of key criteria. Consequently, this thesis proposes that any prospective semantic explanation will only be considered a valid and productive contribution to Johnson-Laird's mental model theory if, and only if, it is able to answer in the affirmative to the following questions:

1) Is the theory based on a correspondence between word and world?

---

[5] Interestingly, this leaves us in a position Johnson-Laird was in the 1980s and 1990s, with the details and minutia of the relation between mental models and the world not being clear.

2) Does the theory explain how a specific proposition corresponds to a specific state of affairs through a congruent process akin to that which Peirce terms iconicity? That is, does the theory answer in the affirmative to the following:

   a. Do propositions *mirror*, are *analogous to* or are *isomorphic with* a state of affairs in the world?[6]

   b. Do propositions contain or mirror the same components that make up the relevant state of affairs?

   c. Do the relations between the components of a proposition mirror the relations between the parts found in its corresponding state of affairs?

3) Does the theory provide a full account of the foundation of meaning in counterfactuals?

4) Is the theory able to account for negation?

5) Is the theory free from any form of purely mental symbolism?

6) Is the theory absent from hypotheses which lead to the creation of other tensions within Johnson-Laird's theory as a whole (e.g. to his theory of reasoning etc.)?

Now this criteria of success has been proposed, it needs to be both demonstrated and tested. To this end, a raft of semantic theories propose themselves, however, this discussion will attempt to authenticate its measure by using a paradigm correspondence theory of meaning that can be found in one of the most influential philosophical works of the twentieth century: Wittgenstein's *Tractatus Logico-Philosophicus*. This work is famous for its introduction of the concept of the 'picture theory' of meaning into the philosophy of language and is an archetypal account of correspondence by congruence.[7] However, before we can begin an investigation into the picture theory of meaning's validity for Johnson-Laird, it should be noted there are a number of competing and highly divergent interpretations of the *Tractatus* and that the one that will be taken by this thesis can in no way be considered an uncontroversial reading of Wittgenstein. When looking to the relevant literature, Positivist (e.g. Carnap 1959) as well as Resolute (both weak (e.g. Diamond 1981; Conant 2000; Kuusela 2011) and strong (e.g. Rickets 1996; Read and Deans 2003; Goldfarb 2011)) readings of the *Tractatus* can be found alongside the 'standard' Ineffable interpretation this thesis will favour and which is

---

[6] Taking cognitive science (as defined by this research) and Johnson-Laird's conception of 'world'.

[7] Although, as Stern (1995, p36) points out, the standard translation of Wittgenstein's "*Bild*" is "picture", this may in fact be a misnomer that could create confusion. This is because the term "picture" inspires a potentially incorrect image of the operation of Wittgenstein's use of "*Bild*" that does not fully capture the isomorphic character it is said in 4.014 to possess. Therefore, Stern recommends that either the term "*Bild*" is maintained when discussing the theory, or that, following the usual translation of Hertz's *The Principles of Mechanics* (a work that greatly inspired the writing of the TL-P (see 4.04, 6.361)), the term should be translated into "model", which does capture the original nature It can easily be perceived why following this proposal would be preferable given the project of this chapter, but, as "picture theory" is the standard translation, confusion may be generated when quoting other works. As such, this discussion will follow convention and continue use of the incorrect "picture" phraseology.

exemplified by the likes of Hacker (1997) and Mounce (2003).[8] Finally, it needs to be highlighted that a connection between the works of Wittgenstein and Johnson-Laird is not a novel, nor a controversial, observation and has been articulated a number of times in the past. For example, both McGinn (1989, pp176-177) and Waskan (2006, pp171-172) have observed a relation between Wittgenstein and Craik (and therefore, by extension, with Johnson-Laird). Additionally, Johnson-Laird (2002, p80) has himself identified a connection between his own work and Wittgenstein's, stating:

> Early formulations of iconicity were Maxwell's (1910) analysis of diagram and Wittgenstein's (1922) "picture" theory of meaning[.]

This is a sentiment he has expressed on a number of different occasions (see also Johnson-Laird 1993, p4; 2005, p186; 2006, p435), but one which, unfortunately, misses the subtle differences between the nuanced positions held by the two correspondence theorists (Peirce and Wittgenstein).[9]

Now these preliminaries have been established, it may be enlightening to first outline Wittgenstein's picture theory before we compare it against our criteria of success. As such, this discussion will continue with a description of the ineffable *Tractatus Logico-Philosophicus* (hereinafter, the T-LP)*.*

Wittgenstein's picture theory of meaning

When first introduced to the ideas contained within the TL-P, it is important to be aware that the picture theory of meaning is based on two important hypotheses:

i)   Wittgenstein's insights into the logical structure of representation.

ii)   His metaphysical views on the structure of the world, views I will refer to, following standard phraseology from the secondary literature, as logical atomism.[10]

Language as representation

The germ for the particular theory of semantics which is the backbone to the TL-P can be found in a remark from Wittgenstein's notebooks made on 29/09/1914, in which it is stated:

> The general concept of the proposition carries with it a quite general concept of the coordination of proposition and situation: The solution to all my questions must be *extremely* simple.
>
> In the proposition a world is as it were put together experimentally. (As when in the law-court in Paris a motor-car accident is represented by means of dolls, etc.)

---

[8] This metaphysical reading of the TL-P is – pragmatically – being favoured as it is one that contains a complete correspondence theory of meaning and only because of this fact. This thesis is making no comment on exegesis nor authorial intent.

[9] Interestingly, Johnson-Laird has previously demonstrated an aversion to the work of Wittgenstein within his writings. An aversion identified by Wason (1996) and described by Garnham (1996, p51) as a "dislike".

[10] It should be highlighted that Wittgenstein may not agree with this designation due to certain theoretical divergences between Russell, from whom the name derives, and himself.

Here Wittgenstein believes he has hit upon a discovery which will eventually lead to an understanding of the nature of propositions and meaning.[11] In becoming aware of the relationship between language and the modelling process employed by the Parisian law courts, Wittgenstein has had the insight that both are equivalent forms of communication, that both are equivalent methods used to describe a possible state of affairs, but which are doing so via differing media (one by physical symbolic objects and the other by written or spoken symbolism). The inevitable conclusion of such a line of thought was the realisation that an investigation into how models and other common forms of depiction are able to describe the world will lead, *ex hypothesi*, to insights into how language is also able to do the same (4.014 – 4.016).[12] To express this line of thought succinctly: Wittgenstein's conclusion that language and representation are intimately related is initially derived from the following logical process:

(1) Communication is representation.

(2) All representations operate using the same laws of depiction.

(3) Language is a form of communication.

∴

(4) Language is a form of representation (following (3) and (1)).

∴

(5) Language must operate using common laws of depiction (following (4) and (2)).

∴

(6) An investigation into the laws of depiction will constitute an investigation into the nature of language.

Later, in the TL-P itself, Wittgenstein builds upon these thoughts, detailing a further argument to support his thesis that propositions are pictures and, therefore, that an investigation into representation constitutes an investigation into language. During phrases 4.016-4.021 and 4.03, Wittgenstein supports his hypothesis by appealing to the fact that a sentence which an individual has never previously encountered, can, unlike a name (4.026), but analogous to any other kind of commonly regarded picture (2.172), still successfully express its sense and is able to be meaningful even though that meaning has never been explained or defined for that person (4.02) (i.e. it is possible for old expressions to communicate a new sense (4.03)). That is, as part of the definition of any picture is that its sense is apparent to an individual without definition (4.172), then, following *modus ponens*, if propositions are able to do the same then they should be classified as pictures. Ergo, as language

---

[11] This particular story about the Parisian law courts and the genesis of Wittgenstein's picture theory entered the zeitgeist through von Wright (1955).

[12] All references structured in this way will, unless specified, allude to the numbered propositions found within the Pears and McGuiness translation of the TL-P.

does behave in this manner (4.021), it is therefore a mode of depiction (4.01). Finally, Wittgenstein also supports this classification of language as a form of depiction by appealing to the fact that the origins of written language are to be found in hieroglyphic script (4.016). This early form of symbolism operates more obviously via depiction than the symbolism of today, but what is not obviously apparent is that our current "alphabetic script developed out of [hieroglyphic script] without losing what was essential to depiction" (*Ibid.*). The implication of this being: Although our common scripts are not obviously built upon the laws of representation, we ought to look to those scripts' histories and origins in hieroglyphs to see that we should not be fooled by appearances, but instead be assured of their fundamental pictorial nature via analysis.

The nature of representation

These initial discoveries lead Wittgenstein into a deeper examination into the nature of representation more generally, which results in him gaining a greater appreciation of its variegated forms of instantiation, becoming aware that "not only paintings, drawings, photographs and other obviously pictorial representations[…] but also maps, sculptures, three-dimensional models, and even such things as musical scores and gramophone records" (Kenny 1975, p54) are examples of pictures. From an analysis of this broad range of exemplars, Wittgenstein draws three main properties common to all forms of representation, and which must be present within any medium of depiction (including language) for it to function as such. These are:

(i) That those individual elements that constitute a representation (e.g. a doll in a 3D model or a word in a proposition) must correspond to an individual element in the represented state of affairs (2.13-2.131).

(ii) That the relationship between the elements in the represented state of affairs must be mirrored in the relationship between their corresponding elements in the representation (2.14, 3.21).

and

(iii) That a representation must be able to *not* represent the world, i.e., must be capable of being false as well as being true (2.21).[13]

To clearly explicate the role and implications of these three properties in and upon language, it will be necessary to divide this section. As such, we will: 1) Looking into the above three properties' repercussions upon true propositions; and 2) investigate their repercussions upon false and negative propositions.

---

13 These three properties have also been articulated as Wittgenstein's analysis that representations must have 1) Content, 2) Structure and 3) Form (see McGinn 1999, pp500-501).

True propositions

The obvious implication upon language of function i) above is that a proposition (i.e. an interpretable sentence) "does not have meaning in itself, but requires some supplementation in order to acquire a complete meaning" (Russell and Whitehead 1999, p44). As such, propositions need to *correspond* with a state of affairs found in the world to acquire meaning. That is, a proposition's truth is determined by "a corresponding complex of the objects" of the proposition (*Ibid.*). Ergo, where language is concerned, for the Ineffable TL-P, truth is based on realist principles and the objective construction of the world. This is explained through the fact that propositions are not discrete individual units (3.143), but are in fact composite structures (2.0122), what makes a proposition a proposition is that it is built from individual simple words that correspond to, and therefore mean, simple objects in the world (3.203, 3.22) (due to this naming relation between sign and object, these base words are called 'names' (3.202)). But this raises the question: How, exactly, does a specific word name a specific object? How is a standardised correspondence between a name and an object established? Within the TL-P, Wittgenstein identifies that this link is established by a random naming relation, i.e. "we make an arbitrary determination" (5.473). Therefore, what correlates a sign to an object (making it a symbol) is a conscious act of christening by a person or group. Furthermore, Wittgenstein contends that for a proposition to be meaningful, it must not only be composed from individual mental entities that refer to an object, but the organisation – the logical form – of the representations in the proposition must mirror the organisation – the logical form – of the entities as they are found in the state of affairs in the world (2.15).

In summary then, names realise the first of the properties identified by Wittgenstein and the organisation of these names the second. That is, the relationship between names in a propositions mirror the organisation of the objects in the state of affairs referenced (3.14, 3.21).[14]

False and negative propositions

However, there appears to be an apparent tension with representations' final property, for how can (iii) now be possible without violating functions (i) and (ii)? How can a picture still be a picture if it is false and if its elements, by definition, then do not mirror what they represent? Wittgenstein realised that this issue only arises if we mistakenly think that a model as a whole behaves exactly like its elements (or, in the case of language, if a proposition's behaviour is akin to a name's). To solve this problem, we need to realise that the semantic behaviour of the two are not identical; we need to realise that, as Wittgenstein states in his *Notebooks* (p95), "a proposition [or a picture as a whole] is a

---

[14] Interestingly, this reliance on an 'arbitrary determination' for fixing the reference of names (i.e., property (i)) *and* logical form (i.e., property (ii)) makes the picture theory, technically, an amalgamation of correspondence by correlation *and* correspondence by congruence.

standard to which facts behave; with names [or pictorial elements] it is otherwise". As a result, whereas names (or pictorial elements) derive their meaning [*Bedeutung*] from a direct correspondence with an item found in the world (3.203), propositions (or whole pictures) get their sense [*Sinn*] by being compared to reality (3.3). This is why "names are like points: propositions like arrows" (3.144), names and other pictorial elements are defined and specific, whereas propositions and other pictures are freer (3.221). As such, whereas these aforementioned pictorial elements have to have a veridical connection with reality, propositions and full pictures do not; they are able to depict *possible* realities (2.201-2.203) which can then be compared to the world or, in the language of the TL-P, can be "laid against reality like a measure" (2.1521). Pictures, therefore, are not limited to how we find the world like names are, they can experimentally combine the world's simple objects, which do obtain, to generate states of affairs, which do not, thereby solve our apparent problem of how pictures are capable of being false. This former fact (that it is names which provide meaning [*Bedeutung*]) can be seen in that both a proposition and its negation have the same constituents and therefore the same meaning [*Bedeutung*] (4.024), yet, although *p* and not-*p* mean the same, they have a completely differing sense that, after being compared to reality (2.223), can be either true or false.

This distinction, between the behaviour of names and propositions, is founded upon work we have already encountered by Gottlob Frege, who, whilst realising that the object a name refers to (i.e. its referent) provides it with its meaning, identified that there are situations where reference [*Bedeutung*] is not enough, where sense [*Sinn*] is also required. As such, along with a metaphysical constituent for semantic content operating within the picture theory of meaning, a psychological aspect is introduced, i.e. Wittgenstein's picture theory employs both narrow and broad content. It is important to note here, though, that Wittgenstein does not view the meaning of pictures (i.e. propositions) or its elements (i.e. names) as being entirely and independently constituted by either its sense (narrow content) or referent (broad content) respectively. It is just meant that for the provision of semantic content, it is necessary that *both* sense [*Sinn*] (2.221) and referent [*Bedeutung*] (3.203) are present. When Wittgenstein is translated as saying that "only propositions have sense" (3.3), this should not be taken to mean that he holds the opinion that names are a separate entity who derive their meaning totally from their referent and have nothing to do with sense. This is seen in the original German, where it is stated: "Nur der Satz hat *Sinn*; nur im Zusammenhang des Satzes hat ein Name *Bedeutung*" (my emphasis). Implying, in accordance with the context principle from Frege (1980, px) (i.e. "never to ask for the meaning of a word in isolation, but only in the context of a proposition"), that words are meaningless while on their own (irrespective of either sense or reference). The point here is that both *Bedeutung* and *Sinn* are only present in "the nexus of a proposition" (3.3). That is, in

the words of Kenny (1973, p5), "the understanding of names and the understanding of propositions stand or fall together". Instead of any putative independence, propositions actually have both meaning and sense [*Bedeutung und Sinn*]: Its constituents provide the meaning and the picture as a whole its sense (3.3).

All this provides the groundwork for an explanation of *negative* statements. Wittgenstein employs this distinction between sense and reference to explain statements such as 'There is not a car in this room', by arguing that this type of proposition refers to the same state of affairs as its un-negated counterpart ('There is a car in this room'), but asserting that the two propositions possess a different sense (4.0621). Wittgenstein further states that this difference in sense is created by an operation that reverses the truth-values of the elementary proposition (5.2341), thereby reversing the false sense of 'There is a car in this room' to the true statement 'There is not a car in this room'. However, what needs to be answered here is: What *signifies* that this operation is to be performed on the sense/truth functions of a proposition? That is: How do we know when an operation is to be carried out? Looking to the TL-P we find:

> **4.0621** But it is important that the signs '*p*' and '*~p*' *can* say the same thing. For it shows that nothing in reality corresponds to the sign '~'.
>
> The occurrence of negation in a proposition is not enough to characterize its sense '~~p = p'.
>
> The propositions 'p' and '$~p$' have opposite sense, but there corresponds to them one and the same reality[...]
>
> **5.2341** The sense of a truth function of p is a function of the sense of p.
>
> Negation, logical addition, logical multiplication, etc. are operations.
>
> (Negation reverses the sense of a proposition.)

Consequently, it appears that it is the *symbol* '~', or 'not' in natural language, that signifies that an operation on the sense of a proposition is to be performed. As such, Wittgenstein, like Peirce, employs a sign for negation.[15]

---

[15] There is one further conclusion that this necessity for a proposition to possess both meaning and sense leads to and that may be of interest. This reliance on sense, and the need that a proposition must be compared to reality to uncover its truth-values, suggests that only declarative sentences, i.e., those with truth values such as 'The book is on the desk', are covered by Wittgenstein's theory of semantics. Although this may be the case, a lot of seemingly meaningful communication relies on non-declarative sentences. For example, statements such as 'Shut the door!' or 'Is the door shut?' seem to be meaningful yet they cannot be said to be either true or false. Consequently, Wittgenstein's use of sense in this regard seems to exclude a large number of generally accepted propositions from meaningful discourse. A possible solution to this deficit may be found in Lewis (1970), who argues that such non-declarative sentences can be, after analysis, reduced to declarative sentences by the use of performative verbs. For example, 'Shut the door!' will be reduced to 'I tell you to shut the door' and 'Is the door shut?' to 'I ask you if the door is shut'. This is an approach that accords with comments found within the TL-P itself:

> **4.002**[...] Language disguises the thought; so that from the external form of the clothes one cannot infer the form of the thought they clothe, because the external form of the clothes is constructed with quite another object than to let the form of the body be recognized.

However, along with true and false propositions, which can *say* their meaning, Wittgenstein identifies a further form of proposition: Those that *show* their meaning. This term, that of 'showing' or of being 'shown', highlights a theory at the very heart of the TL-P and one whose understanding is vital if the message of that work is to be comprehended. In fact, Wittgenstein said of the doctrine in his *Letters from Cambridge* (p98):

> The main point is the theory of what can be expressed (gesagt) by prop[osition]s – ie by language[...] and what can not be expressed by prop[osition]s, but only shown (gezeigt); which, I believe, is the cardinal problem of philosophy.

To clarify, this cardinal problem of philosophy, and arguably the main point of the TL-P, concerns those propositions that attempt to express things that are fundamentally *ineffable* and entails a number of sub-theses that deserve explication in their own right.

Saying and Showing

When gaining an understanding of what the concept of 'showing' is, it is important to appreciate that its meaning only becomes clear when we see it in its right context: On one arm of a dichotomy opposed to 'saying'. That is to say, within language there are two ways in which items can be expressed: Those that can be said and those which can only be shown. The character of the former type (things that can be said) is quite apparent: They are the things which can be detailed and described using the operation of language and words we have detailed above. As for the latter (those that can only be shown), this is not the case: They cannot be put into words. As such, when we wish to express such thoughts, we must employ nonsensical language in such a manner as to portray these ineffable concepts. Consequently, these are merely pseudo-propositions that "say nothing" (4.461), instead the very structure of the language used to express them (i.e., logical form) informs us about those aspects of the world they are attempting to convey. In this sense, the expressions are informative as they let us 'read off' such things as logical truths from them.[16] This is the view that, according to Hacker (2001, pp14-15), certain sentences:

> fail to conform with the logical syntax of language and are a self-conscious attempt to say what can only be, and indeed is, shown by features of the relevant symbolism[…] hence they lack sense. They are not nonsense, but senseless, i.e. have, in a manner of speaking, zero sense.

Or, as Wittgenstein expresses it in 'Some remarks on Logical Form' (p169):

---

[16] Within the literature the term 'pseudo-proposition' is used to denote a string of words which *appears* to be a united and interpretable declarative sentence that can communicate a sense [*Sinn*] from one individual to another but which in fact is not. That is to say, a pseudo-proposition is a meaningless string of words which only *appears* to be a proposition. Although these forms of proposition appear to be readily interpretable (i.e. can easily be mistaken to express a sense and contain semantic content) on first appearance it is important to recognise that after analysis it is revealed that such strings in fact are not (i.e. they express no sense and are linguistically meaningless).

the forms of the entities are contained in the form of the proposition which is about these entities. For the sentence, together with the mode of projection which projects reality into the sentence, determines the logical form of the entities.

Nonsense

To fully appreciate this operation of showing, it will be illuminating to analyse Wittgenstein's nuanced use of the term 'nonsense' in the TL-P. Within the work, Wittgenstein uses a number of terms to signify nonsensical propositions:

1) *Sinnlos***:** Denote expressions that are best described as 'saying nothing'. In the TL-P the phrase is used in connection with logic, tautologies and contradictions (see 4.461; 5.132; 5.5351). As such, it appears to be a 'positive' form of nonsense, identifying propositions which, because of the subject matter, are not able to possess sense.

2) *Unsinn/unsinnig***:** Best described as 'utter nonsense'. More commonly used in the TL-P than *Sinnlos*, *Unsinn/unsinnig* appears to be the harshest distinction of a nonsense term and refers to expressions which are just plain, straightforward, nonsense (for *Unsinn* see p4; 5.5303; 5.5351; 5.5422; 5.571, for *unsinnig* see 3.34; 4.003; 4.124; 4.1272; 4.1274; 4.4611; 5.73; 5.5351; 6.51; 6.54).[17]

As such, although some propositions do not picture a real or possible state of affairs, some do so while being *Sinnlos* while others are *Unsinn/unsinnig*. Both types of propositions are technically nonsensical, however the latter can tell us nothing (e.g., "Socrates is identical" (5.473)), while the former's logical form can *show* us certain truths that cannot be put into words (e.g., "it is either raining or not raining" (4.461)).

The operation of representation

While Wittgenstein covers *what* representations do, he also explains *how* representations are able to do so. From this investigation Wittgenstein uncovers a range of characteristics which any representation must possess in order to represent. These are:

4) **Have structure**: The particular connexion of axiomatic elements that comprise a representation, e.g. the relationship between the dolls in the Parisian law court (2.15).

5) **Possess pictorial form**: The way a depiction's structure mirrors that of a state of affairs found within the world. For example, a model mirrors the three-dimensional and spatial relation of a scene's objects, while a painting depicts the relation of its colours (2.15-2.17).

---

[17] However, some read Wittgenstein's use of 'nonsense' as more nuanced than is presented here. These theorists, such as Lugg (2003, p343), argue that *unsinng* propositions do not express sheer nonsense, but are in fact a form of showing, averring that propositions are:
> *sinnlos* when taken to 'represent [*stellen*]' the 'scaffolding of the world', and as *unsinnig* when taken to 'describe' it (6.124)[...] in both case the thoughts in question have to be presented, displayed, exhibited (rather than represented, asserted, said).

6) **Possess representational form**: The media from which a picture is created. For example, a painting has a representational form based upon colour and an architect's model one from spatial extension (2.173).

7) **Possess logical form**: The fundamental structure which any depiction must have in order to represent a state of affairs via any medium (i.e. via any pictorial and representational form). For example, a sculpture, a painting and a sentence that all picture the same state of affairs have different pictorial and representational forms, but all have the same logical form. That is how they are all pictures of the same event, yet still picture differently (2.18).

8) **Be a logical picture**: Be a representation whose pictorial form *is* logical form. Therefore, as every picture that represents the world must do so using logical form (no matter its pictorial form), every picture is, perforce, a logical one (2.181-2.182).

9) **Possess logico-pictorial form**: Logico-pictorial form is what everything that can be considered a representation must have, to be a picture *qua* picture. It is a conflation expressing the necessity that before a representation can represent it must first possess logical form and some sort of pictorial form in common with that being represented, thereby highlighting the need for the possession of all five previous functions to represent (2.2).

Therefore, the conclusions that Wittgenstein draws from his investigation into the operation of picturing, are that for an image's elements to mirror those of a state of affairs, its representational form must show its fundamental structure via its pictorial form. This fundamental structure must, in turn, possess logical form in common with the aforementioned state of affairs. If this is the case, the representation has attained logico-pictorial form, is considered a logical picture, has a sense and successfully depicts a portion of the world (either rightly or wrongly).

Summary

A summary of Wittgenstein's analysis of representation can be found in Kenny (1975, pp62-63) who highlights the following eight crucial theses that are the result of the TL-P's analysis:

(1)   A proposition is essentially composite.

(2)   The elements which compose a proposition are correlated by human decision with elements or reality.

(3)   The combination of such correlated elements into a proposition presents – without further human intervention – a possible state of affairs.

(4)   A proposition stands in an internal relation to the possible state of affairs which it presents.

(5)   This internal relationship can only be shown, it cannot be informatively stated.

(6)   A proposition is true or false in virtue of agreement or disagreement with reality.

(7)   A proposition must be independent of the actual state of affairs which makes it true or makes it false

(8)   No proposition is *a priori* true.

Once this, primarily congruent or isomorphic, operation of language has been analysed and detailed, the next question that the picture theory needs to answer is: What, exactly, is it that these representations are depicting? That is: What are these worldly objects with which names are supposed to be correlated? For, it is obvious that the TL-P's account of meaning can only be accurate if it investigates and explains that which is represented along with representation. These are questions whose answers can be found within the TL-P by a metaphysical doctrine identified within the secondary literature as the theory of 'logical atomism'.

Logical atomism

Now we have explained the first of Wittgenstein's conceptions we identified as comprising the central hypotheses of the TL-P, we can move on to detailing his second: His metaphysical theory of the make-up of reality, i.e., of the composition of the world. Within the secondary literature, this theory has been labelled 'logical atomism'. As previously stated, this appellation is one taken from the philosophical writings of Wittgenstein's mentor Bertrand Russell and, as such, may be one that leads to some confusion in some readers over the ontology of the TL-P and may even be one that Wittgenstein would reject. However, no matter this debate, this section will focus solely upon illuminating the metaphysical and ontological theses and arguments found in the TL-P. This shall be achieved by:

1) Covering Wittgenstein's physicalist metaphysics.

2) Explaining the atoms that are the foundation of this physicalism.

3) Elucidating the nature of states of affairs that are generated by these atoms.

4) Discussing the metaphysical status of relations, properties and concepts.

5) Detailing the types of state of affairs a proposition can mirror.

Wittgenstein's metaphysics

Wittgenstein's metaphysical views are, sequentially, the first to be covered in the TL-P, occupying the focus of the propositions numbered from 1 to 2.1.[18] The most fundamental of these are the first two of the main body of the work:

---

[18] Interestingly, this positioning of Wittgenstein's metaphysics in the TL-P may, in itself, be somewhat illuminating. For, the debate over whether the work is realist (i.e., states that meaning is derived from the world – see Hacker 1997) or anti-realist (i.e., that we only know the world via language – see Ishiguro 1969) essentially hangs on causational order (i.e. does knowledge of the world lead to a knowledge of language or does a knowledge of language lead to knowledge of the world) and that Wittgenstein chose to put comments concerning the

1 The world is all that is the case.

1.1 The world is the totality of facts, not of things.

These laconic sentences establish Wittgenstein's materialist ontology, which is fundamental to the rest of his thought within the TL-P. These two aphorisms establish that there are two types of entity to be found in the world, facts and things, from which Wittgenstein moves on to state that, although there is this conceptual dualism, it is based on an ontological monism. This can be seen when examining the following quotes:

2 What is the case – a fact – is the existence of states of affairs.

2.01 A state of affairs (a state of things) is a combination of objects (things).

This shows us that, for Wittgenstein, all that exists are physically real objects and their relations to one another (for how are non-physical 'things', such as some form of Cartesian soul, able to combine with those of a physical nature?) Wittgenstein supports this hypothesis by showing that as relativism is not the case and reality is objective (employing a *reductio ad absurdum* argument to show that relativism can only inevitably slide into nonsense (2.0211-2.0212)), there must be something providing the world's fixed nature and that this substance that provides reality's form must itself be unchanging and fixed, i.e., that there must exist simple elementary and unchanging objects to provide the subsistence to the world's unalterable form (2.023 and 2.026-2.027). This chain of reasoning can be summarised as follows:

(1) The world is objective (2.0211-2.0212).

(2) The form of physical objects is unchanging (2.027).

∴

(3) The world's form is unalterable (Following (1), see 2.022).

∴

(4) Some substance must underlie and provide this unalterable form (Following (3), see 2.024).

∴

(5) Simple (i.e., atomic) physical objects provide the substance of the world (following (2) and (4), see 2.023 and 2.026-2.027).

Wittgenstein further supports this axiomatic ontology concerning objects by appealing to the fact that as the world is composed of molecular facts (states of affairs), these facts have to be broken down, but by following this course of action we must reach a point where reality can no longer be decomposed, for otherwise an infinite regress will result. Simple objects are said to be precisely those

---

world before those concerning language seems to suggest that he views knowledge of the world as prior to that of language, thereby implying a realist metaphysics.

physical entities that provide this metaphysical foundation (see 2.0201-2.021). This is a view which can be viewed as a foreshadowing of Wittgenstein's own observation from the *Philosophical Investigations* (hereafter, the PI) that "explanations must come to an end somewhere" (PI 1).[19] Accordingly, the above logical process culminates in one final proposition:

∴ (6) Objects cannot be molecular; must be atomic (Following (5), see 2.021).

Atoms

Although facts, or states of affairs, hold more importance in Wittgenstein's ontology (1.1), it is important to begin an elucidation of the metaphysics of the TL-P by first focusing, and shedding light, upon the elements that compose those facts, i.e. on objects or things. As such, the fundamental concept that underlies the whole of Wittgenstein's view on the make-up of the world is the belief that reality is constructed out of the discrete atoms, or 'simples', our last section's analysis resulted in (2.02).[20]

However, what still needs to be clarified is: What, for the TL-P, *are* these simple objects which are the backbone of Wittgenstein's ontology? Are they the everyday medium-sized objects with which people usually interact (such as tables and chairs)? Are they the microscopic atoms of physics that are now known to make up these medium sized objects? Or, are they some form of strange and foreign entity found in the world of the quantum? Of these questions Wittgenstein is unfortunately elusive, never providing a definitive answer. Throughout the TL-P, Wittgenstein uses a range of entities as *examples* of what constitute an object: From sense-data (2.0131) to everyday objects (3.1431).[21] Nevertheless, Wittgenstein's ultimate position on the matter in the TL-P is to assert that, as we are only aware of the existence of objects via *a priori* reasoning, any talk of examples would be both nonsensical and groundless and, in fact, "to ask whether a formal concept exists is nonsensical. For no proposition can be the answer to such a question" (4.1274). This *a priori* reasoning is one which originates in thoughts found in the *Notebooks 1914-1916* (p60), where Wittgenstein expresses the idea in a less terse manner:

> It seems that the idea of the SIMPLE is already to be found contained in that of the complex and in the idea of analysis, and in such a way that we come to this idea quite apart from any examples of simple objects, or of propositions which mention them, and we realize the existence of the simple object--*a priori*--as a logical necessity.[22]

---

[19] All references structured in this way will, unless specified, allude to the numbered propositions found within the G.E.M Anscombe translation of the PI.

[20] These simple entities, which are the building blocks of the world, are the atoms that provide the theory as a whole with the 'atomism' of its title.

[21] See also PI 47

[22] There is, however, a residual issue here: If the atoms of the world are in fact smaller than the average sized objects to which we attach names and commonly discuss via propositions, then those names actually signify a state of affairs and therefore must be classed as a

States of affairs

As previously mentioned, the importance of these simple atomic objects is not found in their presence alone but in their relations with others and the resulting structure this creates, i.e., objects are not ends but means (1.1). This is a point eloquently discussed by Griffin (1964, p31):

> The world is not just things; we would not characterise the world by listing objects. We would not, I think, because[…] objects are unalterable: they are unalterable both in the sense that they remain the same through change [2.0271] and in the sense that they are present in any possible world [2.022 and 2.023]. That is, no matter what changes we imagine something's going through, the objects involved are what have not changed; and no matter how we imagine the world as different from what it is, the objects will be common both to this and the imagined world. If being unalterable in this sense is a necessary condition of being an object, then of course it is of no help characterizing a world to list the objects. Such lists would be identical for all possible worlds, so nothing characteristic could be given by them.

As such, it is in an important – and, indeed, the central – facet of these axiomatic components of the world that they are able to combine, relate or form bounds with their peers, without which the molecular composition of the world (both as we know it and its potential counterfactuals), would not be possible (2.0121 and 2.0141). This last observation is, in itself, one important enough to the picture theory to need to be highlighted one last time: A state of affairs, a situation or a fact is generated from/supervenes upon nothing but relations between atomical objects (2.01). From here, further complex molecular facts are generated through the relations between the more fundamental states of affairs, which in turn are related to one another to create yet more multifaceted facts, and so on until we reach the highest level, the world in totality (see 4.221-4.2211 and 1.1).

Relations, properties and concepts

This previous discussion has introduced something new into our ontology that we must clarify before we can continue: That of relations. When discussing the TL-P's view on relations, it is important to highlight that, as 4.12, 4.26 and 4.1272 make clear, we are also simultaneously talking of properties and concepts as the three operate on similar principles. Therefore, a discussion of one is able to constitute, with minor alterations, a discussion of all three. Where relations are concerned, the TL-P states:

> **3.1431** The essence of a propositional sign is very clearly seen if we imagine one composed of spatial objects (such as tables, chairs, and books) instead of written signs.
>
> Then the spatial arrangement of these things then will express the sense of the proposition.

---

proposition, not a name. But, according to Frege's context principle (3.3), a proposition cannot be a proper name. A potential solution to this problem may be found at 5.26, wherein it is stated that "We can describe the world completely by means of fully generalised propositions, i.e. without first correlating any name with a particular object[…] we simply add, after an expression like, 'There is one and only one *x* such that …', the words, 'and that *x* is *a*'". This means that, if average sized structures such as tables and chairs are composed from a multitude of atoms, then what we consider names for them are actually generalised propositions of the nature Wittgenstein describes. When we subject the names for objects such as tables and chairs to analysis, it is discovered that they are actually structured like the example given, thereby naming and showing the relations of the simples that make up our table or our chair.

3.1432 Instead of, 'The complex sign "*a*R*b*" says that *a* stands to *b* in the relation R', we ought to put '*That* "*a*" stands to "*b*" in a certain relationship says *that a*R*b*.'[…]

4.122 […] It is impossible, however, to assert by means of propositions that such internal properties and relations obtain: rather, this makes itself manifest in the propositions that represent the relevant states of affairs and are concerned with the relevant objects.

Meaning that, within a proposition, if the corresponding sign for object *a* is found to the left of the propositional element 'bigger than', while the name for *b* is to the right of the same element, this then *shows* (i.e. does not *say*) that *a* is bigger than *b* in the exact same way it would in a painting if *a* occupied more space than *b.* The only divergence in the two instances (between language and painting) is that the differences in the representational form employed by language can obscure this function of its elements.[23]

Language and the world

Now that the mechanics behind the substance of the world as we perceive it have been made clear (i.e. that states of affairs are generated by relationships between unchanging axiomatic physical entities), we are now in a position to clarify the three possibilities what Wittgenstein has uncovered produces:

1) **Potential states of affairs that do obtain:** These are situations/relationships that are both allowable by the objects in the world and are actually present. That is to say, the real facts found in the world or all that is the case. These facts are known as the true. For example, the fact that 'The Earth is closer to the sun than Mars' (See 1-1.12; 2.04).

2) **Potential states of affairs that do not obtain**: These are situations/relationships that are allowable by the objects in the world but which are not actually present. That is to say, the potential but non-existent facts, or all that is not the case. These facts are known as counterfactuals or the false. For example, the fact that 'The Earth is farther from the sun than Mars' (See 1.12; 2.022; 2.05; 3.24).

3) **States of affairs that cannot obtain:** This refers to situations that cannot exist through the limitations of the objects found in the world (i.e. those that denote relationships that are impossible to create with the given physical objects). These facts are not false, as may first be assumed, they are unimaginable, they are utter nonsense [*Unsinn/unsinnig*]. For example, 'Earth and Mars are prime numbers' (See pp3-4; 4.003).

---

[23] Although this putative non-reality of relations, properties and concepts has historically been attributed to the picture theory by some (see Anscombe 2001), others disagree with this, stating instead that Wittgenstein held a realism concerning such phenomena and argue that the TL-P holds that relations etc. can be referred to and *said* in propositions (see Hacker 1997, pp67-72).

Summary

All these explanations, taken in conjunction, should furnish a reader with the answers to the questions with which our prior section on 'Language as representation' closed. That is, they should supply an understanding of what it is, for the TL-P, that language is representing. As we have seen, language is said to depict the world, which itself consists of states of affairs that supervene, ultimately, upon axiomatically simple objects. As such, within the nexus of a proposition, names depict these simple objects (therefore providing meaning [*Bedeutung*]), and propositions as a whole depict a fact, providing sense [*Sinn*]. To put it succinctly: Within language, or the propositions that comprise it, names represent or picture objects (3.202-3.203), relations between these words draws the relations between the objects in the world and propositions as a whole represent or picture facts or state of affairs (3.21), while properties containing relations, properties and concepts can only be shown (3.1431-3.1432; 4.122).

The picture theory, Johnson-Laird and our criteria of success

Now that this correspondence theory of meaning has been defined, we are in a position to lay it against our criteria of success and attempt to establish whether or not the picture theory would make a suitable addition to Johnson-Laird's theory of mental models. Once this is done, the following is found:

1) **Is the theory based on a correspondence between word and world?**

   Yes. The picture theory relies upon a correspondence between word and world (see 2.1511; 3.22).

2) **Does the theory explain how a specific proposition corresponds to a specific state of affairs through a congruent process akin to that which Peirce terms iconicity?**

   Yes and no. Although it relies on correspondence by correlation to align the atomic 'names' of language with the 'atoms' of the world, the picture theory also operates through a process reminiscent of Peircean iconicity. This latter point can be seen in its answer to the following questions:

   a. **Do propositions *mirror*, or are *analogous to,* or are *isomorphic with* a state of affairs in the world?**

      Yes. It relies on a form of congruence between proposition and reality, termed, in the language of the TL-P, as 'logical form' (see 2.151; 2.16; 2.161; 2.18).

   b. **Do propositions contain or mirror the same components that make up the relevant state of affairs?**

Yes. It explains logical form via propositions containing the same components as the relevant state of affairs (see 2.13; 2.131)**.**

c. **Do the relations between the components of a proposition mirror the relations between the parts found in its corresponding state of affairs?**

Yes. The picture theory completes its account of logical form through the components of the propositions mirroring the relations of the components of the state of affairs (see 2.14; 2.21)**.**

3) **Does the theory provide a full account of the foundation of meaning in counterfactuals?**

Yes. It relies on the Fregean concept of sense and reference to explain how counterfactuals are still meaningful while not corresponding to a specific worldly state of affairs (2.201-2.203; 3.144).

4) **Is the theory able to account for negation?**

Yes. It uses Fregean sense and operations on truth values to explain negation. A negative proposition corresponds to the exact same state of affairs as its un-negated counterpart. A negating symbol signifies that an operation has been performed on the proposition to reverse its truth functions (4.0621; 5.2341).

5) **Is the theory from any form of purely mental symbolism?**

No. Following from 4), the picture theory uses a sign to negate propositions.

6) **Is the theory absent from hypotheses which lead to the creation of other tensions within Johnson-Laird's theory as whole (e.g. to his theory of reasoning etc.)?**

No. It relies on Fregean sense as well as reference to account for meaning (2.201-2.203; 3.144). Therefore, the picture theory employs both narrow and broad content. Furthermore, as Putnam (1988, pp19-56) shows, narrow content constitutes an internalist account of content, which, as we have already seen, is antithetical to Johnson-Laird's externalist correspondence theory of meaning and leads to – now familiar – problems (e.g., "the meaning of a symbol cannot be another symbol").[24] Additionally, the picture theory's use of purely mental symbolism (see 5)) clashes with Johnson-Laird's naturalism constraint. Finally, the atomic concept of language adhered to by Wittgenstein, wherein propositions are composed

---

[24] In fact, such difficulties with narrow content are what led Fodor (2010) to abandon his commitment to a dualistic notion of content in place of a monistic, purely broad, conception.

from 'names' that are *correlated* with atoms in the world, is antithetical to the *analogical*, i.e., non-decompositional, mental models.[25]

Consequently, according to our criteria of success, the ineffable reading of Wittgenstein's TL-P will not make a valid, productive nor successful inclusion into the work of Johnson-Laird and will not save the latter from its deficiencies. For, Wittgenstein's picture theory of meaning:

1) Uses a sign for negation.

2) Employs both narrow and broad content.

3) Gas an atomic concept of language.

These are issues because, as we have seen, 1) leads to the homunculus and symbolic fallacies and clashes with Johnson-Laird's naturalism constraint; 2) creates a dissonance with Johnson-Laird's externalist semantics and counteracts his motivations for initially adopting a correspondence theory of meaning; and 3) clashes with his analogical RTM.[26]

Criteria of success tested

Although the above attempts to demonstrate our criteria of success, we have not yet endeavoured to perform our second task: Showing that the criteria are an accurate determinant of a theory's validity and its productive contribution to Johnson-Laird's philosophy of mind and language. Although, as Popper (1957; 1968; 1974; 1994) shows, proving this is not possible, we may be able to use others' research to corroborate our findings and maintain our criteria as a 'provisional conjecture'. To do this, we may be able to look to Jonathan Waskan to support our findings. When we do this, we find that Waskan (2006, pp171-172; p292) similarly dismisses Wittgenstein's theory of meaning as it does not meet his own criteria of success for a semantics of analogical models. For, the picture theory does not manage "to simultaneously (i) maintain consistency with basic brain facts and (ii) support a distinction between sentential and imagistic representations" (*Ibid.*). As such, it seems that it may be possible to establish some confidence in our criteria of success and its potential usefulness in future research. However, due to the minimal scale of this argument, future research into both the success of our criteria of success and their application with other correspondence theories of meaning is still recommended.

---

[25] However, if this latter facet of the picture theory is translated as an 'arbitrary' correlation between natural language terms and mental models, the difficulties highlighted here may be dissolved.

[26] Again, this last point may be dissolved if the naming relation in the picture theory, when introduced to Johnson-Laird's thought, is stipulated as being between natural language terms and 'arbitrary' parts of mental models.

Summary

Having recommended the abandonment of Peirce's iconism from Johnson-Laird's philosophy during our Findings chapter, this discussion has attempted to in some way rectify the deficit such a move creates. In doing so, we have proffered, demonstrated and tested a criteria of success for a replacement correspondence theories of meaning, using Wittgenstein's picture theory as an example. From this process, it was determined that: a) The metaphysical TL-P will not work as a valid nor a productive addition to Johnson-Laird's philosophy; and b) our criteria of success stands as a 'provisional conjecture'.

§3 Is the introduction of a correspondence theory of meaning into Johnson-Laird's broader philosophy of cognition logically consistent?

After investigating the consistency of Johnson-laird's referentialist RTM, the third and final section of this thesis' Finding chapter concluded that this correspondence-based incarnation of the mental model theory is inconsistent. However, this investigation and this conclusion raise a couple of interesting points that merit attention here. These points concern:

1) A concern over the implicit commitments a correspondence theory of meaning necessitates, regardless of whether these commitments are, or can be made to be, consistent with Johnson-Laird's wider philosophy.

2) The implications this finding of inconsistency may have on Johnson-Laird's RTM project as a whole.

Unacceptable commitments

First, one concern that may be raised regarding the third section of our Findings chapter relates to those commitments entailed by a correspondence theory of meaning. For, even those assumptions that we have determined to be consistent with Johnson-Laird's philosophy may be, in their own right, unacceptable or too unintuitive for Johnson-Laird, or any other theorist, to freely admit into their philosophy without reservation. For example, it was found that the following hypotheses are both implicit to corresponded theories and consistent with Johnson-Laird's philosophy:

a)   A commitment to externalist semantics.

b)   A rejection of internalist hypotheses and arguments.

c)   An ontological realism concerning states of affairs/facts.

d)   That a reliable and consistent mode of reference fixation is possible.

e)   An epistemic scepticism of meaning.

f)   An eternalist stance on the metaphysical status of time.

g)   That properties and relations have some form of ontological reality.

h)   That two signs who possess the same referent have the same meaning.

i)   That meaning is mutable.

j)   That the apparent informativeness of content found in substitutivity/Frege cases can be explained.

k)   That the challenges faced to the implementation problem by Twin and Frege cases can be resolved.

l)   That slingshot arguments can be answered.

However, this list includes some rather daunting positions many would find hard to accept, or be forced to accept (e.g., eternalism, determinism and an epistemic scepticism of meaning). This is an issue that needs to be highlighted within this thesis, for consistency alone may be no measure of the true acceptability of such a referentialist program as is undertaken by Johnson-Laird. That is to say, it may be the case that any correspondence theory of meaning introduces too many unpalatable metaphysical commitments into a philosophy and, for this alone, may well be found to be unacceptable by some.

Moreover, one reaction to this research's findings may be the adaptation of Johnson-Laird's current philosophy to make it consistent with those commitments of externalism that were identified as creating dissonance. However, these implicit, inconsistent, hypotheses (i.e., that a) the limitations of genre or pictorial form can be accounted for, that b) the one-way nature of reference can be explained, c) that Brentano's question/problem can be answered and d) an adherence to determinism) again include a number of somewhat controversial positions that may be simply unacceptable to some. Consequently, although it is possible that Johnson-Laird's philosophy can be made consistent with the implicit hypotheses entailed by a correspondence theory, these theories in their own right may not be conscionable. That is, correspondence, in its own right, may contain too high a level of residue difficulty to be an acceptable theory of meaning.

Degenerative science

This research's findings may be, for a number of reasons, far more troublesome for Johnson-Laird than they first appear. Firstly, our Findings seem to show that although Johnson-Laird's research programme is adapting to problematic facts, that seems to be all that it is doing: It seems to possess only a focus on difficulties and is engaged in only a reactionary response to those difficulties. Secondly, it was found that Johnson-Laird is engaged in a reactionary response that is making no further hypotheses but only proliferating further inconsistencies. These may be issues for Johnson-Laird, for under the sophisticated falsificationist epistemology of Lakatos (1984), if a research programme has ran out of explanatory power and is only concerned with anomalous facts, that programme is said to

184

be a degenerative science and its abandonment is recommended. As such, the increased inconsistency and problematic facts identified by this research in the current phase of Johnson-Laird's theory may lead to more serious conclusions than simply the need to make further updates.

Before we continue, however, a brief explication of Lakatos' epistemology is required. To clarify, Lakatos' sophisticated falsification is founded on the distinction between progressive and degenerative research programmes; where a research programme is defined as a succession of continuous theories that are adaptations to known facts, which in turn should anticipate, i.e., predict, novel ones. Each consecutive replacement theory constitutes a problemshift within a research programme (*Ibid.*, pp38-47). To retain a continuity between successive theories and ensure the one theory constitutes a problemshift to one research programmes but not another, a research programme entails two factors (*Ibid.*, p48):

1) A hard core of central theories that cannot be changed.
2) A protective belt of hypotheses that surround the hard core and that can be updated so as to protect the hard core.

For a theory to be progressive, "each step [in a research programme must be…] consistently content increasing" (*Ibid.*, p49); that is, its problemshifts must have "heuristic" (i.e., explanatory) power (*Ibid.*, p69). If this is not the case, then the problemshift and/or the programme are deemed to be degenerative and its abandonment recommended (however, an otherwise progressive research programme may have degenerative phases or problemshifts within its progress (*Ibid.*, pp68-69; p71)). A degenerative programme may be identified via a number of "typical signs" (*Ibid.*, p77):

(1) Is instantiated by problemshifts with no explanatory/predictive power (*Ibid.*, pp68-69).
(2) A focus on only anomalous facts (*Ibid.*, p68).
(3) A feeling of a Kuhnian 'crisis' (*Ibid.*).
(4) A proliferation of inconsistencies (*Ibid.*, p77).
(5) A reactionary response to problematic facts. That is, the replacement of problemshifts with those that do not predict further problemshifts (*Ibid.*, p88).

Nevertheless, a research programme cannot be declared degenerative and be eliminated by any single investigation, for such a study can be incorporated into the research programme and be used as a base for further problemshifts. As such, there is no "instant learning[…] rationality works much slower than most people tend to think" (*Ibid.*, p87) and any investigation that finds a research programme to possess the typical signs of degeneration is merely a piece of "constructive criticism" to that programme (*Ibid.*, p92). Consequently no investigation can eliminate a programme, they can only be replaced by progressive rival programmes (*Ibid.*, p69).

Applying this epistemology and its typology to the work of Johnson-Laird, it is found that the research programme is his analogical RTM; the hard core of his theory are the base theories of RTM covered in our literature review (i.e., propositional attitudes, functionalism, computation, and Marr's tri-level hypothesis); and his supplementary hypotheses, including his theory of semantic content, constitute the protective belt of his research programme. Finally, Johnson-Laird's recent referentialist move constitutes the most current problemshift of his analogical RTM research programme. As such, it is interesting to note that our Literature Review found that the motivations for this shift were the arguments encapsulated by the CRA and SGP and not for any explanatory power offered by a correspondence theory (i.e., (2) and (5) above) and our findings discovered that the problemshift merely proliferates inconsistencies (i.e., (4) above). As such, and looking to Lakatos' typical signs of a degenerating research programme, this suggests that Johnson-Laird's RTM is in a degenerative position as he seems to be only in the business of updating auxiliary hypotheses to protect his analogical RTM rather than increasing explanatory power (cf. *Ibid.*, p112). However, this is in no way a confirmed conclusion, only an observation. As such, a potential avenue for future research may be an investigation into Johnson-Laird's theories to identify if this accusation is accurate, i.e., to identify if Johnson-Laird's research program is progressing and still making novel predictions or if it is merely reacting to problematic facts and should be labelled a degenerative science. If such a study finds that this move does represent a degenerative problemshift for Johnson-Laird, "the hard core [of his programme] might have to be abandoned" (*Ibid.*, p49). However, it should be remembered that:

1) A degenerative problemshift does not equate to a degenerative research programme.
2) Any one study will never result in the abandonment of Johnson-Laird's research programme, it can only serve as "constructive criticism" and inspiration for further problemshifts (*Ibid.*, p92).

Consequently, even if such a study as here proposed concludes that Johnson-Laird's current problemshift is degenerative, Johnson-Laird may still replace his semantics with some other semantic account in a progressive problemshift and continue on a valid and progressive research programme.

As such, under the sophisticated falsificationist epistemology of Lakatos (*Ibid.*), this research's findings may be used to identify Johnson-Laird's referentialist RTM as a degenerative problemshift in his research programme. However, what needs to be established are the limits of this observation. That is, it would be of interest for future research to investigate whether this finding merely represents a degenerative phase in an otherwise healthy and progressive research programme, or whether this points to a programme in decline.

Summary

To summarise, the third section of our Findings chapter raised and investigated two points of discussion:

1) The introduction of a correspondence theory into Johnson-Laird's philosophy may introduce an unacceptable level of residual difficulties. That is, as correspondence necessitates a commitment to such controversial hypotheses as eternalism, determinism and an epistemic scepticism of meaning, it may present itself to some as too controversial a position to accept into Johnson-Laird's philosophy, even if it is, or can be made to be, consistent.

2) That, when viewed under the sophisticated falsificationist epistemology of Lakatos (*Ibid.*), it seems that Johnson-Laird's move to referentialism may stand as a degenerative problemshift. That is, our findings may be viewed as a demonstration that Johnson-Liard's recent developments to his analogical RTM are merely reactionary changes that proliferate inconsistencies and that have no heuristic (i.e., predictive) power. This is a serious accusation, as this may be indicative of a degenerative research programme, which, if so, may lead to a need to abandon one or more of the core theories of Johnson-laird's RTM (i.e., propositional attitudes, functionalism, computation or Marr's tri-level hypothesis). As such, future research into this observation is recommended.

Conclusion

This chapter has tasked itself with reflecting on the findings of our last chapter and investigating any implications or areas and topics of interest they raise. Through this process, a number of observations supplementary to our findings were made:

1) Johnson-Laird's externalism is dependent on a system's extra-linguistic connection with its environment and this extra-linguistic connection is necessarily opposed to the mereological fallacy.

2) Johnson-Laird has not clarified whether the externalism he has introduced into his philosophy of mind is of a weak or a strong nature. Further, the difficulties implicit in either type of externalism have not currently been addressed by Johnson-Laird.

3) The removal of Peirce's iconism from Johnson-Laird's philosophy leaves a deficit that needs to be filled if his referentialist RTM is to validly continue.

4) The context of Johnson-Laird's broader philosophy demarks a criteria of success any replacement correspondence theory of meaning must meet if it is to be consistently introduced.

5) Wittgenstein's picture theory of meaning, detailed within his TL-P, fails to meet this criteria of success and does not stand as a viable replacement correspondence theory for Johnson-Laird.

6) Correspondence theories, generally, necessitate a number of daunting hypotheses to operate successfully. Furthermore, these implicit hypotheses are of such a nature that some may reject them instantly.

7) This investigations' findings may be evidence that Johnson-Laird's RTM project may be in a degenerative phase.

As well as these observations, several further proposals for future research were also made (additional to the need to verify and triangulate this study's findings and investigate other referentialist RTM approaches identified during the Evaluation of our findings):

1) An investigation into the validity of the criteria of success established by this thesis for putative replacements of Peircean iconism, as well as further studies measuring other correspondence theories of meaning against this criteria.

2) Further research to establish the accuracy of the claim that Johnson-Laird is currently in a degenerative phase. Or, research into whether the findings of this research concern a degenerative problemshift or a degenerative research programme.

Conclusion

We began this research by observing a gap in the current cognitive science literature; in observing that there is currently an absence of detailed scrutiny of the psychologist Philip Johnson-Laird's shift toward an externalist representational theory of mind. This conclusion was reached through a review of the cognitive science literature on the recent reaction to semantic arguments against the representational theory of mind (a suite of arguments encapsulated by John Searle's Chinese room argument). From this review, it was identified that, due to such arguments, many theorists have been motivated to shift from an internalist, syntactical account of meaning toward an externalist, referentialist semantics. From here, and through a review of correspondence theories of meaning in their own right, we further observed that this shift constitutes both a substantial and, to some, controversial move that necessitates considerable analysis before its adoption should be confidently asserted as valid; a level of analysis that is currently lacking in the literature. As such, we were led to the recommendation that research into the validity of this move be conducted. However, due to the nuanced differences found in the myriad referentialist positions within the representational theory of mind paradigm, such a broad piece of research was found to be beyond the scope of this thesis. Instead, an investigation into just one of these positions would be practicable. In this case, the position held by the psychologist Philip Johnson-Laird was singled out as the focus of such an investigation, as he is both an outspoken and long term advocate of the introduction of a correspondence theory of meaning into the account of semantic content offered by cognitive science. Consequently, we were led to the following research aim:

> **Aim:** To assess the validity of the hypothesis that the introduction of a Peircean correspondence theory of meaning into Johnson-Laird's mental model theory of mind saves the latter from the critiques of its commentators and to test the success of the inclusion of a Peircean referentialist semantics into such a representational theory of mind.

With this gap in the literature and research objective clarified, it was then established that the best method to employ in meeting this aim would be the philosophic method of decompositional analysis and synthesis. This is a method more commonly employed within the discipline of philosophy and entails the breaking down of a theory, argument etc., into its component parts by analysis, before a reconstruction of these parts through synthesis. This method is therefore a two-step process designed to make perspicuous the atomic hypotheses that make up a position as well as their interaction and

relations. In this instance, this method was applied in relation to the following research questions; questions designed specifically to interrogate Johnson-Laird's hypotheses in relation to our aim:

1) Does Philip Johnson-Laird's updated semantics circumvent the semantic objections of the representational theory of mind's commentators?

2) Is Charles Sanders Peirce's iconic correspondence theory of meaning (i.e., the particular correspondence theory of meaning implemented by Johnson-Laird) internally and externally consistent?

3) Is the introduction of a correspondence theory of meaning, generally, into Philip Johnson-Laird's broader philosophy of cognition logically consistent?

With this framework established, it was argued that, if the application of the method of philosophic analysis and synthesis produced affirmative answers to all three of the above questions, Philip Johnson-Laird's referentialist shift would be said to be valid and successful. However, if just one negative response were generated to any one of them, the move would be deemed to be invalid and unsuccessful.

Finally, it should be clarified that this method was deemed to be both appropriate and valid in this instance as:

(i)   It is an acceptable approach under the post-positivist paradigm (i.e., it will operate within Johnson-Laird's own framework and will produce findings acceptable under his own lights.

(ii)   That due to the object of study here (i.e., semantic content), the 'traditional' quantitative and qualitative methods employed within this post-positivist paradigm are not suitable, leading us to analysis and synthesis by inference to the best explanation.

(iii)   As the semantic arguments that inspired the move to externalism derive from philosophical analysis, it seems only appropriate to reapply the method when investigating how successful this move is at subverting them.

Findings

Now that these preliminaries have contextualised our research, we are in a position to summarise the findings our application of philosophical analysis and synthesis has produced. From this investigation, a number of results were derived:

1) The introduction of a correspondence theory of meaning, and such a theory's implicit language entry and exit rules, is a successful counter to the semantic arguments against the representational theory of mind, as summarised by the Chinese room argument.

2) Philip Johnson-Laird's introduction of C.S. Peirce's correspondence theory of meaning involves the reintroduction of the homunculus and symbolic fallacies through the inclusion of mental signifiers (for such things as negation and other abstract concepts) and the concept of an Interpretant found within C.S. Peirce's theory of icons.

3) Mental signifiers (for such things as negation and other abstract concepts) conflict with Philip Johnson-Laird's own naturalism constraint, i.e.:

> a *natural* mental model of discourse has a structure that corresponds directly to the structure of the state of affairs that the discourse describe. (Johnson-Laird 1983, p125)

4) A tension exists between C.S. Peirce and Philip Johnson-Laird over the role of mental images in cognition, with the former holding that they are the basis for cognition and reasoning, while the latter is explicit in his denial of this position.

5) The limitations of genre/pictorial form are not accounted for within Philip Johnson-Laird's iconic mental models.

6) The one-way nature of reference within a two-way theory of iconism is not explained within Philip Johnson-Laird's iconic mental models.

7) Brentano's problem is not accounted for within Philip Johnson-Laird's physicalist iconic mental models.

8) Philip Johnson-Laird's adherence to free will is antithetical to correspondence theories of meaning: The intuitive notion that future tensed propositions are meaningful cannot be simultaneously held with Johnson-Laird's position that there are no determinate, concrete future states of affairs.

Consequently, when these findings are applied to the research questions of this investigation, specified earlier, the following is uncovered:

Q1) Does Johnson-Laird's updated account of meaning circumvent the semantic objections of the representational theory of mind's commentators?

A1) Yes. The introduction of language entry and exit rules, through a correspondence theory of meaning, ensures that a system's symbols are grounded and that such a system is able to avoid the arguments from such cases as the Chinese room argument (see point 1) above).

Q2) Is Peirce's structural correspondence theory of meaning internally and externally consistent?

A2) No. As Peirce's iconism employs the concept of both an Interpretant and purely mental symbolism, the theory is internally logically inconsistent (see point 2) above). Furthermore, as Peirce and Johnson-Laird differ on the role of mental imagery in cognition and the necessity of the naturalism constraint, Peirce's theory is also found to be externally logically inconsistent (see points 3) and 4) above).

Q3) Is the introduction of a correspondence theory of meaning into Johnson-Laird's broader philosophy of cognition logically consistent?

A3) No. Johnson-Laird fails to account for the limitations of genre/pictorial form, the one-way nature of reference within a two-way theory of iconism or Brentano's problem before his introduction of a correspondence theory. Consequently, such an introduction is therefore logically inconsistent (see points 5), 6) and 7) above). Furthermore, there is a direct and explicit tension between Johnson-Laird's commitment to free will and the reliance of correspondence theories on determinism to account for future tensed propositions (see point 8) above).

As such, and in reference to our stipulation that 'if it is discovered that the theory concerned can answer in the affirmative to all three of our research questions, it will be considered a logically consistent position and will remain as a provisional conjecture; if a negative response is produced to any one of them, the inclusion of an externalist account of meaning into Johnson-Laird's theory of mind will be found to be inconsistent and the theory will be said to have been falsified', we are led to the finding that, as Philip Johnson-Laird's hypothesis cannot answer in the affirmative to all three of these questions, his referentialist move should be considered invalid and unsuccessful under both the Quinean and Popperian epistemologies.

Implications of findings

A number of implications can be derived from the finding that Johnson-Laird's philosophy of mind is, as it currently stands, logically inconsistent and should be considered falsified. These implications are pertinent to both Philip Johnson-Laird's philosophy specifically and to externalist semantics generally. Those inferences concerning Johnson-Laird specifically, are:

1) For Johnson-Laird to retain the possibility of progressing on with a productive research programme, he needs to abandon his iconic correspondence theory of meaning and rethink his account of semantics.

2) Johnson-Laird's wider philosophy of mind demarks a criteria of success, which must be met by any putative replacement correspondence theory of meaning, if this replacement is to produce a valid and logically consistent philosophy of mind.

The implications that can be drawn from the above finding that are pertinent to externalist semantics generally are:

1) Externalist accounts of meaning are dependent on a system's extra-linguistic connection with its environment and that this extra-linguistic connection is necessarily related to the mereological fallacy.

2) It is necessary for a theorist to clarify whether the externalism they hold is of a weak or a strong nature and that the complications implicit in either type of externalism must also be addressed by that theorist.

3) Correspondence theories, generally, necessitate a number of daunting hypotheses in order to operate successfully (for example, eternalism, determinism and an epistemic scepticism of meaning) and these implicit hypotheses alone are of such a nature that some theorists may reject an externalist semantics or a referentialist shift out of hand.

Evaluation

With these conclusions and inferences drawn, this investigation proceeded to demonstrate that a sufficient level of critical scrutiny had been applied to such findings, thereby evidencing that the level of confidence placed on them is justified. In this regard, it should be stressed that the method of philosophic analysis and synthesis employed in this study has produced enlightening data concerning a notoriously ephemeral object of study: Meaning. Furthermore, as this method, and its findings concerning logical consistency, are informative under both the Popperian and Quinean epistemologies – in that they are repeatable, falsifiable and naturalised with cognitive science – this research was capable of contributing to the literature at face value (i.e., without the need to engage in further, paradigmatic, argument). Finally, this investigation was deemed to be a success as its design, through its focus on establishing logical consistency, was able to meet our research aim and assess the validity, or lack thereof, of the claim that the introduction of a Peircean correspondence theory of meaning into Philip Johnson-Laird's mental model theory of mind saves the latter from the critiques of its commentators.

It is important to note, however, that there were also some limitations to this investigation. For example, although the confining of focus to the thought and argument of the psychologist Philip

Johnson-Laird was deemed a necessity, this limitation caused problems concerning generalizability. For, it was found that the nuance found in the literature between the related referentialist positions within the representational theory of mind means that the conclusions reached by this research are only applicable to the work of Johnson-Laird and no other, at least without much further work. That is, our findings cannot be generalised to the theories of others working within the referentialist representational theory of mind (e.g. Brooks 1990; Harnard 1990b; Waskan 2006; Fodor 2010; Rapaport 2011; Schweizer 2012). Finally, it was also determined that further research into the arguments of Johnson-Laird, repeating and triangulating this study's findings, is necessary before this investigation's conclusions can be assuredly asserted.

Recommendations

In carrying out this investigation, a number of areas that would benefit from further research were identified. It was therefore recommended that:

1) Further studies be conducted to verify and triangulate this study's findings. That is, further research is needed to confirm our findings in their own right.

2) Investigations into the success of other referentialist representational theories of mind should be conducted. That is, further research into the success of the work of those others operating within the cognitive science paradigm who have made the referentialist shift (e.g., William Rapaport, Paul Schweizer and Jonathan Waskan) would be beneficial.

3) Investigation into the validity of the criterion of success, for putative replacement to Peirce's iconism within Johnson-Laird's philosophy, established by this thesis be carried out.

4) Further research applying this criterion of success to other, replacement correspondence theories of meaning (i.e., those that may putatively fill the gap left in Johnson-Laird's philosophy by the removal of Peirce) be conducted.

5) Studies into the accuracy of the claim that Johnson-Laird's research programme is currently in a degenerative phase be undertaken. That is, research into whether the findings of this thesis signify a degenerative problemshift or a degenerative research programme for Johnson-Laird.

Final remarks

To conclude, this research has met its aim and identified that the introduction of Charles Sanders Peirce's theory of iconism into Philip Johnson-Laird's analogical mental models is not successful and ought to be reconsidered. However, what this research cannot make comment on is the success, or otherwise, of the myriad similar semantic shifts recently taken in the literature. Rather, and in closing,

the message that should be taken from this investigation is that correspondence theories of meaning are substantial and complex hypotheses and their inclusion into any wider philosophy is no trivial matter: This research has shown that if the sizable nature of these accounts of semantics are not fully respected and their adoption within a wider philosophy of mind is taken without due levels of analysis, then important philosophical inconsistencies may be unknowingly sustained.

Appendix

Individual Citations

Table 1a

Citation	Title	Author(s)	Publication	Journal/Book title (if appropriate)
1	Improving end-user proficiency: Effects of conceptual training and nature of interaction	Santhanam, R., Sein, M.K.	1994	Information Systems Research 5 (4) , pp. 378-399
2	Designing human-machine interfaces to match the user's mental models	Fuchs-Frohnhofen, P., Hartmann, E.A., Brandt, D., Weydandt, D.	1996	Control Engineering Practice 4 (1) , pp. 13-18
3	Planning, encoding, and overcoming conflict in partial occlusion drawing: A neo-Piagetian model and an experimental analysis	Morra, S., Angi, A., Tomat, L.	1996	Journal of Experimental Child Psychology 61 (3) , pp. 276-301
4	Causal model-based knowledge acquisition tools: Discussion of experiments	Charlet, J., Reynaud, C., Krivine, J.-P.	1996	International Journal of Human Computer Studies 44 (5) , pp. 629-652
5	A Theory of Decision Support System Design for User Calibration	Kasper, G.M.	1996	Information Systems Research 7 (2) , pp. 215-232
6	Illustrations et modèles mentaux dans la compréhension de textes	[Illustrations and mental models in text comprehension]	Gyselinck, V.	1996	Annee Psychologique 96 (3) , pp. 495-516
7	User instructions for the elderly: What the literature tells us	Van Hees, M.M.W.	1996	Journal of Technical Writing and Communication 26 (4) , pp. 521-536
8	Étude expérimentale de l'usage des organisateurs paralinguistiques de mise en retrait dans les documents électroniques [Using weakening paralinguistic organizers in electronic documents: An experimental approach]	Caro, S., Bisseret, A.	1997	Travail Humain 60 (4) , pp. 409-437
9	Structure of operators' mental models in coping with anomalies occurring in nuclear power plants	Takano, K., Sasou, K., Yoshimura, S.	1997	International Journal of Human Computer Studies 47 (6) , pp. 767-789
10	Teams embedded in organizations: Some implications	Ilgen, D.R.	1999	American Psychologist 54 (2) , pp. 129-139
11	Cognitive systems engineering: New wine in new bottles	Hollnagel, E., Woods, D.D.	1999	International Journal of Human Computer Studies 51 (2) , pp. 339-356
12	10 Cognitive modelling of vision and speech understanding	Hildebrandt, B., Moratz, R., Rickheit, G., Sagerer, G.	1999	Advances in Psychology 128 (C) , pp. 213-236
13	Scientific theories and naive theories as forms of mental representation: psychologism revived	Brewer, W.F.	1999	Science and Education 8 (5) , pp. 489-505
14	Children's development of a theory of mind	Cadinu, M.R., Kiesner, J.	2000	European Journal of Psychology of Education 15 (2) , pp. 93-111
15	L'intégration spatiale d'éléments textuels et illustratifs améliore-t-elle la performance ?	Jamet, E.	2000	Revue d'Intelligence Artificielle 14 (1-2) , pp. 167-187
16	Decision Constraints and Consideration-Set Formation in Consumer Durables	Punj, G., Brookes, R.	2001	Psychology and Marketing 18 (8) , pp. 843-863
17	Cognitive perspectives on unresolved loss: Insights from the study of PTSD	Fearon, R.M.P., Mansell, W.	2001	Bulletin of the Menninger Clinic 65 (3) , pp. 380-396
18	The role of prototypes in the mental representation of temporally related events	Colcombe, S.J., Wyer Jr., R.S.	2002	Cognitive Psychology 44 (1) , pp. 67-103
19	Visuospatial memory and phonological loop in learning from multimedia	Gyselinck, V., Cornoldi, C., Dubois, V., De Beni, R., Ehrlich, M.-F.	2002	Applied Cognitive Psychology 16 (6) , pp. 665-685
20	The influence of pre-decisional constraints on information search and consideration set formation in new automobile purchases	Punj, G., Brookes, R.	2002	International Journal of Research in Marketing 19 (4) , pp. 383-400
21	Age changes of semantic composition of associative reactions and effect on them of emotional sign of the verbal stimulus	Pakhomova, A.S.	2003	Journal of Evolutionary Biochemistry and Physiology 39 (1) , pp. 110-118
22	Les stratégies de traitement des phrases relatives chez le lecteur débutant	[Processing relative clauses]	Gola-Asmussen, C., Bastien-Toniazzo, M.	2003	Enfance 55 (2) , pp. 159-173
23	Why did John Herschel fail to understand polarization? The differences between object and event concepts	Chen, X.	2003	Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences 34 (3) , pp. 491-513
24	When scientific models represent	Bailer-Jones, D.M.	2003	International Studies in the Philosophy of Science 17 (1) , pp. 59-74
25	Object and Event Concepts: A Cognitive Mechanism of Incommensurability	Chen, X.	2003	Philosophy of Science 70 (5) , pp. 962-974
26	On the ontology of digital identification	Hogben, G., Wilikens, M., Vakalis, I.	2003	Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 2889 , pp. 579-593
27	The quality of human-automation cooperation in human-system interface for nuclear power plants	Skjerve, A.B.M., Skraaning Jr., G.	2004	International Journal of Human Computer Studies 61 (5) , pp. 649-677
28	Framing-effects approach: A theoretical and methodological critique	Scheufele, B.	2004	Communications 29 (4) , pp. 401-428
29	The role played by the right hemisphere in the organization of complex textual structures	Marini, A., Carlomagno, S., Caltagirone, C., Nocentini, U.	2005	Brain and Language 93 (1) , pp. 46-54
30	Understanding instructional leadership through the mental models of three elementary school principals	Ruff, W.G., Shoho, A.R.	2005	Educational Administration Quarterly 41 (3) , pp. 554-577
31	Exploiting concept mapping in a semantic web environment	Giouvanakis, T., Fragidis, G., Kehris, E., Samaras, H.	2005	Proceedings - 5th IEEE International Conference on Advanced Learning Technologies, ICALT 2005 2005 , art. no. 1508720, pp. 432-434
32	Sprachverarbeitung: Theorienkonkurrenz oder gegenstandsflexibilität?	[Language processing: Competing theories or flexibility of the linguistic object ?]	Groeben, N.	2005	Zeitschrift fur Psychologie 213 (3) , pp. 152-158
33	Cognitive evaluation of human-robot systems: A method for analyzing cognitive change in human-robot systems	Stubbs, K., Bernstein, D., Crowley, K., Nourbakhsh, I.	2006	Proceedings - IEEE International Workshop on Robot and Human Interactive Communication , art. no. 4107786, pp. 59-65
34	Content-based analysis of modes in design engineering	Saariluoma, P., Nevala, K., Karvinen, M.	2006	Design Computing and Cognition '06 , pp. 325-344
35	General introduction: Current developments in cognitive psychology, neuroscience, and the philosophy of mind	[No author name available]	2006	Advances in Psychology 138 (1) , pp. 5-22
36	From the choice of the patients for a transitive verb to its polysemy	Cordier, F., Pariollaud, F.	2007	Current Psychology Letters: Behaviour, Brain and Cognition 1 (21)
37	La solución de problemas cognitiuos. Una reflexión cognitiva sociocultural	[The solution of cognitive problems. A cognitive and socio-cultural reflection]	Montealegre, R.	2007	Avances en Psicología Latinoamericana 25 (2) , pp. 20-39
38	Reasoning about knowledge in human-automation systems (preliminary report)	Moses, Y., Shamo, M.K.	2007	Collection of Technical Papers - 2007 AIAA InfoTech at Aerospace Conference 1 , pp. 72-87
39	Nuevos avances en el estudio científico de la comprensión de textos	[New advances in the scientific study of text comprehension]	Marmolejo-Ramos, F.	2007	Universitas Psychologica 6 (2) , pp. 331-343
40	Frontal eye fields involved in shifting frame of reference within working memory for scenes	Wallentin, M., Roepstorff, A., Burgess, N.	2008	Neuropsychologia 46 (2) , pp. 399-408
41	Toward a (pragmatic) science of strategic intervention: Design propositions for scenario planning	Hodgkinson, G.P., Healey, M.P.	2008	Organization Studies 29 (3) , pp. 435-457
42	Computational memory architectures for autobiographic agents interacting in a complex virtual environment: A working model	Ho, W.C., Dautenhahn, K., Nehaniv, C.L.	2008	Connection Science 20 (1) , pp. 21-65
43	The role of working memory components in multimedia comprehension	Gyselinck, V., Jamet, E., Dubois, V.	2008	Applied Cognitive Psychology 22 (3) , pp. 353-374
44	Closed loop state space analysis and simulation for cognitive systems	Oberheid, H., Gamrad, D., Söffker, D.	2008	Proceedings - International Conference on Application of Concurrency to System Design, ACSD , art. no. 4574593, pp. 39-44
45	The language of schizophrenia: An analysis of micro and macrolinguistic abilities and their neuropsychological correlates	Marini, A., et al	2008	Schizophrenia Research 105 (1-3) , pp. 144-155
46	Straight-party voting: What do voters think?	Campbell, B.A., Byrne, M.D.	2009	IEEE Transactions on Information Forensics and Security 4 (4) , art. no. 5256335 , pp. 718-728
47	Affect, behavioural schemas and the proving process	Selden, A., McKee, K., Selden, J.	2010	International Journal of Mathematical Education in Science and Technology 41 (2) , pp. 199-215
48	A social-cognitive framework of multidisciplinary team innovation	Paletz, S.B.F., Schunn, C.D.	2010	Topics in Cognitive Science 2 (1) , pp. 73-95

49	Contrast and assimilation effects of processing fluency	Shen, H., Jiang, Y., Adaval, R.	2010	Journal of Consumer Research 36 (5), pp. 876-889
50	Synthetic environments as visualization method for product design	Meijer, F.,et al	2010	Proceedings of SPIE - The International Society for Optical Engineering 7527 , art. no. 752712
51	The breakdown of coordinated decision making in distributed systems	Bearman, C., Paletz, S.B.F., Orasanu, J., Thomas, M.J.W.	2010	Human Factors 52 (2), pp. 173-188
52	Between understanding and misunderstanding	Verdonik, D.	2010	Journal of Pragmatics 42 (5), pp. 1364-1379
53	Scenistic methods in training: Definitions and theory grounding	Lyons, P.	2010	Journal of European Industrial Training 34 (5), pp. 416-431
54	Merging internal and external processes: Examining the mental model convergence process through team communication	Kennedy, D.M., McComb, S.A.	2010	Theoretical Issues in Ergonomics Science 11 (4), pp. 340-358
55	The nature of the refutation text effect: An investigation of attention allocation	Broughton, S.H., Sinatra, G.M., Reynolds, R.E.	2010	Journal of Educational Research 103 (6), pp. 407-423
56	Inclusion and interaction: Designing interaction for inclusive populations	Langdon, P., Thimbleby, H.	2010	Interacting with Computers 22 (6), pp. 439-448
57	Developing a model of cognitive interaction for analytical inclusive design evaluation	Langdon, P., Persad, U., John Clarkson, P.	2010	Interacting with Computers 22 (6), pp. 510-529
58	Toward a (pragmatic) science of strategic intervention: Design propositions for scenario planning	Hodgkinson, G.P., Healey, M.P.	2008	Organization Studies 29 (3), pp. 435-457
59	Naturalistic cognition: A research paradigm for human-centered design	Storkerson, P.	2010	Journal of Research Practice 6 (2)
60	Team creative processes: The importance of complementary and shared mental models	Kennedy, D.M.	2011	Proceedings of the Annual Hawaii International Conference on System Sciences , art. no. 5718448
61	The Influence of Connectives on Young Readers' Processing and Comprehension of Text	Cain, K., Nash, H.M.	2011	Journal of Educational Psychology 103 (2), pp. 429-441
62	Emulation as an integrating principle for cognition	Colder, B.	2011	Frontiers in Human Neuroscience (MAY), pp. 2
63	The effects of worked examples on CAD performance: An application of the four-component instructional design model to CAD instruction	Barnes, S., Wiebe, E.N., Branoff, T.J.	2011	ASEE Annual Conference and Exposition, Conference Proceedings
64	Subject-oriented elicitation of distributed business process knowledge	Oppl, S.	2011	Communications in Computer and Information Science 213 CCIS , pp. 16-33
65	Evidence maps: Communicating risk assessments in societal controversies: The case of engineered nanoparticles	Wiedemann, P., Schütz, H., Spangenberg, A., Krug, H.F.	2011	Risk Analysis 31 (11), pp. 1770-1783
66	Evolution of a predictive internal model in an embodied and situated agent	Gigliotta, O., Pezzulo, G., Nolfi, S.	2011	Theory in Biosciences 130 (4), pp. 259-276
67	A multi-level approach to the analysis of narrative language in aphasia	Marini, A., Andreetta, S., del Tin, S., Carlomagno, S.	2011	Aphasiology 25 (11), pp. 1372-1392
68	Distributed Cognition as a Lens to Understand the Effects of Scaffolds: The Role of Transfer of Responsibility	Belland, B.R.	2011	Educational Psychology Review 23 (4), pp. 577-600
69				
70	Stakeholder-driven collaborative modeling of subject-oriented business processes	Wachholder, D., Oppl, S.	2012	Lecture Notes in Business Information Processing 104 LNBIP , pp. 145-162
71	Leveraging the affordances of Youtube: The role of pedagogical knowledge and mental models of technology functions for lesson planning with technology	Krauskopf, K., Zahn, C, Hesse, F.W.	2012	Computers and Education 58 (4), pp. 1194-1206
72	Guiding decisions on authorization policies: A participatory approach to decision support	Bartsch, S., Sasse, M.A.	2012	Proceedings of the ACM Symposium on Applied Computing, pp. 1502-1507
73	Thinking Outside the Box? Applying Design Theory to Public Policy	Considine, M.	2012	Politics and Policy 40 (4), pp. 704-724
74	Exploring visual perception and children's interpretations of picture books	Yu, X.	2012	Library and Information Science Research 34 (4), pp. 292-299
75	Persuasive environments: Argumentation models in serious games	Pisan, Y., Tan, C.T.	2012	ICCSE 2012 - Proceedings of 2012 7th International Conference on Computer Science and Education , art. no. 6295308 , pp. 1319-1322
76	Lexical entrainment and success in student engineering groups	Friedberg, H., Litman, D., Paletz, S.B.F.	2012	2012 IEEE Workshop on Spoken Language Technology, SLT 2012 - Proceedings , art. no. 6424258 , pp. 404-409
77	An ACT-R approach to reasoning about spatial relations with preferred and alternative mental models	Brüssow, S., Frorath, M., Ragni, M., Fangmeier, T.	2012	Proceedings of the 11th International Conference on Cognitive Modeling, ICCM 2012 , pp. 19-24
78	Multiple logics within argument: How defeasible and classical reasoning work together	Stenning, K.	2012	Frontiers in Artificial Intelligence and Applications 245 (1) , pp. 14-20
79	Diversity of reason: Equivalence relations over description logic explanations	Bail, S., Parsia, B., Sattler, U.	2012	CEUR Workshop Proceedings 846 , pp. 48-58
80	Extracting narrative timelines as temporal dependency structures	Kolomiyets, O., Bethard, S., Moens, M.-F.	2012	50th Annual Meeting of the Association for Computational Linguistics, ACL 2012 - Proceedings of the Conference 1 , pp. 88-97
81	Concept design as knowledge construction - The case of user-centered design	Liikkanen, L.A., Laakso, M.	2012	Proceedings of International Design Conference, DESIGN DS 70 , pp. 1341-1350
82	Language, Linguistics and Cognition (Chapter)	Baggio, G., van Lambalgen, M., Hagoort, P.	2012	Philosophy of Linguistics
83	Seeking a unified framework for cerebellar function and dysfunction: From circuit operations to cognition	D'Angelo, E., Casali, S.	2012	Frontiers in Neural Circuits (DEC)
84	Seeking a unified framework for cerebellar function and dysfunction: From circuit operations to cognition	D'Angelo, E., Casali, S.	2013	Frontiers in Neural Circuits (JAN)
85	Aging and predicting inferences: A diffusion model analysis	McKoon, G., Ratcliff, R.	2013	Journal of Memory and Language 68 (3), pp. 240-254
86	Reuse and recycle: The development of adaptive expertise, routine expertise, and novelty in a large research team	Paletz, S.B.F., Kim, K.H., Schunn, C.D., Tollinger, I., Vera, A.	2013	Applied Cognitive Psychology 27 (4), pp. 415-428
87	How a picture facilitates the process of learning from text: Evidence for scaffolding	Eitel, A., Scheiter, K., Schüler, A., Nyström, M., Holmqvist, K.	2013	Learning and Instruction 28 , pp. 48-63
88	Affordances of Augmented Reality in Science Learning: Suggestions for Future Research	Cheng, K.-H., Tsai, C.-C.	2013	Journal of Science Education and Technology 22 (4), pp. 449-462
89	Climate and narrative: Environmental knowledge in everyday life	Lejano, R.P., Tavares-Reager, J., Berkes, F.	2013	Environmental Science and Policy 31, pp. 61-70
90	Learning and open educational resources: A psychological perspective	Terras, M.M., Ramsay, J., Boyle, E.	2013	E-Learning and Digital Media 10 (2), pp. 161-173

Web of Knowledge -

164 citations

Table 1b

Citation	Title	Author(s)	Publication	Journal/Book title (if appropriate)
1	How a picture facilitates the process of learning from text: Evidence for scaffolding	Eitel, Alexander; Scheiter, Katharina; Schueler, Anne; et al.	2013	LEARNING AND INSTRUCTION Volume: 28 Pages: 48-63
2	Affordances of Augmented Reality in Science Learning: Suggestions for Future Research	Cheng, Kun-Hung; Tsai, Chin-Chung	2013	JOURNAL OF SCIENCE EDUCATION AND TECHNOLOGY Volume: 22 Issue: 4 Pages: 449-462
3	Climate and narrative: Environmental knowledge in everyday life	Lejano, Raul P.; Tavares-Reager, Joana; Berkes, Fikret	: AUG 2013	ENVIRONMENTAL SCIENCE & POLICY Volume: 31 Pages: 61-70
4	Reuse and Recycle: The Development of Adaptive Expertise, Routine Expertise, and Novelty in a Large Research Team	Paletz, Susannah B. F.; Kim, Kevin H; Schunn, Christian D.; et al.	: JUL 2013	APPLIED COGNITIVE PSYCHOLOGY Volume: 27 Issue: 4 Pages: 415-428
5	Aging and predicting inferences: A diffusion model analysis	McKoon, Gail; Ratcliff, Roger	: APR 2013	JOURNAL OF MEMORY AND LANGUAGE Volume: 68 Issue: 3 Pages: 240-254
6	Seeking a unified framework for cerebellar function and dysfunction: from circuit operations to cognition	D'Angelo, Egidio; Casali, Stefano	: JAN 10 2013	FRONTIERS IN NEURAL CIRCUITS Volume: 6 Article Number: 116
7	Visuality for Architects: Architectural Creativity and Modern Theories of Perception and Imagination	Mitrovic, B	: 2013	VISUALITY FOR ARCHITECTS: ARCHITECTURAL CREATIVITY AND MODERN THEORIES OF PERCEPTION AND IMAGINATION Pages: 1-120
8	Shared Mental Models for Collaborative Mechanical Design in Mechanical Engineering	Du, Junmin	: 2013	ADVANCED RESEARCH ON INTELLIGENT SYSTEMS AND MECHANICAL ENGINEERING Book Series:Advanced Materials Research Volume: 644 Pages: 378-381
9	Exploring visual perception and children's interpretations of picture books	Yu, Xinyu	: OCT 2012	LIBRARY & INFORMATION SCIENCE RESEARCH Volume: 34 Issue: 4 Pages: 292-299
10	Leveraging the affordances of Youtube: The role of pedagogical knowledge and mental models of technology functions for lesson planning with technology	Krauskopf, Karsten; Zahn, Carmen; Hesse, Friedrich W.	: MAY 2012	COMPUTERS & EDUCATION Volume: 58 Issue: 4 Pages: 1194-1206

#	Title	Authors	Date	Source
11	Effectiveness of holistic mental model confrontation in driving conceptual change	Gadgil, Soniya; Nokes-Malach, Timothy J.; Chi, Michelene T. H.	: FEB 2012	LEARNING AND INSTRUCTION Volume: 22 Issue: 1 Pages: 47-61
12	LANGUAGE, LINGUISTICS AND COGNITION	Baggio, Giosue; van Lambalgen, Michiel; Hagoort, Peter	: 2012	PHILOSOPHY OF LINGUISTICS Book Series: Handbook of the Philosophy of Science Pages: 325-355
13	LEXICAL ENTRAINMENT AND SUCCESS IN STUDENT ENGINEERING GROUPS	Friedberg, Heather; Litman, Diane; Paletz, Susannah B. F.	: 2012	2012 IEEE WORKSHOP ON SPOKEN LANGUAGE TECHNOLOGY (SLT 2012) Pages: 404-409
14	Stakeholder-Driven Collaborative Modeling of Subject-Oriented Business Processes	Wachholder, Dominik; Oppl, Stefan	: 2012	S-BPM ONE - SCIENTIFIC RESEARCH Book Series: Lecture Notes in Business Information Processing Volume: 104 Pages: 145-162
15	Evolution of a predictive internal model in an embodied and situated agent	Gigliotta, Onofrio; Pezzulo, Giovanni; Nolfi, Sefano	: DEC 2011	THEORY IN BIOSCIENCES Volume: 130 Issue: 4 Pages: 259-276
16	Distributed Cognition as a Lens to Understand the Effects of Scaffolds: The Role of Transfer of Responsibility	Belland, Brian R.	: DEC 2011	EDUCATIONAL PSYCHOLOGY REVIEW Volume: 23 Issue: 4 Pages: 577-600
17	Evidence Maps: Communicating Risk Assessments in Societal Controversies: The Case of Engineered Nanoparticles	Wiedemann, Peter; Schuetz, Holger; Spangenberg, Albena; et al.	: NOV 2011	RISK ANALYSIS Volume: 31 Issue: 11 Pages: 1770-1783
18	Emulation as an integrating principle for cognition	Colder, Brian	: MAY 27 2011	FRONTIERS IN HUMAN NEUROSCIENCE Volume: 5 Article Number: 54
19	The Influence of Connectives on Young Readers' Processing and Comprehension of Text	Cain, Kate; Nash, Hannah M.	: MAY 2011	JOURNAL OF EDUCATIONAL PSYCHOLOGY Volume: 103 Issue: 2 Pages: 429-441
20	Subject-Oriented Elicitation of Distributed Business Process Knowledge	Oppl, Stefan	: 2011	S-BPM ONE - LEARNING BY DOING - DOING BY LEARNING Book Series: Communications in Computer and Information Science Volume: 213 Pages: 16-33
21	An Observation-Based Method for Measuring the Sharedness of Mental Models in Teams	Badke-Schaub, Petra; Neumann, Andre; Lauche, Kristina	: 2011	COORDINATION IN HUMAN AND PRIMATE GROUPS Pages: 177-197
22	A multi-level approach to the analysis of narrative language in aphasia	Marini, Andrea; Andreetta, Sara; del Tin, Silvana; et al.	: 2011	APHASIOLOGY Volume: 25 Issue: 11 Special Issue: SI Pages: 1372-1392
23	How Audiences Decide: A Cognitive Approach to Business Communication	Young, RO	: 2011	HOW AUDIENCES DECIDE: A COGNITIVE APPROACH TO BUSINESS COMMUNICATION Pages: 1-398
24	Inclusion and interaction: Designing interaction for inclusive populations	Langdon, Patrick; Thimbleby, Harold	: NOV 2010	INTERACTING WITH COMPUTERS Volume: 22 Issue: 6 Special Issue: SI Pages: 439-448
25	Developing a model of cognitive interaction for analytical inclusive design evaluation	Langdon, Patrick; Persad, Umesh; Clarkson, P. John	: NOV 2010	INTERACTING WITH COMPUTERS Volume: 22 Issue: 6 Special Issue: SI Pages: 510-529
26	Between understanding and misunderstanding	Verdonik, Darinka	: MAY 2010	JOURNAL OF PRAGMATICS Volume: 42 Issue: 5 Pages: 1364-1379
27	The Breakdown of Coordinated Decision Making in Distributed Systems	Bearman, Christopher; Paletz, Susannah B. F.; Orasanu, Judith; et al.	: APR 2010	HUMAN FACTORS Volume: 52 Issue: 2 Special Issue: SI Pages: 173-188
28	Contrast and Assimilation Effects of Processing Fluency	Shen, Hao; Jiang, Yuwei; Adaval, Rashmi	: FEB 2010	JOURNAL OF CONSUMER RESEARCH Volume: 36 Issue: 5 Pages: 876-889
29	Shifting focus from the determinants to the origin: the foundations of a dynamic view of managerial discretion	Kleindienst, Ingo; Hutzschenreuter, Thomas	: 2010	HANDBOOK OF RESEARCH ON STRATEGY PROCESS Pages: 22-42
30	Synthetic Environments as visualization method for product design	Meijer, Frank; van den Broek, Egon L.; Schouten, Theo E.; et al.	: 2010	HUMAN VISION AND ELECTRONIC IMAGING XV Book Series: Proceedings of SPIE Volume: 7527 Article Number: 752712
31	The Nature of the Refutation Text Effect: An Investigation of Attention Allocation	Broughton, Suzanne H.; Sinatra, Gale M.; Reynolds, Ralph E.	: 2010	JOURNAL OF EDUCATIONAL RESEARCH Volume: 103 Issue: 6 Pages: 407-423
32	A Social-Cognitive Framework of Multidisciplinary Team Innovation	Paletz, Susannah B. F.; Schunn, Christian D.	: JAN 2010	TOPICS IN COGNITIVE SCIENCE Volume: 2 Issue: 1 Pages: 73-95
33	Language, Space, and Social Relationships: A Foundational Cultural Model in Polynesia	Bennardo, G	: 2009	LANGUAGE, SPACE, AND SOCIAL RELATIONSHIPS: A FOUNDATIONAL CULTURAL MODEL IN POLYNESIA Book Series: Language Culture and Cognition Issue: 9 Pages: 1-371
34	ITERATION IN SYSTEMS ANALYSIS AND DESIGN Cognitive Processes and Representational Artifacts	Berente, Nicholas; Lyytinen, Kalle	: 2009	SYSTEMS ANALYSIS AND DESIGN: TECHNIQUES, METHODOLOGIES, APPROACHES, AND ARCHITECTURES Book Series: Advances in Management Information Systems Volume: 15 Pages: 37-57
35	The language of schizophrenia: An analysis of micro and macrolinguistic abilities and their neuropsychological correlates	Marini, Andrea; Spoletini, Ilaria; Rubino, Ivo Alex; et al.	: OCT 2008	SCHIZOPHRENIA RESEARCH Volume: 105 Issue: 1-3 Pages: 144-155
36	nanoHUB.org: Advancing education and research in nanotechnology	Klimeck, Gerhard; McLennan, Michael; Brophy, Sean B.; et al.	: SEP-OCT 2008	COMPUTING IN SCIENCE & ENGINEERING Volume: 10 Issue: 5 Pages: 17-23
37	Advancing Engineering Education in P-12 Classrooms	Brophy, Sean; Klein, Stacy; Portsmore, Merredith; et al.	: JUL 2008	JOURNAL OF ENGINEERING EDUCATION Volume: 97 Issue: 3 Pages: 369-387
38	The role of working memory components in multimedia comprehension	Gyselinck, Valerie; Jamet, Eric; Dubois, Veronique	: APR 2008	APPLIED COGNITIVE PSYCHOLOGY Volume: 22 Issue: 3 Pages: 353-374
39	Toward a (pragmatic) science of strategic intervention: Design propositions for scenario planning	Hodgkinson, Gerard P.; Healey, Mark P.	: MAR 2008	ORGANIZATION STUDIES Volume: 29 Issue: 3 Pages: 435-457
40	Pedagogical Tools for the Training of Subtitlers	Taylor, Christopher	: 2008	AUDIOVISUAL TRANSLATION: LANGUAGE TRANSFER ON SCREEN Pages: 214-228
41	Computational memory architectures for autobiographic agents interacting in a complex virtual environment: a working model	Ho, Wan Ching; Dautenhahn, Kerstin; Nehaniv, Chrystopher L.	: 2008	CONNECTION SCIENCE Volume: 20 Issue: 1 Pages: 21-65
42	Frontal eye fields involved in shifting frame of reference within working memory for scenes	Wallentin, Mikkel; Roepstorff, Andreas; Burgess, Neil	: 2008	NEUROPSYCHOLOGIA Volume: 46 Issue: 2 Pages: 399-408
43	Constructing Mental Models during Learning from Science Text Eye Tracking Methodology Meets Conceptual Change	Mikkila-Erdmann, M.; Penttinen, M.; Anto, E.; et al.	: 2008	UNDERSTANDING MODELS FOR LEARNING AND INSTRUCTION: ESSAYS IN HONOR OF NORBERT M. SEEL Pages: 63-79
44	The 'Ins' and 'Outs' of Learning: Internal Representations and External Visualizations	Rapp, David N.; Kurby, Christopher A.	: 2008	VISUALIZATION: THEORY AND PRACTICE IN SCIENCE EDUCATION Book Series: Models and Modeling in Science Education Volume: 3 Pages: 29-52
45	Closed Loop State Space Analysis and Simulation for Cognitive Systems	Oberheid, Hendrik; Gamrad, Dennis; Soeffker, Dirk	: 2008	2008 8TH INTERNATIONAL CONFERENCE ON APPLICATION OF CONCURRENCY TO SYSTEM DESIGN, PROCEEDINGS Book Series: International Conference on Application of Concurrency to System Design. Proceedings Pages: 39-44
46	Thought experimenting as mental modeling: Empiricism without logic	Nersessian, Nancy J.	: 2007	CROATIAN JOURNAL OF PHILOSOPHY Volume: 7 Issue: 20 Pages: 125-161
47	General Introduction: Current Developments in Cognitive Psychology, Neuroscience, and the Philosophy of Mind	Held, C; Knauff, M; Vosgerau, G	: 2006	MENTAL MODELS AND THE MIND: CURRENT DEVELOPMENTS IN COGNITIVE PSYCHOLOGY, NEUROSCIENCE, AND PHILOSOPHY OF MIND Book Series: Advances in Psychology Volume: 138 Pages: 5-22
48	Understanding instructional leadership through the mental models of three elementary school principals	Ruff, WG; Shoho, AR	: AUG 2005	EDUCATIONAL ADMINISTRATION QUARTERLY Volume: 41 Issue: 3 Pages: 554-577
49	The role played by the right hemisphere in the organization of complex textual structures	Marini, A; Carlomagno, S; Caltagirone, C; et al.	: APR 2005	BRAIN AND LANGUAGE Volume: 93 Issue: 1 Pages: 46-54
50	INTRODUCTION: MAKING SOCIAL MEANINGS IN CONTEXTS Writing, Literate Activities and Learning Processes in Sociohistorical Communities	Kostouli, Triantafillia	: 2005	WRITING IN CONTEXT(S): TEXTUAL PRACTICES AND LEARNING PROCESSES IN SOCIOCULTURAL SETTINGS Book Series: Studies in Writing Volume: 15 Pages: 1+
51	Language processing: Competing theories or flexibility of the linguistic object?	Groeben, N	: 2005	ZEITSCHRIFT FUR PSYCHOLOGIE Volume: 213 Issue: 3 Pages: 152-158
52	Exploiting concept mapping in a semantic web environment	Giouvanakis, T; Fragidis, G; Kehris, E; et al.	: 2005	5th IEEE International Conference on Advanced Learning Technologies, Proceedings Pages: 432-434
53	Framing-effects approach: A theoretical and methodological critique	Scheufele, Bertram	: DEC 2004	COMMUNICATIONS-EUROPEAN JOURNAL OF COMMUNICATION RESEARCH Volume: 29 Issue: 4 Pages: 401-428
54	The quality of human-automation cooperation in human-system interface for nuclear power plants	Skjerve, ABM; Skraaning, G	: NOV 2004	INTERNATIONAL JOURNAL OF HUMAN-COMPUTER STUDIES Volume: 61 Issue: 5 Pages: 649-677
55	Object and event concepts: A cognitive mechanism of incommensurability	Chen, X	: DEC 2003	PHILOSOPHY OF SCIENCE Volume: 70 Issue: 5 Pages: 962-974
56	Why did John Herschel fail to understand polarization? The differences between object and event concepts	Chen, X	: SEP 2003	STUDIES IN HISTORY AND PHILOSOPHY OF SCIENCE Volume: 34A Issue: 3 Pages: 491-513
57	The influence of pre-decisional constraints on information search and consideration set formation in new automobile purchases	Punj, G; Brookes, R	: DEC 2002	INTERNATIONAL JOURNAL OF RESEARCH IN MARKETING Volume: 19 Issue: 4 Pages: 383-400 Article Number: PII S0167-8116(02)00100-3
58	Visuospatial memory and phonological loop in learning from multimedia	Gyselinck, V; Cornoldi, C; Dubois, V; et al.	: SEP 2002	APPLIED COGNITIVE PSYCHOLOGY Volume: 16 Issue: 6 Pages: 665-685
59	The role of prototypes in the mental representation of temporally related events	Colcombe, SJ; Wyer, RS	: FEB 2002	COGNITIVE PSYCHOLOGY Volume: 44 Issue: 1 Pages: 67-103

60	Narrative-based representations of social knowledge: Their construction and use in comprehension, memory, and judgment	Wyer, RS; Adaval, R; Colcombe, SJ	: 2002	ADVANCES IN EXPERIMENTAL SOCIAL PSYCHOLOGY, VOL 34 Book Series: ADVANCES IN EXPERIMENTAL SOCIAL PSYCHOLOGY Volume: 34 Pages: 131-197
61	Decision constraints and consideration-set formation in consumer durables	Punj, G; Brookes, R	: AUG 2001	PSYCHOLOGY & MARKETING Volume: 18 Issue: 8 Pages: 843-863
62	Cognitive perspectives on unresolved loss: Insights from the study of PTSD	Fearon, RMP; Mansell, W	: SUM 2001	BULLETIN OF THE MENNINGER CLINIC Volume: 65 Issue: 3 Pages: 380-396
63	The construction of mental representations during reading	Vellutino, FR	: SUM 2000	AMERICAN JOURNAL OF PSYCHOLOGY Volume: 113 Issue: 2 Pages: 306-318
64	Children's development of a theory of mind	Cadinu, MR; Kiesner, J	: JUN 2000	EUROPEAN JOURNAL OF PSYCHOLOGY OF EDUCATION Volume: 15 Issue: 2 Pages: 93-111
65	Given-new versus new-given? An analysis of reading times for spatial descriptions	Baguley, T; Payne, SJ	: 2000	SPATIAL COGNITION: FOUNDATIONS AND APPLICATIONS Book Series: ADVANCES IN CONSCIOUSNESS RESEARCH Volume: 26 Pages: 317-328
66	Cognitive systems engineering: New wine in new bottles	Hollnagel, E; Woods, DD	: AUG 1999	INTERNATIONAL JOURNAL OF HUMAN-COMPUTER STUDIES Volume: 51 Issue: 2 Pages: 339-356
67	Teams embedded in organizations - Some implications	Ilgen, DR	: FEB 1999	AMERICAN PSYCHOLOGIST Volume: 54 Issue: 2 Pages: 129-139
68	The human operator in complex networked systems	Tschiersch, I; Brandt, D	: 1998	(SAFEPROCESS'97): FAULT DETECTION, SUPERVISION AND SAFETY FOR TECHNICAL PROCESSES 1997, VOLS 1-3 Pages: 1261-1265
69	Structure of operators' mental models in coping with anomalies occurring in nuclear power plants	Takano, K; Sasou, K; Yoshimura, S	: DEC 1997	INTERNATIONAL JOURNAL OF HUMAN-COMPUTER STUDIES Volume: 47 Issue: 6 Pages: 767-789
70	An empirical study of computer system learning: Comparison of co-discovery and self-discovery methods	Lim, KH; Ward, LM; Benbasat, I	: SEP 1997	INFORMATION SYSTEMS RESEARCH Volume: 8 Issue: 3 Pages: 254-272
71	Illustrations and mental models in text comprehension	Gyselinck, V	: SEP 1996	ANNEE PSYCHOLOGIQUE Volume: 96 Issue: 3 Pages: 495-516
72	A theory of decision support system design for user calibration	Kasper, GM	: JUN 1996	INFORMATION SYSTEMS RESEARCH Volume: 7 Issue: 2 Pages: 215-232
73	Causal model-based knowledge acquisition tools: Discussion of experiments	Charlet, J; Reynaud, C; Krivine, JP	: MAY 1996	INTERNATIONAL JOURNAL OF HUMAN-COMPUTER STUDIES Volume: 44 Issue: 5 Pages: 629-652
74	Planning, encoding, and overcoming conflict in partial occlusion drawing: A neo-Piagetian model and an experimental analysis	Morra, S; Angi, A; Tomat, L	: APR 1996	JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY Volume: 61 Issue: 3 Pages: 276-301
75	Designing human-machine interfaces to match the user's mental models	FuchsFrohnhofen, P; Hartmann, EA; Brandt, D; et al.	: JAN 1996	CONTROL ENGINEERING PRACTICE Volume: 4 Issue: 1 Pages: 13-18
76	Gradient and visual speaker normalization in the perception of fricatives	Hildebrandt, B; Moratz, R; Rickheit, G; et al.	: 1996	NATURAL LANGUAGE PROCESSING AND SPEECH TECHNOLOGY: RESULTS OF THE 3RD KONVENS CONFERENCE Pages: 2-13
77	Points of view and point of view sequences	Herrmann, T	: 1996	SPRACHE & KOGNITION Volume: 15 Issue: 4 Pages: 159-177
78	Representations by machines + Cognitive psychology, semiotics, and information science	Glaser, WR	: 1996	ZEITSCHRIFT FUR SEMIOTIK Volume: 18 Issue: 2-3 Pages: 169-189
79	HOW DO CHILDREN REPRESENT INTELLIGENT TECHNOLOGY	VANDUUREN, M; SCAIFE, M	: SEP 1995	EUROPEAN JOURNAL OF PSYCHOLOGY OF EDUCATION Volume: 10 Issue: 3 Pages: 289-301
80	2 MODES OF MENTAL REPRESENTATION AND PROBLEM SOLUTION IN SYLLOGISTIC REASONING	FORD, M	: JAN 1995	COGNITION Volume: 54 Issue: 1 Pages: 1-71
81	INTERNAL REPRESENTATION AND POINT-OF-VIEW - THE EFFECT OF ROUTE DIRECTION	HERRMANN, T; BUHL, HM; SCHWEIZER, K	: 1995	ZEITSCHRIFT FUR PSYCHOLOGIE Volume: 203 Issue: 1 Pages: 1-23
82	IMPROVING END-USER PROFICIENCY - EFFECTS OF CONCEPTUAL TRAINING AND NATURE OF INTERACTION	SANTHANAM, R; SEIN, MK	: DEC 1994	INFORMATION SYSTEMS RESEARCH Volume: 5 Issue: 4 Pages: 378-399
83	SEMANTIC ANALYSIS AND IN-DEPTH UNDERSTANDING OF TECHNICAL TEXTS	CAVAZZA, M; ZWEIGENBAUM, P	: JUL-SEP 1994	APPLIED ARTIFICIAL INTELLIGENCE Volume: 8 Issue: 3 Pages: 425-453
84	MENTAL MODELS OF THE DAY-NIGHT CYCLE	VOSNIADOU, S; BREWER, WF	: JAN-MAR 1994	COGNITIVE SCIENCE Volume: 18 Issue: 1 Pages: 123-183
85	THE USE OF SUPERFICIAL AND MEANING-BASED REPRESENTATIONS IN INTERPRETING PRONOUNS - EVIDENCE FROM SPANISH	CARREIRAS, M; GARNHAM, A; OAKHILL, J	: MAR 1993	EUROPEAN JOURNAL OF COGNITIVE PSYCHOLOGY Volume: 5 Issue: 1 Pages: 93-116
86	WHAT ARE CONCEPTS - ISSUES OF REPRESENTATION AND ONTOLOGY	BREWER, WF	: 1993	CATEGORIZATION BY HUMANS AND MACHINES Book Series: PSYCHOLOGY OF LEARNING AND MOTIVATION-ADVANCES IN RESEARCH AND THEORY Volume: 29 Pages: 495-533
87	PROCESS HIGHLIGHTERS IN A COMPUTER-SIMULATION - FACILITATION OF THEORY-ORIENTED PROBLEM-SOLVING	CHANDLER, TN; CHAILLE, C	: 1993	JOURNAL OF EDUCATIONAL COMPUTING RESEARCH Volume: 9 Issue: 2 Pages: 237-263
88	RECENT VIEWS OF CONCEPTUAL STRUCTURE	KOMATSU, LK	: NOV 1992	PSYCHOLOGICAL BULLETIN Volume: 112 Issue: 3 Pages: 500-526
89	LEVELS OF REPRESENTATION AND DOMAIN-SPECIFIC KNOWLEDGE IN COMPREHENSION OF SCIENTIFIC TEXTS	TARDIEU, H; EHRLICH, MF; GYSELINCK, V	: AUG-NOV 1992	LANGUAGE AND COGNITIVE PROCESSES Volume: 7 Issue: 3-4 Pages: 335-351
90	DISCOURSE ORGANIZATION IN THE COMPREHENSION OF TEMPORAL-ORDER IN NARRATIVE TEXTS	OHTSUKA, K; BREWER, WF	: JUL-SEP 1992	DISCOURSE PROCESSES Volume: 15 Issue: 3 Pages: 317-336
91	INFERENCE DURING READING	MCKOON, G; RATCLIFF, R	: JUL 1992	PSYCHOLOGICAL REVIEW Volume: 99 Issue: 3 Pages: 440-466
92	HOW DO YOUNG-CHILDREN LEARN TO BE CONSUMERS - A SCRIPT-PROCESSING APPROACH	PERACCHIO, LA	: MAR 1992	JOURNAL OF CONSUMER RESEARCH Volume: 18 Issue: 4 Pages: 425-440
93	EXPLORING ACADEMIC LITERACY - AN EXPERIMENT IN COMPOSING	GEISLER, C	: FEB 1992	COLLEGE COMPOSITION AND COMMUNICATION Volume: 43 Issue: 1 Pages: 39-54
94	CAUSAL MODEL-BASED KNOWLEDGE ACQUISITION TOOLS - DISCUSSION OF EXPERIMENTS	CHARLET, J; KRIVINE, JP; REYNAUD, C	: 1992	LECTURE NOTES IN ARTIFICIAL INTELLIGENCE Volume: 599 Pages: 318-336
95	HOW TO DECREASE PRAGMATIC INFERENCES - THE IMPACT OF DIFFERENT TECHNIQUES	DARDENNE, B	: OCT 1991	CAHIERS DE PSYCHOLOGIE COGNITIVE-CURRENT PSYCHOLOGY OF COGNITION Volume: 11 Issue: 5 Pages: 555-584
96	CONTROLLING THE PROCESS OF LEARNING FROM AN EXAMPLE THROUGH ADAPTIVE GENERALIZATION OF EPISODIC MEMORY	SAWARAGI, T; IWAI, S; KATAI, O	: JAN 25 1991	FUZZY SETS AND SYSTEMS Volume: 39 Issue: 2 Pages: 133-162
97	CONCEPTUAL MODELS AND USER INTERACTION WITH COMPUTERS	CARDINALE, LA	: 1991	COMPUTERS IN HUMAN BEHAVIOR Volume: 7 Issue: 3 Pages: 163-169
98	THE PROCESSING OF COHESION DEVICES IN TEXT COMPREHENSION	EHRLICH, MF	: 1991	PSYCHOLOGICAL RESEARCH-PSYCHOLOGISCHE FORSCHUNG Volume: 53 Issue: 2 Pages: 169-174
99	ANALOGICAL LEARNING AND TRANSFER IN LANGUAGE-IMPAIRED CHILDREN	KAMHI, AG; GENTRY, B; MAUER, D; et al.	: FEB 1990	JOURNAL OF SPEECH AND HEARING DISORDERS Volume: 55 Issue: 1 Pages: 140-148
100	USE OF ANALOGIES IN PROBLEM-SOLVING	HESSE, FW; KLECHA, D	: 1990	COMPUTERS IN HUMAN BEHAVIOR Volume: 6 Issue: 1 Pages: 115-129
101	EFFECTS OF SEARCH AND SUBJECT EXPERTISE ON INFORMATION SEEKING IN A HYPERTEXT ENVIRONMENT	MARCHIONINI, G; LIN, X; DWIGGINS, S	: 1990	PROCEEDINGS OF THE ASIS ANNUAL MEETING Volume: 27 Pages: 129-142
102	THE INFLUENCE OF TEXTUAL CONTENT, TITLE AND PREVIOUS KNOWLEDGE ON TEXT COMPREHENSION	SPIEL, C	: 1990	ZEITSCHRIFT FUR EXPERIMENTELLE UND ANGEWANDTE PSYCHOLOGIE Volume: 37 Issue: 3 Pages: 505-518
103	PROBLEM-SOLVING, RECALL, AND MAPPING RELATIONS IN ISOMORPHIC TRANSFER AND NONISOMORPHIC TRANSFER AMONG PRESCHOOLERS AND ELEMENTARY-SCHOOL CHILDREN	GHOLSON, B; DATTEL, AR; MORGAN, D; et al.	: OCT 1989	CHILD DEVELOPMENT Volume: 60 Issue: 5 Pages: 1172-1187
104	CHILDRENS MENTAL ROTATIONS OF ABSTRACT DISPLAYS	BIALYSTOK, E	: FEB 1989	JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY Volume: 47 Issue: 1 Pages: 47-71
105	DEVELOPMENTAL DIFFERENCES IN THE USE OF VERBATIM VERSUS SPATIAL REPRESENTATIONS IN THE RECALL OF SPATIAL DESCRIPTIONS - A PROBABILISTIC MODEL AND AN EXPERIMENTAL-ANALYSIS	MORRA, S	: FEB 1989	JOURNAL OF MEMORY AND LANGUAGE Volume: 28 Issue: 1 Pages: 37-55
106	COGNITIVE SCIENCE - DEFINITION, STATUS, AND QUESTIONS	HUNT, E	: 1989	JOURNAL OF MEMORY AND LANGUAGE Volume: 28 Issue: 1 Pages: 37-55
107	THE LOGIC OF PLAUSIBLE REASONING - A CORE THEORY	COLLINS, A; MICHALSKI, R	: JAN-MAR 1989	ANNUAL REVIEW OF PSYCHOLOGY Volume: 40 Pages: 603-629
108	INFORMATION-SEEKING STRATEGIES OF NOVICES USING A FULL-TEXT ELECTRONIC ENCYCLOPEDIA	MARCHIONINI, G	: JAN 1989	COGNITIVE SCIENCE Volume: 13 Issue: 1 Pages: 1-49
109	LANGUAGE AND COGNITIVE PROCESSES Volume: 4 Issue: 3-4 Special Issue: SI Pages: SI263-SI286	Oakhill, Jane; Garnham, Alan; Vonk, Wietske	: 1989	JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE Volume: 40 Issue: 1 Pages: 54-66
110	SYMBOLS, ICONS, AND SIGN PROCESSES	DROSTE, FG	: 1989	LANGUAGE AND COGNITIVE PROCESSES Volume: 4 Issue: 3-4 Special Issue: SI Pages: SI263-SI286
111	COGNITIVE-PSYCHOLOGY AND TEXT-PROCESSING - FROM TEXT REPRESENTATION TO TEXT-WORLD	DENHIERE, G; BAUDET, S	: 1989	SEMIOTICA Volume: 77 Issue: 1-3 Pages: 271-293
112	PRESCHOOL-CHILDREN CAN LEARN TO TRANSFER - LEARNING TO LEARN AND LEARNING FROM EXAMPLE	BROWN, AL; KANE, MJ	: OCT 1988	COGNITIVE PSYCHOLOGY Volume: 20 Issue: 4 Pages: 493-523
113	DECISION PROBLEM STRUCTURING - GENERATING OPTIONS	KELLER, LR; HO, JL	: SEP-OCT 1988	IEEE TRANSACTIONS ON SYSTEMS MAN AND CYBERNETICS Volume: 18 Issue: 5 Pages: 715-728
114	WORKING MEMORY (OR THE M-OPERATOR) AND THE PLANNING OF CHILDRENS DRAWINGS	MORRA, S; MOIZO, C; SCOPESI, A	: AUG 1988	JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY Volume: 46 Issue: 1 Pages: 41-73

115	IMPACT OF PRIOR KNOWLEDGE OF INFORMATIONAL CONTENT AND ORGANIZATION ON LEARNING SEARCH PRINCIPLES IN A DATABASE	LINDE, L; BERGSTROM, M	: APR 1988	CONTEMPORARY EDUCATIONAL PSYCHOLOGY Volume: 13 Issue: 2 Pages: 90-101
116	PROBLEM-SOLVING, RECALL, ISOMORPHIC TRANSFER, AND NONISOMORPHIC TRANSFER AMONG 3RD-GRADE AND 4TH-GRADE CHILDREN	GHOLSON, B; EYMARD, LA; LONG, D; et al.	: JAN 1988	COGNITIVE DEVELOPMENT Volume: 3 Issue: 1 Pages: 37-53
117	PROBLEM-SOLVING, RECALL, AND ISOMORPHIC TRANSFER AMONG 3RD-GRADE AND 6TH-GRADE CHILDREN	GHOLSON, B; EYMARD, LA; MORGAN, D; et al.	: APR 1987	JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY Volume: 43 Issue: 2 Pages: 227-243
118	CONCEPT, KNOWLEDGE, AND THOUGHT	ODEN, GC	: 1987	ANNUAL REVIEW OF PSYCHOLOGY Volume: 38 Pages: 203-227
119	A QUALITATIVE MODEL OF HUMAN INTERACTION WITH COMPLEX DYNAMIC-SYSTEMS	HESS, RA	: JAN-FEB 1987	IEEE TRANSACTIONS ON SYSTEMS MAN AND CYBERNETICS Volume: 17 Issue: 1 Pages: 33-51
120	2 WAYS OF COPING WITH INDETERMINATE SPATIAL DISCOURSE	BREDART, S	: 1987	PSYCHOLOGICA BELGICA Volume: 27 Issue: 2 Pages: 143-151
121	THE COMPREHENSION OF ANAPHORIC RELATIONS	STEVENSON, RJ; VITKOVITCH, M	: OCT-DEC 1986	LANGUAGE AND SPEECH Volume: 29 Pages: 335-360 Part: 4
122	DEFAULT REASONING AND SELF-KNOWLEDGE	WILKS, Y	: OCT 1986	PROCEEDINGS OF THE IEEE Volume: 74 Issue: 10 Pages: 1399-1404
123	3 COMPONENTS OF UNDERSTANDING A PROGRAMMERS MANUAL - VERBATIM, PROPOSITIONAL, AND SITUATIONAL REPRESENTATIONS	SCHMALHOFER, F; GLAVANOV, D	: JUN 1986	JOURNAL OF MEMORY AND LANGUAGE Volume: 25 Issue: 3 Pages: 279-294
124	WHAT IS DOMAIN SKILL - FRAMES IN INFORMATION SEEKING	LINDE, LK	: APR 1986	BEHAVIORAL SCIENCE Volume: 31 Issue: 2 Pages: 89-102
125	YOUNG CHILDRENS MENTAL MODELS DETERMINE ANALOGICAL TRANSFER ACROSS PROBLEMS WITH A COMMON GOAL STRUCTURE	BROWN, AL; KANE, MJ; ECHOLS, CH	: APR 1986	COGNITIVE DEVELOPMENT Volume: 1 Issue: 2 Pages: 103-121
126	HOW CAN COGNITIVE-PSYCHOLOGY HELP SOLVE AN ARTIFICIAL-INTELLIGENCE PROBLEM	FARKAS, A	: 1986	COMPUTERS AND ARTIFICIAL INTELLIGENCE Volume: 4 Pages: 315-320
127	TOO LITTLE AND LATENT	MORTON, J	: 1985	BEHAVIORAL AND BRAIN SCIENCES Volume: 8 Issue: 1 Pages: 26-27
128	A REPRESENTATION SYSTEM OF COGNITIVE PROCEDURES	SAJDA, J	: 1985	COMPUTERS AND ARTIFICIAL INTELLIGENCE Volume: 4 Issue: 5 Pages: 407-419
129	THE CONTENT OF AWARENESS IS A MODEL OF THE WORLD	YATES, J	: 1985	PSYCHOLOGICAL REVIEW Volume: 92 Issue: 2 Pages: 249-284
130	IMPLICIT KNOWLEDGE, QUESTION ANSWERING, AND THE REPRESENTATION OF EXPOSITORY TEXT	GRAESSER, AC; GOODMAN, SM	: 1985	PSYCHOLOGY OF READING AND READING INSTRUCTION Pages: 109-171
131	COGNITIVE-ABILITIES, CONDITIONS OF LEARNING, AND THE EARLY DEVELOPMENT OF READING SKILL	EVANS, MA; CARR, TH	: 1985	READING RESEARCH QUARTERLY Volume: 20 Issue: 3 Pages: 327-350
132	LANGUAGE, THOUGHT, AND MENTAL MODELS	DROSTE, FG	: 1985	SEMIOTICA Volume: 56 Issue: 1-2 Pages: 31-98
133	SEMIOTIC PRINCIPLES AND SEMANTIC THEORIES	DROSTE, FG	: 1985	SEMIOTICA Volume: 54 Issue: 3-4 Pages: 429-458
134	HUMAN AND COMPUTER REASONING	JOHNSONLAIRD, PN	: 1985	TRENDS IN NEUROSCIENCES Volume: 8 Issue: 2 Pages: 54-57
135	FILM, PERCEPTION AND COGNITIVE PSYCHOLOGY	BROOKS, V	: 1985	MILLENNIUM FILM JOURNAL Issue: 14-1 Pages: 105-126
136	LEXICAL INFORMATION IN SENTENCE COMPREHENSION	TABOSSI, P	: 1985	QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY SECTION A-HUMAN EXPERIMENTAL PSYCHOLOGY Volume: 37 Issue: 1 Pages: 83-94
137	SYLLOGISTIC INFERENCE	JOHNSONLAIRD, PN; BARA, BG	: 1984	COGNITION Volume: 16 Issue: 1 Pages: 1-61
138	MODELS OF CONCEPTS	COHEN, B; MURPHY, GL	: 1984	COGNITIVE SCIENCE Volume: 8 Issue: 1 Pages: 27-58
139	THE ROLE OF A MENTAL MODEL IN LEARNING TO OPERATE A DEVICE	KIERAS, DE; BOVAIR, S	: 1984	COGNITIVE SCIENCE Volume: 8 Issue: 3 Pages: 255-273
140	EXPLAINING INTERPRETATION, THE CASE OF JAMES,HENRY THE 'TURN OF THE SCREW'	FREUNDLIEB, D	: 1984	POETICS TODAY Volume: 5 Issue: 1 Pages: 79-95
141	SOME EMPIRICAL JUSTIFICATION FOR A THEORY OF NATURAL PROPOSITIONAL LOGIC	BRAINE, MDS; REISER, BJ; RUMAIN, B	: 1984	PSYCHOLOGY OF LEARNING AND MOTIVATION-ADVANCES IN RESEARCH AND THEORY Volume: 18 Pages: 313-371
142	CAUSALITY AND SYNTAX - SOME LINGUISTIC DELINEATIONS IN KANT	DROSTE, FG	: 1984	TIJDSCHRIFT VOOR FILOSOFIE Volume: 46 Issue: 3 Pages: 445-465
143	SCHEMATA IN COGNITIVE ANTHROPOLOGY	CASSON, RW	: 1983	ANNUAL REVIEW OF ANTHROPOLOGY Volume: 12 Pages: 429-462
144	DIGIT MEMORY OF GRAND EXPERTS IN ABACUS-DERIVED MENTAL CALCULATION	HATANO, G; OSAWA, K	: 1983	COGNITION Volume: 15 Issue: 1-3 Pages: 95-110
145	WHATS WRONG WITH STORY GRAMMARS	GARNHAM, A	: 1983	COGNITION Volume: 15 Issue: 1-3 Pages: 145-154
146	BELIEFS ABOUT BELIEFS REPRESENTATION AND CONSTRAINING FUNCTION OF WRONG BELIEFS IN YOUNG CHILDRENS UNDERSTANDING OF DECEPTION	WIMMER, H; PERNER, J	: 1983	COGNITION Volume: 13 Issue: 1 Pages: 103-128
147	ANALOGICAL MODES OF REASONING AND PROCESS MODELING	FUNT, BV	: 1983	COMPUTER Volume: 16 Issue: 10 Pages: 99-104
148	COGNITIVE SYSTEMS-ENGINEERING - NEW WINE IN NEW BOTTLES	HOLLNAGEL, E; WOODS, DD	: 1983	INTERNATIONAL JOURNAL OF MAN-MACHINE STUDIES Volume: 18 Issue: 6 Pages: 583-600
149	SUBJECTIVE REPRESENTATION OF VELOCITY IN VERBS OF MOTION	WEBER, G; COLONIUS, H	: 1983	PSYCHOLOGICAL RESEARCH-PSYCHOLOGISCHE FORSCHUNG Volume: 45 Issue: 1 Pages: 73-90
150	ON THE NATURE OF INTELLIGENCE	HUNT, E	: 1983	SCIENCE Volume: 219 Issue: 4581 Pages: 141-146
151	itle: COGNITIVE AUTONOMY OF THE DEVELOPMENT OF MORAL EVALUATION OF ACHIEVEMENT	WIMMER, H; WACHTER, J; PERNER, J	: 1982	CHILD DEVELOPMENT Volume: 53 Issue: 3 Pages: 668-676
152	A TEMPORAL LOGIC FOR REASONING ABOUT PROCESSES AND PLANS	MCDERMOTT, D	: 1982	COGNITIVE SCIENCE Volume: 6 Issue: 2 Pages: 101-155
153	COMPUTATIONAL PROCESS MODELING OF SPATIAL COGNITION AND BEHAVIOR	SMITH, TR; PELLEGRINO, JW; GOLLEDGE, RG	: 1982	GEOGRAPHICAL ANALYSIS Volume: 14 Issue: 4 Pages: 305-325
154	ARTIFICIAL-INTELLIGENCE AND THE STUDY OF LANGUAGE	GOODWIN, JW; HEIN, U	: 1982	JOURNAL OF PRAGMATICS Volume: 6 Issue: 3-4 Pages: 241-280
155	ON BEING A MODEL	MEY, JL	: 1982	JOURNAL OF PRAGMATICS Volume: 6 Issue: 3-4 Pages: 281-292
156	STORIES ARE TO ENTERTAIN - A STRUCTURAL-AFFECT THEORY OF STORIES	BREWER, WF; LICHTENSTEIN, EH	: 1982	JOURNAL OF PRAGMATICS Volume: 6 Issue: 5-6 Pages: 473-486
157	SPATIAL DESCRIPTIONS AND REFERENTIAL CONTINUITY	EHRLICH, K; JOHNSONLAIRD, PN	: 1982	JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR Volume: 21 Issue: 3 Pages: 296-306
158	THE MENTAL REPRESENTATION OF SPATIAL DESCRIPTIONS	MANI, K; JOHNSONLAIRD, PN	: 1982	MEMORY & COGNITION Volume: 10 Issue: 2 Pages: 181-187
159	9TH BARTLETT-MEMORIAL-LECTURE - THINKING AS A SKILL	JOHNSONLAIRD, PN	: 1982	QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY SECTION A-HUMAN EXPERIMENTAL PSYCHOLOGY Volume: 34 Issue: FEB Pages: 1-29
160	FRAMEWORKS FOR A SCIENCE OF TEXTS - THE COMPLEAT SEMIOTICIAN	GRIMAUD, M	: 1982	SEMIOTICA Volume: 39 Issue: 3-4 Pages: 193-241
161	MENTAL MODELS AS REPRESENTATIONS OF TEXT	GARNHAM, A	: 1981	MEMORY & COGNITION Volume: 9 Issue: 6 Pages: 560-565
162	COMPREHENSION AS THE CONSTRUCTION OF MENTAL MODELS	JOHNSONLAIRD, PN	: 1981	PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY OF LONDON SERIES B-BIOLOGICAL SCIENCES Volume: 295 Issue: 1077 Pages: 353-374
163	LINGUISTIC CONTEXT AND THE PRIMING OF SEMANTIC INFORMATION	TABOSSI, P; JOHNSONLAIRD, PN	: 1980	QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY Volume: 32 Issue: NOV Pages: 595-603

2) Johnson-Laird, P., 2002, "Peirce, logic diagrams, and the elementary operations of reasoning" in *Thinking and Reasoning* , Vol. 8, No. 1, pp69-95.

Scopus - Article not listed (see http://www.scopus.com/source/sourceInfo.url?sourceId=15330&origin=sbrowse)
    0 citations
Web of Knowledge - Article not listed (see http://apps.webofknowledge.com/UA_GeneralSearch_input.do?SID=R1b3H3CBFBVef3RMmtr&product=UA&search_mode=GeneralSearch&errorQid=11#searchErrorMessage)
    0 citations

Google Scholar (see http://scholar.google.co.uk/scholar?cites=17818853831669484165&as_sdt=2005&sciodt=0,5&hl=en)
37 citations

Table 1c

Citation	Title	Author(s)	Publication	Journal/Book title (if appropriate)
1	The tasks for the practical logic	Zhaldak, N.	2013	
2	Architectural Design Students' Explorations through Conceptual Diagrams	F Dogan	2013	The Design Journal, vol. 16, no. 1
3	Carpenter, tractors and microbes for the development of logical–mathematical thinkingthe way 10th graders and pre-service teachers solve thinking challenges	A Gazit	2013	International Journal of Mathematical Education in Science and Technology, Volume 43, Issue 8,

4	THE ABSTRACTION EFFECT ON LOGIC RULES APPLICATION	G Macbeth, E Razumiejczyk, G Campitelli	2013	Education Science and Psychology, No.2(24)
5	Theories of the syllogism: A meta-analysis.	S Khemlani, PN Johnson-Laird	2012	Psychological bulletin, Vol 138(3)
6	Negation: A theory of its meaning, representation, and use	Sangeet Khemlania, Isabel Orenesb & P. N. Johnson-Laird	2012	Journal of Cognitive Psychology, Volume 24, Issue 5
7	A Moving Picture of Thought	JF Sowa	2012	CTF'12 Conference in Düsseldorf
8	Effect of difference in word formulation and mathematical characteristics of story problems on mathematics preservice teachers and practising teachers	D Patkin, A Gazit	2011	International Journal of Mathematical Education in Science and Technology, Volume 42, Issue 1
9	Existential Graphs: What a diagrammatic logic of cognition might look like	AV Pietarinen	2011	History and Philosophy of Logic, Volume 32, Issue 3
10	Diagrams, iconicity, and abductive discovery	S Paavola	2011	Semiotica, Vol. 2011, Iss. 186
11	Negation: A Theory of its Meaning, Representation, and Use	S Khemlani, I Orenes, PN Johnson-Laird	2011	Journal of Cognitive Psychology
12	Cognitive architectures for conceptual structures	JF Sowa	2011	Conceptual Structures for Discovering Knowledge, Lecture Notes in Computer Science Volume 6828
13	Mental Models in the Visualization of Conflict Escalation and Entrapment: Biases and Alternatives	A Bartoli, A Nowak, L Bui-Wrzosinska	2011	IACM 24th Annual Conference Paper
14	Cognitive illusions and the lying machine: a blueprint for sophistic mendacity	MH Clark	2010	
15	Biological and Psycholinguistic Influences on Architectures For Natural Language Processing	JF Sowa	2010	Biologically Inspired Cognitive Architectures 2010: Proceedings of the First Annual Meeting of the BICA Society
16	Challenges and Opportunities for Existential Graphs	AV Pietarinen	2010	Nordic Studies in Pragmatism
17	Human factors and data fusion as part of control systems resilience	DI Gertman	2009	Human System Interactions, 2009. HSI '09. 2nd Conference on
18	Conditional modulation and mental models	C Juhos	2009	
19	Mending or abandoning cognitivism	A Gomila	2008	Symbols, embodiment, and meaning
20	Teaching Analysis to Professional Writing Students: Heuristics Based on Expert Theories	SN Smith	2008	
21	Cephalic Organization: Animacy and Agency	J Schulkin	2008	Contemporary Pragmatism, Volume 5, Number 1,
22	Cognitive Adaptation: Insights from a Pragmatist Perspective	J Schulkin	2008	Contemporary Pragmatism, Volume 5, Number 1
23	Charles S. Peirce 'Theorie natürlicher Sprache und ihre Relevanz für die Linguistik: Logik, Semantik, Pragmatik	DH Rellstab	2007	
24	Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults	J Halberda	2006	Cognitive psychology
25	Mental models & the mind: current developments in cognitive psychology, neuroscience and philosophy of mind	C Held, M Gottfried, GV Vosgerau, M Knauff	2006	Advances in Psychology: Mental models and the mind
26	The perceptual nature of mental models	G Vosgerau	2006	Advances in Psychology, Volume 138
27	Deduktion und logisches Denken	M Knauff	2006	Enzyklopädie der Psychologie
28	The collapse illusion effect: A semantic-pragmatic illusion of truth and paradox	S Elqayam	2006	Thinking & Reasoning, Volume 12, Issue 2,
29	The semiotic self: developing interpretive identity as a dramatic process	M Michel	2006	
30	The Perceptual Nature of Mental Models	Carsten Held, Markus Knau, Gottfried Vosgerau	2006	Mental Models in Cognitive Psychology, Neuroscience, and Philosophy of Mind
31	Reasoning about relations.	GP Goodwin, PN Johnson-Laird	2005	Psychological review, Vol 112(2)
32	Mental models and thought	PN Johnson-Laird	2005	The Cambridge handbook of thinking and Reasoning
33	Design problem solving with conceptual diagrams	F Dogan, NJ Nersessian	2005	
34	Mental models, model-theoretic semantics, and the psychosemantic conception of truth	S Elqayam	2005	Philosophia scientiae, Vol. 9, No. 2.
35	The history of mental models	PN Johnson-Laird	2004	Psychology of reasoning: Theoretical and Historical Perspectives
36	From models to rules: Mechanization of reasoning as a way to cope with cognitive overloading in combinatorial problems	P Cherubini, A Mazzocco	2004	Acta Psychologica, Volume 116, Issue 3
37	The role of conceptual diagrams in the architectural design process: case studies of the First Unitarian Church	F Dogan	2003	

3) Barres P.E., Johnson-Laird P.N. 2003, "On imagining what is true (and what is false)" in Thinking and Reasoning, 9 (1), pp. 1-42.

Scopus
(See http://www.scopus.com/results/citedbyresults.url?sort=plf-f&cite=2-s2.0-0345012051&src=s&imp=t&sid=54A12D3C75203B48FBE40A0F6C589666.zQKnzAySRvJOZYcdfIziQ%3a600&sot=cite&sdt=a&sl=0&origin=resultslist&txGid=54A12D3C75203B48FBE40A0F6C589666.zQKnzAySRvJOZYcdfIziQ%3a60

20 citations

<u>Table 1d</u>

Citation	Title	Author(s)	Publication	Journal/Book title (if appropriate)
1	Toward a unified theory of reasoning	Johnson-Laird, P.N., Khemlani, S.S.	2014	Psychology of Learning and Motivation - Advances in Research and Theory 59 , pp. 1-42
2	The compatibility heuristic in non-categorical hypothetical reasoning: Inferences between conditionals and disjunctions	Espino, O., Byrne, R.M.J.	2013	Cognitive Psychology 67 (3) , pp. 98-129
3	It is not the case that if you understand a conditional you know how to negate it	Espino, O., Byrne, R.M.J.	2012	Journal of Cognitive Psychology 24 (3) , pp. 329-334
4	The consistency of disjunctive assertions	Johnson-Laird, P.N., Lotstein, M., Byrne, R.M.J.	2012	Memory and Cognition 40 (5), pp. 769-778
5	Reasoning from connectives and relations between entities	Mackiewicz, R., Johnson-Laird, P.N.	2012	Memory and Cognition 40 (2), pp. 266-279
6	Logic, models, and paradoxical inferences	Orenes, I., Johnson-Laird, P.N.	2012	Mind and Language 27 (4), pp. 357-377
7	The Cognitive Dynamics of Negated Sentence Verification	Dale, R., Duran, N.D.	2011	Cognitive Science 35 (5), pp. 983-996
8	Mistaking the instance for the rule: A critical analysis of the truth-table evaluation paradigm	Schroyens, W.	2010	Quarterly Journal of Experimental Psychology 63 (2), pp. 246-259
9	A meta-analytic review of thinking about what is true, possible, and irrelevant in reasoning from or reasoning about conditional propositions	Schroyens, W.	2010	European Journal of Cognitive Psychology 22 (6), pp. 897-921
10	At least one problem with "some" formal reasoning paradigms	Schmidt, J.R., Thompson, V.A.	2008	Memory and Cognition 36 (1), pp. 217-229
11	Action dynamics reveal parallel competition in decision making	McKinstry, C., Dale, R., Spivey, M.J.	2008	Psychological Science 19 (1), pp. 22-24
12	The collapse illusion effect: A semantic-pragmatic illusion of truth and paradox	Elqayam, S.	2006	Thinking and Reasoning 12 (2), pp. 144-180
13	The negated conditional: A litmus test for the suppositional conditional?	Handley, S.J., Evans, J.St.B.T., Thompson, V.A.	2006	Journal of Experimental Psychology: Learning Memory and Cognition 32 (3) , pp. 559-569
14	A hyper-emotion theory of psychological illnesses	Johnson-Laird, P.N., Mancini, F., Gangemi, A.	2006	Psychological Review 113 (4) , pp. 822-841
15	Mental models, sentential reasoning, and illusory inferences	Johnson-Laird, P.N.	2006	Advances in Psychology 138 (1), pp. 27-51
16	Naïve deontics: A theory of meaning, representation, and reasoning	Bucciarelli, M., Johnson-Laird, P.N.	2005	Cognitive Psychology 50 (2), pp. 159-193
17	Co-reference and reasoning	Walsh, C.R., Johnson-Laird, P.N.	2004	Memory and Cognition 32 (1), pp. 96-106
18	Reasoning from inconsistency to consistency	Johnson-Laird, P.N., Girotto, V., Legrenzi, P.	2004	Psychological Review 111 (3) , pp. 640-661
19	The probability of conditionals	Girotto, V., Johnson-Laird, P.N.	2004	Psychologia 47 (4), pp. 207-225
20	Counterexamples in sentential reasoning	Johnson-Laird, P.N., Hasson, U.	2003	Memory and Cognition 31 (7), pp. 1105-1113

Web of Knowledge - Article not listed (see http://apps.webofknowledge.com/UA_GeneralSearch.do#searchErrorMessage)
0 citations

37 citations

Table 1e

Citation	Title	Author(s)	Publication	Journal/Book title (if appropriate)
1	Toward a Unified Theory of Reasoning	PN Johnson-Laird, SS Khemlani	2014	Psychology of Learning and Motivation - Advances in Research and Theory 59 , pp. 1-42
2	The compatibility heuristic in non-categorical hypothetical reasoning: Inferences between conditionals and disjunctions	O Espino, RMJ Byrne	2013	Cognitive Psychology, Volume 67, Issue 3, 98-129
3	Mental models and consistency	PN Johnson-Laird	2012	Cognitive Consistency: A Unifying Concept in Social Psychology
4	Logic, models, and paradoxical inferences	I Orenes, PN JOHNSON-LAIRD	2012	Mind and Language 27 (4), pp. 357-377
5	The consistency of disjunctive assertions	PN Johnson-Laird, M Lotstein, RMJ Byrne	2012	Memory and Cognition 40 (5), pp. 769-778
6	It is not the case that if you understand a conditional you know how to negate it	O Espino, RMJ Byrne	2012	Journal of Cognitive Psychology, Volume 24, Issue 3, 329-334
7	Reasoning from connectives and relations between entities	R Mackiewicz, PN Johnson-Laird	2012	Memory & Cognition, Vol. 40, No. 2, 266-279
8	reasoning	R Dale, ND Duran	2011	Cognitive Science, Vol 35 Issue 5, 983-996
9	INFERENTIAL REASONING AND CREDIT CARDS	LF Ackert, BK Church	2011	
10	Reasoning from inconsistency to consistency.	PN Johnson-Laird, V Girotto, P Legrenzi	2011	Psychological Review, Vol 111(3), 640-661.
11	On the temporal dynamics of negated perceptual simulations	S E. Anderson, S Huette, T Matlock, & M. J. Spivey	2010	Meaning, Form, and Body
12	A meta-analytic review of thinking about what is true, possible, and irrelevant in reasoning from or reasoning about conditional propositions	W Schroyens	2010	European Journal of Cognitive Psychology 22 (6), pp. 897-921
13	Mistaking the instance for the rule: A critical analysis of the truth-table evaluation paradigm	W Schroyens	2010	Quarterly Journal of Experimental Psychology 63 (2), pp. 246-259
14	Conditionals and probability	V Girotto, PN Johnson-Laird	2010	Cognition and Conditionals: Probability and Logic in Human Thinking
15	Comprehending negated sentences with binary states and locations	SE Anderson, S Huette, T Matlock...	2010	mindmodelling.org
16	Possibilities: A Framework for Modeling Students' Deductive Reasoning in Physics	JDH Gaffney	2010	
17	Modele umysłowe i błędy w rozumo-waniu	SWP Społecznej	2010	Studia z Kognitywistyki: i Filozofii Umysłu
18	Action dynamics reveal parallel competition in decision making	C McKinstry, R Dale, MJ Spivey	2008	Psychological Science, vol. 19 no. 1 22-24
19	Mental models and deductive reasoning	PN Johnson-Laird	2008	Reasoning: Studies in Human Inference and Its Foundations
20	"At least one" problem with "some" formal reasoning paradigms	JR Schmidt, VA Thompson	2008	Memory & cognition, Vol. 31, No. 1, pp 217-229
21	Suchprozesse der Seele: die Psychologie des Erwägens	G Jüttemann	2008	Suchprozesse der Seele: die Psychologie des Erwägens
22	Il ruolo del ragionamento nella psicopatologia secondo la Hyper Emotion Theory	F Mancini, A Gangemi, PN Johnson-Laird	2007	Giornale italiano di psicologia, DOI: 10.1421/25829
23	A hyper-emotion theory of psychological illnesses.	PN Johnson-Laird, F Mancini, A Gangemi	2006	Psychological Review 113 (4) , pp. 822-841
24	The negated conditional: A litmus test for the suppositional conditional?	Handley, Simon J.; Evans, Jonathan St. B. T.; Thompson, Valerie A.	2006	Journal of Experimental Psychology: Learning, Memory, and Cognition, Vol 32(3), 559-569
25	Mental models, sentential reasoning, and illusory inferences	PN Johnson-Laird	2006	Advances in Psychology 138 (1) , pp. 27-51
26	Towards a flexible Bayesian and deontic logic of testing descriptive and prescriptive rules	M von Sydow	2006	
27	The collapse illusion effect: A semantic-pragmatic illusion of truth and paradox	S Elqayam	2006	Thinking & Reasoning, Volume 12, Issue 2, 144-180
28	Mental Models, Sentential Reasoning, and Illusory Inferences	P.N. Johnson-Laird	2006	Advances in Psychology, Volume 138, 27-51
29	Naïve deontics: A theory of meaning, representation, and reasoning	M Bucciarelli, PN Johnson-Laird	2005	Cognitive Psychology 50 (2) , pp. 159-193
30	Reasoning: Studies in Human Inference and Its Foundations	PN Johnson-Laird	2005	Reasoning: Studies in Human Inference and Its Foundations
31	Systematically Deviant Judgements of Validity	F Battaglini	2005	Utrecht Institute of Linguistics OTS, Yearbook 2005
32	The history of mental models	PN Johnson-Laird	2004	Psychology of Reasoning: Theoretical and Historical Perspectives
33	The probability of conditionals	V Girotto, PN Johnson-Laird	2004	Psychologia 47 (4), pp. 207-225
34	Co-reference and reasoning	CR Walsh, PN Johnson-Laird	2004	Memory and Cognition 32 (1) , pp. 96-106
35	Counterexamples in sentential reasoning	PN Johnson-Laird, U Hasson	2003	Memory & Cognition, Volume 31, Issue 7, pp 1105-1113
36	An antidote to illusory inferences	C Santamaría, PN Johnson-Laird	2000	Thinking & Reasoning, Volume 6, Issue 4, 313-333
37	A mental-model-theoretic account of Free Choice Permission	F Battaglini		Availible online (http://jmprieto.us.es/folli/cds/2004/content/readers/61.pdf)

4) Khemlani S., Orenes I., Johnson-Laird P.N., 2012, "Negation: A theory of its meaning, representation, and use" in Journal of Cognitive Psychology, 24 (5), pp. 541-559.

2 citations

Table 1f

Citation	Title	Author(s)	Publication	Journal/Book title (if appropriate)
1	Toward a Unified Theory of Reasoning	Johnson-Laird, P. N.; Khemlani, Sangeet S.	2013	Psychology of Learning and Motivation Volume: 59 Pages: 1-42
2	The compatibility heuristic in non-categorical hypothetical reasoning: Inferences between conditionals and disjunctions	Espino, O., Byrne, R.M.J.	2013	Cognitive Psychology 67 (3) , pp. 98-129

1 citation

Table 1g

Citation	Title	Author(s)	Publication	Journal/Book title (if appropriate)
1	Toward a Unified Theory of Reasoning	Johnson-Laird, P. N.; Khemlani, Sangeet S.	2013	Psychology of Learning and Motivation Volume: 59 Pages: 1-42

10 citations

Table 1h

Citation	Title	Author(s)	Publication	Journal/Book title (if appropriate)
1	The processes of inference	S Khemlani, PN Johnson-Laird	2013	Argument & Computation, Volume 4, Issue 1,
2	The consistency of disjunctive assertions	PN Johnson-Laird, M Lotstein, RMJ Byrne	2013	Memory and Cognition, Vol. 40

3	Scope ambiguities and conditionals	D Over, I Douven, S Verbrugge	2013	Thinking & Reasoning, DOI: 10.1080/13546783.2013.810172
4	The evaluation of the consistency of quantified assertions	M Ragni, S Khemlani, PN Johnson-Laird	2013	Memory and Cognition, DOI: 10.3758/s13421-013-0349-y
5	Toward a Unified Theory of Reasoning	PN Johnson-Laird, SS Khemlani	2013	Psychology of Learning and Motivation Volume: 59 Pages: 1-42
6	The compatibility heuristic in non-categorical hypothetical reasoning: Inferences between conditionals and disjunctions	O Espino, RMJ Byrne	2013	Cognitive Psychology, Volume 67, Issue 3,
7	Theories of the syllogism: A meta-analysis.	S Khemlani, PN Johnson-Laird	2012	Psychological Bulletin, Vol 138(3),
8	The Erotetic Theory of Reasoning	P Koralus, S Mascarenhas	2012	
9	Negating compound sentences	S Khemlani, I Orenes, PN Johnson-Laird	2012	mindmodelling.org
10	The modulation of conditional assertions and its effects on reasoning	AC Quelhas, PN Johnson-Laird, Csongor Juhosa	2010	The Quarterly Journal of Experimental Psychology, Volume 63, Issue 9

Unique Citations

Below are the unique citations. These have been generated by removing any repeat citations, citations from other Philip Johnson-Laird works (as we are only concerned with the critical analysis found within the work of commentators of the theory), citations to which access is not available and any errors (i.e. any reported citations that do not actually cite the concerned works).

Table 2

Citation	Title	Author(s)	Publication	Journal/Book title (if appropriate)
1	A cultural lens on interpersonal conflict and creativity in multicultural environments	Paletz, S.B.F., Miron-Spektor, E., Lin, C.-C.	2014	Psychology of Aesthetics, Creativity, and the Arts
2	Can pictures say no or not? Negation and denial in the visual mode	Oversteegen, E., Schilperoord, J.	2014	Journal of Pragmatics
3	Facilitating shared understanding of work situations using a tangible tabletop interface	Oppl, S., Stary, C.	2014	Behaviour and Information Technology
4	How negation is understood: Evidence from the visual world paradigm	Orenes, I., Beltrán, D., Santamaría, C.	2014	Journal of Memory and Language
5	Neural correlates and network connectivity underlying narrative production and comprehension: A combined fMRI and PET study	AbdulSabur, N.Y., Xu, Y., Liu, S., (...), Carson, J., Braun, A.R.	2014	Cortex
6	The processing of contradictory and non-contradictory negative sentences	Du, P., Liu, D., Zhang, L., Hitchman, G., Lin, C.	2014	Journal of Cognitive Psychology
7	Understanding agile software development practices using shared mental models theory	Yu, X., Petter, S.	2014	Information and Software Technology
8	Cognition, ontologies and description logics	P Warren	2014	
9	Mental Models for the Negation of Conjunctions and Disjunctions.	G Macbeth, E Razumiejczyk...	2014	Europe's Journal of ..., 2014
10	Reasoning with and without Reasons: The Effects of Professional Culture and Information Access in Educational and Clinical Settings	B Saferstein	2014	Journal of Cognition and Neuroethics
11	The Usability of Description Logics	P Warren, P Mulholland, T Collins, E Motta	2014	The Semantic Web: Trends ...,
12	How the design of presentation slides affects audience comprehension: A case for the assertion-evidence approach			
13	Architectural Design Students' Explorations through Conceptual Diagrams	F Dogan	2013	The Design Journal, vol. 16, no. 1
14	Carpenter, tractors and microbes for the development of logical–mathematical thinking–the way 10th graders and pre-service teachers solve thinking challenges	A Gazit	2013	International Journal of Mathematical Education in Science and Technology, Volume 43, Issue 8,
15	Scope ambiguities and conditionals	D Over, I Douven, S Verbrugge	2013	Thinking & Reasoning, DOI:10.1080/13546783.2013.810172
16	THE ABSTRACTION EFFECT ON LOGIC RULES APPLICATION	G Macbeth, E Razumiejczyk, G Campitelli	2013	Education Science and Psychology, No.2(24)
17	The compatibility heuristic in non-categorical hypothetical reasoning: Inferences between conditionals and disjunctions	Espino, O., Byrne, R.M.J.	2013	Cognitive Psychology 67 (3), pp. 98-129
18	The tasks for the practical logic	Zhaldak, N.	2013	
19	How the design of presentation slides affects audience comprehension: A case for the assertion-evidence approach	Garner, J.K., Alley, M.P.	2013	International Journal of Engineering Education
20	Manipulation of mental models of anatomy in interventional radiology and its consequences for design of human-computer interaction	Varga, E., Pattynama, P.M.T., Freudenthal, A.	2013	Cognition, Technology and Work
21	Scope ambiguities and conditionals	Over, D., Douven, I., Verbrugge, S.	2013	Thinking and Reasoning
22	The psychology of creativity: Mental models in design teams (Book Chapter)	Casakin, H., Badke-Schaub, P.	2013	Psychology of Creativity: Advances in Theory, Research and Application
23	Transfer of Nature of Science Understandings into Similar Contexts: Promises and Possibilities of an Explicit Reflective Approach	Khishfe, R.	2013	International Journal of Science Education
24	How the Design of Presentation Slides Affects Audience Comprehension: A Case for the Assertion-Evidence Approach	Garner, Joanna K.; Alley, Michael P.;	2013	INTERNATIONAL JOURNAL OF ENGINEERING EDUCATION Volume: 29 Issue: 6 Pages: 1564-1579
25	Death of the Internet (Book)	Jakobsson, M.	2012	Death of the Internet
26	Extracting narrative timelines as temporal dependency structures	Kolomiyets, O., Bethard, S., Moens, M.-F.	2012	50th Annual Meeting of the Association for Computational Linguistics, ACL 2012 - Proceedings of the Conference
27	Annotating Story Timelines as Temporal Dependency Structures	Bethard, Steven; Kolomiyets, Oleksandr; Moens, Marie-Francine	2012	Edited ; Calzolari, N; Choukri, K; Declerck, T; et al.; Conference: 8th International Conference on Language Resources and Evaluation (LREC) Location: Istanbul, TURKEY Date: MAY 21-27, 2012 Sponsor(s): Nuance; CELI Language & Informat Technol; European Media Lab GmbH (EML); IMMI; META-Net; Quaero; K-dictionariesLREC 2012 - EIGHTH INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION Pages: 2721-2726
28	The Erotetic Theory of Reasoning	P Koralus, S Mascarenhas	2012	
29	It is not the case that if you understand a conditional you know how to negate it	O Espino, RMJ Byrne	2012	Journal of Cognitive Psychology, Volume 24, Issue 3, 329-334
30	Cognitive architectures for conceptual structures	JF Sowa	2011	Conceptual Structures for Discovering Knowledge, Lecture Notes in Computer Science Volume 6828
31	Diagrams, iconicity, and abductive discovery	S Paavola	2011	Semiotica, Vol. 2011, Iss. 186
32	Effect of difference in word formulation and mathematical characteristics of story problems on mathematics preservice teachers and practising teachers	D Patkin, A Gazit	2011	International Journal of Mathematical Education in Science and Technology, Volume 42, Issue 1
33	Existential Graphs: What a diagrammatic logic of cognition might look like	AV Pietarinen	2011	History and Philosophy of Logic, Volume 32, Issue 3
34	INFERENTIAL REASONING AND CREDIT CARDS	LF Ackert, BK Church	2011	
35	Mental Models in the Visualization of Conflict Escalation and Entrapment: Biases and Alternatives	A Bartoli, A Nowak, L Bui-Wrzosinska	2011	IACM 24th Annual Conference Paper
36	The Cognitive Dynamics of Negated Sentence Verification	Dale, R., Duran, N.D.	2011	Cognitive Science 35 (5), pp. 983-996
37	On the temporal dynamics of negated perceptual simulations	S E. Anderson, S Huette, T Matlock, & M. J. Spivey	2010	Meaning, Form, and Body
38	A meta-analytic review of thinking about what is true, possible, and irrelevant in reasoning from or reasoning about conditional propositions	Schroyens, W.	2010	European Journal of Cognitive Psychology 22 (6), pp. 897-921
39	Biological and Psycholinguistic Influences on Architectures For Natural Language Processing	JF Sowa	2010	Biologically Inspired Cognitive Architectures 2010: Proceedings of the First Annual Meeting of the BICA Society
40	Challenges and Opportunities for Existential Graphs	AV Pietarinen	2010	Nordic Studies in Pragmatism
41	Cognitive illusions and the lying machine: a blueprint for sophistic mendacity	MH Clark	2010	
42	Comprehending negated sentences with binary states and locations	SE Anderson, S Huette, T Matlock...	2010	mindmodelling.org
43	Modele umysłowe i błędy w rozumo-waniu	SWP Społecznej	2010	Studia z Kognitywistyki: i Filozofii Umysłu
44	Possibilities: A Framework for Modeling Students' Deductive Reasoning in Physics	JDH Gaffney	2010	
45	Conditional modulation and mental models	C Juhos	2009	
46	Human factors and data fusion as part of control systems resilience	DI Gertman	2009	Human System Interactions, 2009. HSI '09. 2nd Conference on
47	Carbohydrate-carbohydrate interaction prominence in 3D supramolecular self-assembly	Fuss, Martina; Luna, Monica; Alcantara, David; et al.	2008	JOURNAL OF PHYSICAL CHEMISTRY B Volume: 112 Issue: 37 Pages: 11595-11600
48	"At least one" problem with "some" formal reasoning paradigms	JR Schmidt, VA Thompson	2008	Memory & cognition, Vol. 31, No. 1, pp 217-229
49	Action dynamics reveal parallel competition in decision making	McKinstry, C., Dale, R., Spivey, M.J.	2008	Psychological Science 19 (1), pp. 22-24
50	Cephalic Organization: Animacy and Agency	J Schulkin	2008	Contemporary Pragmatism, Volume 5, Number 1,
51	Cognitive Adaptation: Insights from a Pragmatist Perspective	J Schulkin	2008	Contemporary Pragmatism, Volume 5, Number 1
52	Mending or abandoning cognitivism	A Gomila	2008	Symbols, embodiment, and meaning
53	Suchprozesse der Seele: die Psychologie des Erwägens	G Jüttemann	2008	Suchprozesse der Seele: die Psychologie des Erwägens
54	Teaching Analysis to Professional Writing Students: Heuristics Based on Expert Theories	SN Smith	2008	
55	Psycholinguistic models of speech production in bilingualism and multilingualism (Book Chapter)	Marini, A., Fabbro, F.	2007	Speech and Language Disorders in Bilinguals

56	Medical costs and resource utilization for hemophilia patients with and without HIV or HCV infection	Tencer, Thomas; Friedman, Howard S.; Li-Mcleod, Josephine; et al.	2007	JOURNAL OF MANAGED CARE PHARMACY Volume: 13 Issue: 9 Pages: 790-798
57	Charles S. Peirce 'Theorie natürlicher Sprache und ihre Relevanz für die Linguistik: Logik, Semantik, Pragmatik	DH Rellstab	2007	
58	The effectiveness of online task support vs. instructor-led training (Book Chapter)	Mao, J.-Y., Brown, B.R.	2006	Contemporary Issues in End User Computing
59	Deduktion und logisches Denken	M Knauff	2006	Enzyklopädie der Psychologie
60	Mental models & the mind: current developments in cognitive psychology, neuroscience and philosophy of mind	C Held, M Gottfried, GV Vosgerau, M Knauff	2006	Advances in Psychology: Mental models and the mind
61	The collapse illusion effect: A semantic-pragmatic illusion of truth and paradox	Elqayam, S.	2006	Thinking and Reasoning 12 (2) , pp. 144-180
62	The negated conditional: A litmus test for the suppositional conditional?	Handley, S.J., Evans, J.St.B.T., Thompson, V.A.	2006	Journal of Experimental Psychology: Learning Memory and Cognition 32 (3) , pp. 559-569
63	The Perceptual Nature of Mental Models	C. Held, M. Knau , G. Vosgerau	2006	Mental Models in Cognitive Psychology, Neuroscience, and Philosophy of Mind
64	Towards a flexible Bayesian and deontic logic of testing descriptive and prescriptive rules	M von Sydow	2006	
65	Seeing Change in Time: Video Games to Teach about Temporal Change in Scientific Phenomena	Corredor, Javier; Gaydos, Matthew; Squire, Kurt	Jul-05	JOURNAL OF SCIENCE EDUCATION AND TECHNOLOGY Volume: 23 Issue: 3 Pages: 324-343
66	Standard regression-based methods for measuring recovery after sport-related concussion	McCrea, M; Barr, WB; Guskiewicz, K; et al.	2005	JOURNAL OF THE INTERNATIONAL NEUROPSYCHOLOGICAL SOCIETY Volume: 11 Issue: 1 Pages: 58-69
67	Design problem solving with conceptual diagrams	F Dogan, NJ Nersessian	2005	
68	Mental models, model-theoretic semantics, and the psychosemantic conception of truth	S Elqayam	2005	Philosophia scientiae, Vol. 9, No. 2.
69	Systematically Deviant Judgements of Validity	F Battaglini	2005	Utrecht Institute of Linguistics OTS, Yearbook 2005
70	From models to rules: Mechanization of reasoning as a way to cope with cognitive overloading in combinatorial problems	P Cherubini, A Mazzocco	2004	Acta Psychologica, Volume 116, Issue 3
71	The role of conceptual diagrams in the architectural design process: case studies of the First Unitarian Church	F Dogan	2003	
72	A mental-model-theoretic account of Free Choice Permission	F Battaglini		Availible online (http://jmprieto.us.es/folli/cds/2004/content/readers/61.pdf)
73	Mistaking the instance for the rule: A critical analysis of the truth-table evaluation paradigm	W Schroyens	2010	Quarterly Journal of Experimental Psychology 63 (2) , pp. 246-259
74	How a picture facilitates the process of learning from text: Evidence for scaffolding	Eitel, Alexander; Scheiter, Katharina; Schueler, Anne; et al.	2013	LEARNING AND INSTRUCTION Volume: 28 Pages: 48-63
75	Affordances of Augmented Reality in Science Learning: Suggestions for Future Research	Cheng, Kun-Hung; Tsai, Chin-Chung	2013	JOURNAL OF SCIENCE EDUCATION AND TECHNOLOGY Volume: 22 Issue: 4 Pages: 449-462
76	Climate and narrative: Environmental knowledge in everyday life	Lejano, Raul P.; Tavares-Reager, Joana; Berkes, Fikret	: AUG 2013	ENVIRONMENTAL SCIENCE & POLICY Volume: 31 Pages: 61-70
77	Reuse and Recycle: The Development of Adaptive Expertise, Routine Expertise, and Novelty in a Large Research Team	Paletz, Susannah B. F.; Kim, Kevin H.; Schunn, Christian D.; et al.	: JUL 2013	APPLIED COGNITIVE PSYCHOLOGY Volume: 27 Issue: 4 Pages: 415-428
78	Aging and predicting inferences: A diffusion model analysis	McKoon, Gail; Ratcliff, Roger	: APR 2013	JOURNAL OF MEMORY AND LANGUAGE Volume: 68 Issue: 3 Pages: 240-254
79	Seeking a unified framework for cerebellar function and dysfunction: from circuit operations to cognition	D'Angelo, Egidio; Casali, Stefano	: JAN 10 2013	FRONTIERS IN NEURAL CIRCUITS Volume: 6 Article Number: 116
80	Visuality for Architects: Architectural Creativity and Modern Theories of Perception and Imagination	Mitrovic, B	: 2013	VISUALITY FOR ARCHITECTS: ARCHITECTURAL CREATIVITY AND MODERN THEORIES OF PERCEPTION AND IMAGINATION Pages: 1-120
81	Shared Mental Models for Collaborative Mechanical Design in Mechanical Engineering	Du, Junmin	: 2013	ADVANCED RESEARCH ON INTELLIGENT SYSTEMS AND MECHANICAL ENGINEERING Book Series:Advanced Materials Research Volume: 644 Pages: 378-381
82	Exploring visual perception and children's interpretations of picture books	Yu, Xinyu	: OCT 2012	LIBRARY & INFORMATION SCIENCE RESEARCH Volume: 34 Issue: 4 Pages: 292-299
83	Effectiveness of holistic mental model confrontation in driving conceptual change	Gadgil, Soniya; Nokes-Malach, Timothy J.; Chi, Michelene T. H.	: FEB 2012	LEARNING AND INSTRUCTION Volume: 22 Issue: 1 Pages: 47-61
84	LANGUAGE, LINGUISTICS AND COGNITION	Baggio, Giosue; van Lambalgen, Michiel; Hagoort, Peter	: 2012	PHILOSOPHY OF LINGUISTICS Book Series: Handbook of the Philosophy of Science Pages: 325-355
85	LEXICAL ENTRAINMENT AND SUCCESS IN STUDENT ENGINEERING GROUPS	Friedberg, Heather; Litman, Diane; Paletz, Susannah B. F.	: 2012	2012 IEEE WORKSHOP ON SPOKEN LANGUAGE TECHNOLOGY (SLT 2012) Pages: 404-409
86	Stakeholder-Driven Collaborative Modeling of Subject-Oriented Business Processes	Wachholder, Dominik; Oppl, Stefan	: 2012	S-BPM ONE - SCIENTIFIC RESEARCH Book Series: Lecture Notes in Business Information Processing Volume: 104 Pages: 145-162
87	Evolution of a predictive internal model in an embodied and situated agent	Gigliotta, Onofrio; Pezzulo, Giovanni; Nolfi, Sefano	: DEC 2011	THEORY IN BIOSCIENCES Volume: 130 Issue: 4 Pages: 259-276
88	Distributed Cognition as a Lens to Understand the Effects of Scaffolds: The Role of Transfer of Responsibility	Belland, Brian R.	: DEC 2011	EDUCATIONAL PSYCHOLOGY REVIEW Volume: 23 Issue: 4 Pages: 577-600
89	Evidence Maps: Communicating Risk Assessments in Societal Controversies: The Case of Engineered Nanoparticles	Wiedemann, Peter; Schuetz, Holger; Spangenberg, Albena; et al.	: NOV 2011	RISK ANALYSIS Volume: 31 Issue: 11 Pages: 1770-1783
90	Emulation as an integrating principle for cognition	Colder, Brian	: MAY 27 2011	FRONTIERS IN HUMAN NEUROSCIENCE Volume: 5 Article Number: 54
91	The Influence of Connectives on Young Readers' Processing and Comprehension of Text	Cain, Kate; Nash, Hannah M.	: MAY 2011	JOURNAL OF EDUCATIONAL PSYCHOLOGY Volume: 103 Issue: 2 Pages: 429-441
92	Subject-Oriented Elicitation of Distributed Business Process Knowledge	Oppl, Stefan	: 2011	S-BPM ONE - LEARNING BY DOING - DOING BY LEARNING Book Series: Communications in Computer and Information Science Volume: 213 Pages: 16-33
93	An Observation-Based Method for Measuring the Sharedness of Mental Models in Teams	Badke-Schaub, Petra; Neumann, Andre; Lauche, Kristina	: 2011	COORDINATION IN HUMAN AND PRIMATE GROUPS Pages: 177-197
94	A multi-level approach to the analysis of narrative language in aphasia	Marini, Andrea; Andreetta, Sara; del Tin, Silvana; et al.	: 2011	APHASIOLOGY Volume: 25 Issue: 11 Special Issue: SI Pages: 1372-1392
95	How Audiences Decide: A Cognitive Approach to Business Communication	Young, RO	: 2011	HOW AUDIENCES DECIDE: A COGNITIVE APPROACH TO BUSINESS COMMUNICATION Pages: 1-398
96	Inclusion and interaction: Designing interaction for inclusive populations	Langdon, Patrick; Thimbleby, Harold	: NOV 2010	INTERACTING WITH COMPUTERS Volume: 22 Issue: 6 Special Issue: SI Pages: 439-448
97	Developing a model of cognitive interaction for analytical inclusive design evaluation	Langdon, Patrick; Persad, Umesh; Clarkson, P. John	: NOV 2010	INTERACTING WITH COMPUTERS Volume: 22 Issue: 6 Special Issue: SI Pages: 510-529
98	Between understanding and misunderstanding	Verdonik, Darinka	: MAY 2010	JOURNAL OF PRAGMATICS Volume: 42 Issue: 5 Pages: 1364-1379
99	The Breakdown of Coordinated Decision Making in Distributed Systems	Bearman, Christopher; Paletz, Susannah B. F.; Orasanu, Judith; et al.	: APR 2010	HUMAN FACTORS Volume: 52 Issue: 2 Special Issue: SI Pages: 173-188
100	Contrast and Assimilation Effects of Processing Fluency	Shen, Hao; Jiang, Yuwei; Adaval, Rashmi	: FEB 2010	JOURNAL OF CONSUMER RESEARCH Volume: 36 Issue: 5 Pages: 876-889
101	Shifting focus from the determinants to the origin: the foundations of a dynamic view of managerial discretion	Kleindienst, Ingo; Hutzschenreuter, Thomas	: 2010	HANDBOOK OF RESEARCH ON STRATEGY PROCESS Pages: 22-42
102	Synthetic Environments as visualization method for product design	Meijer, Frank; van den Broek, Egon L.; Schouten, Theo E.; et al.	: 2010	HUMAN VISION AND ELECTRONIC IMAGING XV Book Series: Proceedings of SPIE Volume: 7527 Article Number: 752712
103	The Nature of the Refutation Text Effect: An Investigation of Attention Allocation	Broughton, Suzanne H.; Sinatra, Gale M.; Reynolds, Ralph E.	: 2010	JOURNAL OF EDUCATIONAL RESEARCH Volume: 103 Issue: 6 Pages: 407-423
104	Language, Space, and Social Relationships: A Foundational Cultural Model in Polynesia	Bennardo, G	: 2009	LANGUAGE, SPACE, AND SOCIAL RELATIONSHIPS: A FOUNDATIONAL CULTURAL MODEL IN POLYNESIA Book Series: Language Culture and Cognition Issue: 9 Pages: 1-371
105	ITERATION IN SYSTEMS ANALYSIS AND DESIGN Cognitive Processes and Representational Artifacts	Berente, Nicholas; Lyytinen, Kalle	: 2009	SYSTEMS ANALYSIS AND DESIGN: TECHNIQUES, METHODOLOGIES, APPROACHES, AND ARCHITECTURES Book Series: Advances in Management Information Systems Volume: 15 Pages: 37-57
106	The language of schizophrenia: An analysis of micro and macrolinguistic abilities and their neuropsychological correlates	Marini, Andrea; Spoletini, Ilaria; Rubino, Ivo Alex; et al.	: OCT 2008	SCHIZOPHRENIA RESEARCH Volume: 105 Issue: 1-3 Pages: 144-155
107	nanoHUB.org: Advancing education and research in nanotechnology	Klimeck, Gerhard; McLennan, Michael; Brophy, Sean B.; et al.	: SEP-OCT 2008	COMPUTING IN SCIENCE & ENGINEERING Volume: 10 Issue: 5 Pages: 17-23
108	Advancing Engineering Education in P-12 Classrooms	Brophy, Sean; Klein, Stacy; Portsmore, Merredith; et al.	: JUL 2008	JOURNAL OF ENGINEERING EDUCATION Volume: 97 Issue: 3 Pages: 369-387

#	Title	Authors	Date	Publication
109	The role of working memory components in multimedia comprehension	Gyselinck, Valerie; Jamet, Eric; Dubois, Veronique	: APR 2008	APPLIED COGNITIVE PSYCHOLOGY Volume: 22 Issue: 3 Pages: 353-374
110	Toward a (pragmatic) science of strategic intervention: Design propositions for scenario planning	Hodgkinson, Gerard P.; Healey, Mark P.	: MAR 2008	ORGANIZATION STUDIES Volume: 29 Issue: 3 Pages: 435-457
111	Pedagogical Tools for the Training of Subtitlers	Taylor, Christopher	: 2008	AUDIOVISUAL TRANSLATION: LANGUAGE TRANSFER ON SCREEN Pages: 214-228
112	Frontal eye fields involved in shifting frame of reference within working memory for scenes	Wallentin, Mikkel; Roepstorff, Andreas; Burgess, Neil	: 2008	NEUROPSYCHOLOGIA Volume: 46 Issue: 2 Pages: 399-408
113	Constructing Mental Models during Learning from Science Text Eye Tracking Methodology Meets Conceptual Change	Mikkila-Erdmann, M.; Penttinen, M.; Anto, E.; et al.	: 2008	UNDERSTANDING MODELS FOR LEARNING AND INSTRUCTION: ESSAYS IN HONOR OF NORBERT M. SEEL Pages: 63-79
114	The 'Ins' and 'Outs' of Learning: Internal Representations and External Visualizations	Rapp, David N.; Kurby, Christopher A.	: 2008	VISUALIZATION: THEORY AND PRACTICE IN SCIENCE EDUCATION Book Series: Models and Modeling in Science Education Volume: 3 Pages: 29-52
115	Closed Loop State Space Analysis and Simulation for Cognitive Systems	Oberheid, Hendrik; Gamrad, Dennis; Soeffker, Dirk	: 2008	2008 8TH INTERNATIONAL CONFERENCE ON APPLICATION OF CONCURRENCY TO SYSTEM DESIGN, PROCEEDINGS Book Series: International Conference on Application of Concurrency to System Design. Proceedings Pages: 39-44
116	Thought experimenting as mental modeling: Empiricism without logic	Nersessian, Nancy J.	: 2007	CROATIAN JOURNAL OF PHILOSOPHY Volume: 7 Issue: 20 Pages: 125-161
117	Understanding instructional leadership through the mental models of three elementary school principals	Ruff, WG; Shoho, AR	: AUG 2005	EDUCATIONAL ADMINISTRATION QUARTERLY Volume: 41 Issue: 3 Pages: 554-577
118	The role played by the right hemisphere in the organization of complex textual structures	Marini, A; Carlomagno, S; Caltagirone, C; et al.	: APR 2005	BRAIN AND LANGUAGE Volume: 93 Issue: 1 Pages: 46-54
119	INTRODUCTION: MAKING SOCIAL MEANINGS IN CONTEXTS: Writing, Literate Activities and Learning Processes in Sociohistorical Communities	Kostouli, Triantafillia	: 2005	WRITING IN CONTEXT(S): TEXTUAL PRACTICES AND LEARNING PROCESSES IN SOCIOCULTURAL SETTINGS Book Series: Studies in Writing Volume: 15 Pages: 1-+
120	Language processing: Competing theories or flexibility of the linguistic object?	Groeben, N	: 2005	ZEITSCHRIFT FUR PSYCHOLOGIE Volume: 213 Issue: 3 Pages: 152-158
121	Exploiting concept mapping in a semantic web environment	Giouvanakis, T; Fragidis, G; Kehris, E; et al.	: 2005	5th IEEE International Conference on Advanced Learning Technologies, Proceedings Pages: 432-434
122	Framing-effects approach: A theoretical and methodological critique	Scheufele, Bertram	: DEC 2004	COMMUNICATIONS-EUROPEAN JOURNAL OF COMMUNICATION RESEARCH Volume: 29 Issue: 4 Pages: 401-428
123	The quality of human-automation cooperation in human-system interface for nuclear power plants	Skjerve, ABM; Skraaning, G	: NOV 2004	INTERNATIONAL JOURNAL OF HUMAN-COMPUTER STUDIES Volume: 61 Issue: 5 Pages: 649-677
124	Object and event concepts: A cognitive mechanism of incommensurability	Chen, X	: DEC 2003	PHILOSOPHY OF SCIENCE Volume: 70 Issue: 5 Pages: 962-974
125	Why did John Herschel fail to understand polarization? The differences between object and event concepts	Chen, X	: SEP 2003	STUDIES IN HISTORY AND PHILOSOPHY OF SCIENCE Volume: 34A Issue: 3 Pages: 491-513
126	The influence of pre-decisional constraints on information search and consideration set formation in new automobile purchases	Punj, G; Brookes, R	: DEC 2002	INTERNATIONAL JOURNAL OF RESEARCH IN MARKETING Volume: 19 Issue: 4 Pages: 383-400 Article Number: PII S0167-8116(02)00100-3
127	Visuospatial memory and phonological loop in learning from multimedia	Gyselinck, V; Cornoldi, C; Dubois, V; et al.	: SEP 2002	APPLIED COGNITIVE PSYCHOLOGY Volume: 16 Issue: 6 Pages: 665-685
128	The role of prototypes in the mental representation of temporally related events	Colcombe, SJ; Wyer, RS	: FEB 2002	COGNITIVE PSYCHOLOGY Volume: 44 Issue: 1 Pages: 67-103
129	Narrative-based representations of social knowledge: Their construction and use in comprehension, memory, and judgment	Wyer, RS; Adaval, R; Colcombe, SJ	: 2002	ADVANCES IN EXPERIMENTAL SOCIAL PSYCHOLOGY, VOL 34 Book Series: ADVANCES IN EXPERIMENTAL SOCIAL PSYCHOLOGY Volume: 34 Pages: 131-197
130	Decision constraints and consideration-set formation in consumer durables	Punj, G; Brookes, R	: AUG 2001	PSYCHOLOGY & MARKETING Volume: 18 Issue: 8 Pages: 843-863
131	The construction of mental representations during reading	Vellutino, FR	: SUM 2000	AMERICAN JOURNAL OF PSYCHOLOGY Volume: 113 Issue: 2 Pages: 306-318
132	Given-new versus new-given? An analysis of reading times for spatial descriptions	Baguley, T; Payne, SJ	: 2000	SPATIAL COGNITION: FOUNDATIONS AND APPLICATIONS Book Series: ADVANCES IN CONSCIOUSNESS RESEARCH Volume: 26 Pages: 317-328
133	Cognitive systems engineering: New wine in new bottles	Hollnagel, E; Woods, DD	: AUG 1999	INTERNATIONAL JOURNAL OF HUMAN-COMPUTER STUDIES Volume: 51 Issue: 2 Pages: 339-356
134	Teams embedded in organizations - Some implications	Ilgen, DR	: FEB 1999	AMERICAN PSYCHOLOGIST Volume: 54 Issue: 2 Pages: 129-139
135	The human operator in complex networked systems	Tschiersch, I; Brandt, D	: 1998	(SAFEPROCESS'97): FAULT DETECTION, SUPERVISION AND SAFETY FOR TECHNICAL PROCESSES 1997, VOLS 1-3 Pages: 1261-1265
136	Structure of operators' mental models in coping with anomalies occurring in nuclear power plants	Takano, K; Sasou, K; Yoshimura, S	: DEC 1997	INTERNATIONAL JOURNAL OF HUMAN-COMPUTER STUDIES Volume: 47 Issue: 6 Pages: 767-789
137	An empirical study of computer system learning: Comparison of co-discovery and self-discovery methods	Lim, KH; Ward, LM; Benbasat, I	: SEP 1997	INFORMATION SYSTEMS RESEARCH Volume: 8 Issue: 3 Pages: 254-272
138	Illustrations and mental models in text comprehension	Gyselinck, V	: SEP 1996	ANNEE PSYCHOLOGIQUE Volume: 96 Issue: 3 Pages: 495-516
139	A theory of decision support system design for user calibration	Kasper, GM	: JUN 1996	INFORMATION SYSTEMS RESEARCH Volume: 7 Issue: 2 Pages: 215-232
140	Planning, encoding, and overcoming conflict in partial occlusion drawing: A neo-Piagetian model and an experimental analysis	Morra, S; Angi, A; Tomat, L	: APR 1996	JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY Volume: 61 Issue: 3 Pages: 276-301
141	Designing human-machine interfaces to match the user's mental models	FuchsFrohnhofen, P; Hartmann, EA; Brandt, D; et al.	: JAN 1996	CONTROL ENGINEERING PRACTICE Volume: 4 Issue: 1 Pages: 13-18
142	Gradient and visual speaker normalization in the perception of fricatives	Hildebrandt, B; Moratz, R; Rickheit, G; et al.	: 1996	NATURAL LANGUAGE PROCESSING AND SPEECH TECHNOLOGY: RESULTS OF THE 3RD KONVENS CONFERENCE Pages: 2-13
143	Points of view and point of view sequences	Herrmann, T	: 1996	SPRACHE & KOGNITION Volume: 15 Issue: 4 Pages: 159-177
144	Representations by machines + Cognitive psychology, semiotics, and information science	Glaser, WR	: 1996	ZEITSCHRIFT FUR SEMIOTIK Volume: 18 Issue: 2-3 Pages: 169-189
145	HOW DO CHILDREN REPRESENT INTELLIGENT TECHNOLOGY	VANDUUREN, M; SCAIFE, M	: SEP 1995	EUROPEAN JOURNAL OF PSYCHOLOGY OF EDUCATION Volume: 10 Issue: 3 Pages: 289-301
146	2 MODES OF MENTAL REPRESENTATION AND PROBLEM SOLUTION IN SYLLOGISTIC REASONING	FORD, M	: JAN 1995	COGNITION Volume: 54 Issue: 1 Pages: 1-71
147	INTERNAL REPRESENTATION AND POINT-OF-VIEW - THE EFFECT OF ROUTE DIRECTION	HERRMANN, T; BUHL, HM; SCHWEIZER, K	: 1995	ZEITSCHRIFT FUR PSYCHOLOGIE Volume: 203 Issue: 1 Pages: 1-23
148	IMPROVING END-USER PROFICIENCY - EFFECTS OF CONCEPTUAL TRAINING AND NATURE OF INTERACTION	SANTHANAM, R; SEIN, MK	: DEC 1994	INFORMATION SYSTEMS RESEARCH Volume: 5 Issue: 4 Pages: 378-399
149	SEMANTIC ANALYSIS AND IN-DEPTH UNDERSTANDING OF TECHNICAL TEXTS	CAVAZZA, M; ZWEIGENBAUM, P	: JUL-SEP 1994	APPLIED ARTIFICIAL INTELLIGENCE Volume: 8 Issue: 3 Pages: 425-453
150	MENTAL MODELS OF THE DAY-NIGHT CYCLE	VOSNIADOU, S; BREWER, WF	: JAN-MAR 1994	COGNITIVE SCIENCE Volume: 18 Issue: 1 Pages: 123-183
151	THE USE OF SUPERFICIAL AND MEANING-BASED REPRESENTATIONS IN INTERPRETING PRONOUNS - EVIDENCE FROM SPANISH	CARREIRAS, M; GARNHAM, A; OAKHILL, J	: MAR 1993	EUROPEAN JOURNAL OF COGNITIVE PSYCHOLOGY Volume: 5 Issue: 1 Pages: 93-116
152	WHAT ARE CONCEPTS - ISSUES OF REPRESENTATION AND ONTOLOGY	BREWER, WF	: 1993	CATEGORIZATION BY HUMANS AND MACHINES Book Series: PSYCHOLOGY OF LEARNING AND MOTIVATION-ADVANCES IN RESEARCH AND THEORY Volume: 29 Pages: 495-533
153	PROCESS HIGHLIGHTERS IN A COMPUTER-SIMULATION - FACILITATION OF THEORY-ORIENTED PROBLEM-SOLVING	CHANDLER, TN; CHAILLE, C	: 1993	JOURNAL OF EDUCATIONAL COMPUTING RESEARCH Volume: 9 Issue: 2 Pages: 237-263
154	RECENT VIEWS OF CONCEPTUAL STRUCTURE	KOMATSU, LK	: NOV 1992	PSYCHOLOGICAL BULLETIN Volume: 112 Issue: 3 Pages: 500-526
155	LEVELS OF REPRESENTATION AND DOMAIN-SPECIFIC KNOWLEDGE IN COMPREHENSION OF SCIENTIFIC TEXTS	TARDIEU, H; EHRLICH, MF; GYSELINCK, V	AUG-NOV 1992	LANGUAGE AND COGNITIVE PROCESSES Volume: 7 Issue: 3-4 Pages: 335-351
156	DISCOURSE ORGANIZATION IN THE COMPREHENSION OF TEMPORAL-ORDER IN NARRATIVE TEXTS	OHTSUKA, K; BREWER, WF	: JUL-SEP 1992	DISCOURSE PROCESSES Volume: 15 Issue: 3 Pages: 317-336
157	INFERENCE DURING READING	MCKOON, G; RATCLIFF, R	: JUL 1992	PSYCHOLOGICAL REVIEW Volume: 99 Issue: 3 Pages: 440-466
158	HOW DO YOUNG-CHILDREN LEARN TO BE CONSUMERS - A SCRIPT-PROCESSING APPROACH	PERACCHIO, LA	: MAR 1992	JOURNAL OF CONSUMER RESEARCH Volume: 18 Issue: 4 Pages: 425-440
159	EXPLORING ACADEMIC LITERACY - AN EXPERIMENT IN COMPOSING	GEISLER, C	: FEB 1992	COLLEGE COMPOSITION AND COMMUNICATION Volume: 43 Issue: 1 Pages: 39-54
160	CAUSAL MODEL-BASED KNOWLEDGE ACQUISITION TOOLS - DISCUSSION OF EXPERIMENTS	CHARLET, J; KRIVINE, JP; REYNAUD, C	: 1992	LECTURE NOTES IN ARTIFICIAL INTELLIGENCE Volume: 599 Pages: 318-336
161	HOW TO DECREASE PRAGMATIC INFERENCES - THE IMPACT OF DIFFERENT TECHNIQUES	DARDENNE, B	: OCT 1991	CAHIERS DE PSYCHOLOGIE COGNITIVE-CURRENT PSYCHOLOGY OF COGNITION Volume: 11 Issue: 5 Pages: 555-584
162	CONTROLLING THE PROCESS OF LEARNING FROM AN EXAMPLE THROUGH ADAPTIVE GENERALIZATION OF EPISODIC MEMORY	SAWARAGI, T; IWAI, S; KATAI, O	: JAN 25 1991	FUZZY SETS AND SYSTEMS Volume: 39 Issue: 2 Pages: 133-162
163	CONCEPTUAL MODELS AND USER INTERACTION WITH COMPUTERS	CARDINALE, LA	: 1991	COMPUTERS IN HUMAN BEHAVIOR Volume: 7 Issue: 3 Pages: 163-169
164	THE PROCESSING OF COHESION DEVICES IN TEXT COMPREHENSION	EHRLICH, MF	: 1991	PSYCHOLOGICAL RESEARCH-PSYCHOLOGISCHE FORSCHUNG Volume: 53 Issue: 2 Pages: 169-174
165	ANALOGICAL LEARNING AND TRANSFER IN LANGUAGE-IMPAIRED CHILDREN	KAMHI, AG; GENTRY, B; MAUER, D; et al.	: FEB 1990	JOURNAL OF SPEECH AND HEARING DISORDERS Volume: 55 Issue: 1 Pages: 140-148
166	USE OF ANALOGIES IN PROBLEM-SOLVING	HESSE, FW; KLECHA, D	: 1990	COMPUTERS IN HUMAN BEHAVIOR Volume: 6 Issue: 1 Pages: 115-129
167	EFFECTS OF SEARCH AND SUBJECT EXPERTISE ON INFORMATION SEEKING IN A HYPERTEXT ENVIRONMENT	MARCHIONINI, G; LIN, X; DWIGGINS, S	: 1990	PROCEEDINGS OF THE ASIS ANNUAL MEETING Volume: 27 Pages: 129-142
168	THE INFLUENCE OF TEXTUAL CONTENT, TITLE AND PREVIOUS KNOWLEDGE ON TEXT COMPREHENSION	SPIEL, C	: 1990	ZEITSCHRIFT FUR EXPERIMENTELLE UND ANGEWANDTE PSYCHOLOGIE Volume: 37 Issue: 3 Pages: 505-518
169	CHILDRENS MENTAL ROTATIONS OF ABSTRACT DISPLAYS	BIALYSTOK, E	: FEB 1989	JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY Volume: 47 Issue: 1 Pages: 47-71
170	DEVELOPMENTAL DIFFERENCES IN THE USE OF VERBATIM VERSUS SPATIAL REPRESENTATIONS IN THE RECALL OF SPATIAL DESCRIPTIONS - A PROBABILISTIC MODEL AND AN EXPERIMENTAL-ANALYSIS	MORRA, S	: FEB 1989	JOURNAL OF MEMORY AND LANGUAGE Volume: 28 Issue: 1 Pages: 37-55

#	Title	Authors	Date	Publication
171	COGNITIVE SCIENCE - DEFINITION, STATUS, AND QUESTIONS	HUNT, E	: 1989	JOURNAL OF MEMORY AND LANGUAGE Volume: 28 Issue: 1 Pages: 37-55
172	THE LOGIC OF PLAUSIBLE REASONING - A CORE THEORY	COLLINS, A; MICHALSKI, R	JAN-MAR 1989	ANNUAL REVIEW OF PSYCHOLOGY Volume: 40 Pages: 603-629
173	INFORMATION-SEEKING STRATEGIES OF NOVICES USING A FULL-TEXT ELECTRONIC ENCYCLOPEDIA	MARCHIONINI, G	: JAN 1989	COGNITIVE SCIENCE Volume: 13 Issue: 1 Pages: 1-49
174	The On-line Construction of Discourse Models	Oakhill, Jane; Garnham, Alan; Vonk, Wietske	: 1989	JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE Volume: 40 Issue: 1 Pages: 54-66
175	SYMBOLS, ICONS, AND SIGN PROCESSES	DROSTE, FG	: 1989	LANGUAGE AND COGNITIVE PROCESSES Volume: 4 Issue: 3-4 Special Issue: SI Pages: SI263-SI286
176	COGNITIVE-PSYCHOLOGY AND TEXT-PROCESSING - FROM TEXT REPRESENTATION TO TEXT-WORLD	DENHIERE, G; BAUDET, S	: 1989	SEMIOTICA Volume: 77 Issue: 1-3 Pages: 271-293
177	PRESCHOOL-CHILDREN CAN LEARN TO TRANSFER - LEARNING TO LEARN AND LEARNING FROM EXAMPLE	BROWN, AL; KANE, MJ	: OCT 1988	COGNITIVE PSYCHOLOGY Volume: 20 Issue: 4 Pages: 493-523
178	DECISION PROBLEM STRUCTURING - GENERATING OPTIONS	KELLER, LR; HO, JL	: SEP-OCT 1988	IEEE TRANSACTIONS ON SYSTEMS MAN AND CYBERNETICS Volume: 18 Issue: 5 Pages: 715-728
179	WORKING MEMORY (OR THE M-OPERATOR) AND THE PLANNING OF CHILDRENS DRAWINGS	MORRA, S; MOIZO, C; SCOPESI, A	: AUG 1988	JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY Volume: 46 Issue: 1 Pages: 41-73
180	IMPACT OF PRIOR KNOWLEDGE OF INFORMATIONAL CONTENT AND ORGANIZATION ON LEARNING SEARCH PRINCIPLES IN A DATABASE	LINDE, L; BERGSTROM, M	: APR 1988	CONTEMPORARY EDUCATIONAL PSYCHOLOGY Volume: 13 Issue: 2 Pages: 90-101
181	PROBLEM-SOLVING, RECALL, ISOMORPHIC TRANSFER, AND NONISOMORPHIC TRANSFER AMONG 3RD-GRADE AND 4TH-GRADE CHILDREN	GHOLSON, B; EYMARD, LA; LONG, D; et al.	: JAN 1988	COGNITIVE DEVELOPMENT Volume: 3 Issue: 1 Pages: 37-53
182	PROBLEM-SOLVING, RECALL, AND ISOMORPHIC TRANSFER AMONG 3RD-GRADE AND 6TH-GRADE CHILDREN	GHOLSON, B; EYMARD, LA; MORGAN, D; et al.	: APR 1987	JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY Volume: 43 Issue: 2 Pages: 227-243
183	CONCEPT, KNOWLEDGE, AND THOUGHT	ODEN, GC	: 1987	ANNUAL REVIEW OF PSYCHOLOGY Volume: 38 Pages: 203-227
184	A QUALITATIVE MODEL OF HUMAN INTERACTION WITH COMPLEX DYNAMIC-SYSTEMS	HESS, RA	JAN-FEB 1987	IEEE TRANSACTIONS ON SYSTEMS MAN AND CYBERNETICS Volume: 17 Issue: 1 Pages: 33-51
185	2 WAYS OF COPING WITH INDETERMINATE SPATIAL DISCOURSE	BREDART, S	: 1987	PSYCHOLOGICA BELGICA Volume: 27 Issue: 2 Pages: 143-151
186	THE COMPREHENSION OF ANAPHORIC RELATIONS	STEVENSON, RJ; VITKOVITCH, M	OCT-DEC 1986	LANGUAGE AND SPEECH Volume: 29 Issue: Pages: 335-360 Part: 4
187	DEFAULT REASONING AND SELF-KNOWLEDGE	WILKS, Y	: OCT 1986	PROCEEDINGS OF THE IEEE Volume: 74 Issue: 10 Pages: 1399-1404
188	3 COMPONENTS OF UNDERSTANDING A PROGRAMMERS MANUAL - VERBATIM, PROPOSITIONAL, AND SITUATIONAL REPRESENTATIONS	SCHMALHOFER, F; GLAVANOV, D	: JUN 1986	JOURNAL OF MEMORY AND LANGUAGE Volume: 25 Issue: 3 Pages: 279-294
189	WHAT IS DOMAIN SKILL - FRAMES IN INFORMATION SEEKING	LINDE, LK	: APR 1986	BEHAVIORAL SCIENCE Volume: 31 Issue: 2 Pages: 89-102
190	YOUNG CHILDRENS MENTAL MODELS DETERMINE ANALOGICAL TRANSFER ACROSS PROBLEMS WITH A COMMON GOAL STRUCTURE	BROWN, AL; KANE, MJ; ECHOLS, CH	: APR 1986	COGNITIVE DEVELOPMENT Volume: 1 Issue: 2 Pages: 103-121
191	HOW CAN COGNITIVE-PSYCHOLOGY HELP SOLVE AN ARTIFICIAL-INTELLIGENCE PROBLEM	FARKAS, A	: 1986	COMPUTERS AND ARTIFICIAL INTELLIGENCE Volume: 5 Issue: 4 Pages: 315-320
192	TOO LITTLE AND LATENT	MORTON, J	: 1985	BEHAVIORAL AND BRAIN SCIENCES Volume: 8 Issue: 1 Pages: 26-27
193	A REPRESENTATION SYSTEM OF COGNITIVE PROCEDURES	SAJDA, J	: 1985	COMPUTERS AND ARTIFICIAL INTELLIGENCE Volume: 4 Issue: 5 Pages: 407-419
194	THE CONTENT OF AWARENESS IS A MODEL OF THE WORLD	YATES, J	: 1985	PSYCHOLOGICAL REVIEW Volume: 92 Issue: 2 Pages: 249-284
195	IMPLICIT KNOWLEDGE, QUESTION ANSWERING, AND THE REPRESENTATION OF EXPOSITORY TEXT	GRAESSER, AC; GOODMAN, SM	: 1985	PSYCHOLOGY OF READING AND READING INSTRUCTION Pages: 109-171
196	COGNITIVE-ABILITIES, CONDITIONS OF LEARNING, AND THE EARLY DEVELOPMENT OF READING SKILL	EVANS, MA; CARR, TH	: 1985	READING RESEARCH QUARTERLY Volume: 20 Issue: 3 Pages: 327-350
197	LANGUAGE, THOUGHT, AND MENTAL MODELS	DROSTE, FG	: 1985	SEMIOTICA Volume: 56 Issue: 1-2 Pages: 31-98
198	SEMIOTIC PRINCIPLES AND SEMANTIC THEORIES	DROSTE, FG	: 1985	SEMIOTICA Volume: 54 Issue: 3-4 Pages: 429-458
199	FILM, PERCEPTION AND COGNITIVE PSYCHOLOGY	BROOKS, V	: 1985	MILLENNIUM FILM JOURNAL Issue: 14-1 Pages: 105-126
200	LEXICAL INFORMATION IN SENTENCE COMPREHENSION	TABOSSI, P	: 1985	QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY SECTION A-HUMAN EXPERIMENTAL PSYCHOLOGY Volume: 37 Issue: 1 Pages: 83-94
201	MODELS OF CONCEPTS	COHEN, B; MURPHY, GL	: 1984	COGNITIVE SCIENCE Volume: 8 Issue: 1 Pages: 27-58
202	THE ROLE OF A MENTAL MODEL IN LEARNING TO OPERATE A DEVICE	KIERAS, DE; BOVAIR, S	: 1984	COGNITIVE SCIENCE Volume: 8 Issue: 3 Pages: 255-273
203	EXPLAINING INTERPRETATION, THE CASE OF JAMES,HENRY THE 'TURN OF THE SCREW'	FREUNDLIEB, D	: 1984	POETICS TODAY Volume: 5 Issue: 1 Pages: 79-95
204	SOME EMPIRICAL JUSTIFICATION FOR A THEORY OF NATURAL PROPOSITIONAL LOGIC	BRAINE, MDS; REISER, BJ; RUMAIN, B	: 1984	PSYCHOLOGY OF LEARNING AND MOTIVATION-ADVANCES IN RESEARCH AND THEORY Volume: 18 Pages: 313-371
205	CAUSALITY AND SYNTAX - SOME LINGUISTIC DELINEATIONS IN KANT	DROSTE, FG	: 1984	TIJDSCHRIFT VOOR FILOSOFIE Volume: 46 Issue: 3 Pages: 445-465
206	SCHEMATA IN COGNITIVE ANTHROPOLOGY	CASSON, RW	: 1983	ANNUAL REVIEW OF ANTHROPOLOGY Volume: 12 Pages: 429-462
207	DIGIT MEMORY OF GRAND EXPERTS IN ABACUS-DERIVED MENTAL CALCULATION	HATANO, G; OSAWA, K	: 1983	COGNITION Volume: 15 Issue: 1-3 Pages: 95-110
208	WHATS WRONG WITH STORY GRAMMARS	GARNHAM, A	: 1983	COGNITION Volume: 15 Issue: 1-3 Pages: 145-154
209	BELIEFS ABOUT BELIEFS REPRESENTATION AND CONSTRAINING FUNCTION OF WRONG BELIEFS IN YOUNG CHILDRENS UNDERSTANDING OF DECEPTION	WIMMER, H; PERNER, J	: 1983	COGNITION Volume: 13 Issue: 1 Pages: 103-128
210	ANALOGICAL MODES OF REASONING AND PROCESS MODELING	FUNT, BV	: 1983	COMPUTER Volume: 16 Issue: 10 Pages: 99-104
211	COGNITIVE SYSTEMS-ENGINEERING - NEW WINE IN NEW BOTTLES	HOLLNAGEL, E; WOODS, DD	: 1983	INTERNATIONAL JOURNAL OF MAN-MACHINE STUDIES Volume: 18 Issue: 6 Pages: 583-600
212	OBJECTIVE REPRESENTATION OF VELOCITY IN VERBS OF MOTION	WEBER, G; COLONIUS, H	: 1983	PSYCHOLOGICAL RESEARCH-PSYCHOLOGISCHE FORSCHUNG Volume: 45 Issue: 1 Pages: 73-90
213	ON THE NATURE OF INTELLIGENCE	HUNT, E	: 1983	SCIENCE Volume: 219 Issue: 4581 Pages: 141-146
214	tle: COGNITIVE AUTONOMY OF THE DEVELOPMENT OF MORAL EVALUATION OF ACHIEVEMENT	WIMMER, H; WACHTER, J; PERNER, J	: 1982	CHILD DEVELOPMENT Volume: 53 Issue: 3 Pages: 668-676
215	A TEMPORAL LOGIC FOR REASONING ABOUT PROCESSES AND PLANS	MCDERMOTT, D	: 1982	COGNITIVE SCIENCE Volume: 6 Issue: 2 Pages: 101-155
216	COMPUTATIONAL PROCESS MODELING OF SPATIAL COGNITION AND BEHAVIOR	SMITH, TR; PELLEGRINO, JW; GOLLEDGE, RG	: 1982	GEOGRAPHICAL ANALYSIS Volume: 14 Issue: 4 Pages: 305-325
217	ARTIFICIAL-INTELLIGENCE AND THE STUDY OF LANGUAGE	GOODWIN, JW; HEIN, U	: 1982	JOURNAL OF PRAGMATICS Volume: 6 Issue: 3-4 Pages: 241-280
218	ON BEING A MODEL	MEY, JL	: 1982	JOURNAL OF PRAGMATICS Volume: 6 Issue: 3-4 Pages: 281-292
219	STORIES ARE TO ENTERTAIN - A STRUCTURAL-AFFECT THEORY OF STORIES	BREWER, WF; LICHTENSTEIN, EH	: 1982	JOURNAL OF PRAGMATICS Volume: 6 Issue: 5-6 Pages: 473-486
220	FRAMEWORKS FOR A SCIENCE OF TEXTS - THE COMPLEAT SEMIOTICIAN	GRIMAUD, M	: 1982	SEMIOTICA Volume: 39 Issue: 3-4 Pages: 193-241
221	MENTAL MODELS AS REPRESENTATIONS OF TEXT	GARNHAM, A	: 1981	MEMORY & COGNITION Volume: 9 Issue: 6 Pages: 560-565
222	Planning, encoding, and overcoming conflict in partial occlusion drawing: A neo-Piagetian model and an experimental analysis	Morra, S., Angi, A., Tomat, L.	1996	Journal of Experimental Child Psychology 61 (3), pp. 276-301
223	User instructions for the elderly: What the literature tells us	Van Hees, M.M.W.	1996	Journal of Technical Writing and Communication 26 (4), pp. 521-536
224	Étude expérimentale de l'usage des organisateurs paralinguistiques de mise en retrait dans les documents électroniques | [Using weakening paralinguistic organizers in electronic documents: An experimental approach]	Caro, S., Bisseret, A.	1997	Travail Humain 60 (4), pp. 409-437
225	Structure of operators' mental models in coping with anomalies occurring in nuclear power plants	Takano, K., Sasou, K., Yoshimura, S.	1997	International Journal of Human Computer Studies 47 (6), pp. 767-789
226	10 Cognitive modelling of vision and speech understanding	Hildebrandt, B., Moratz, R., Rickheit, G., Sagerer, G.	1999	Advances in Psychology 128 (C), pp. 213-236
227	Scientific theories and naive theories as forms of mental representation: psychologism revived	Brewer, W.F.	1999	Science and Education 8 (5), pp. 489-505
228	Children's development of a theory of mind	Cadinu, M.R., Kiesner, J.	2000	European Journal of Psychology of Education 15 (2), pp. 93-111
229	L'intégration spatiale d'éléments textuels et illustratifs améliore-t-elle la performance ?	Jamet, E.	2000	Revue d'Intelligence Artificielle 14 (1-2), pp. 167-187
230	Cognitive perspectives on unresolved loss: Insights from the study of PTSD	Fearon, R.M.P., Mansell, W.	2001	Bulletin of the Menninger Clinic 65 (3), pp. 380-396
231	Age changes of semantic composition of associative reactions and effect on them of emotional sign of the verbal stimulus	Pakhomova, A.S.	2003	Journal of Evolutionary Biochemistry and Physiology 39 (1), pp. 110-118
232	Les stratégies de traitement des phrases relatives chez le lecteur débutant | [Processing relative clauses]	Gola-Asmussen, C., Bastien-Toniazzo, M.	2003	Enfance 55 (2), pp. 159-173
233	When scientific models represent	Bailer-Jones, D.M.	2003	International Studies in the Philosophy of Science 17 (1), pp. 59-74
234	On the ontology of digital identification	Hogben, G., Wilikens, M., Vakalis, I.	2003	Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 2889, pp. 579-593
235	Sprachverarbeitung: Theorienkonkurrenz oder gegenstandsflexibilität? | [Language processing: Competing theories or flexibility of the linguistic object?]	Groeben, N.	2005	Zeitschrift fur Psychologie 213 (3), pp. 152-158

236	Cognitive evaluation of human-robot systems: A method for analyzing cognitive change in human-robot systems	Stubbs, K., Bernstein, D., Crowley, K., Nourbakhsh, I.	2006	Proceedings - IEEE International Workshop on Robot and Human Interactive Communication , art. no. 4107786 , pp. 59-65
237	Content-based analysis of modes in design engineering	Saariluoma, P., Nevala, K., Karvinen, M.	2006	Design Computing and Cognition '06 , pp. 325-344
238	From the choice of the patients for a transitive verb to its polysemy	Cordier, F., Pariollaud, F.	2007	Current Psychology Letters: Behaviour, Brain and Cognition 1 (21)
239	La solución de problemas cognitiuos. Una reflexión cognitiva sociocultural	[The solution of cognitive problems. A cognitive and socio-cultural reflection]	Montealegre, R.	2007	Avances en Psicologia Latinoamericana 25 (2) , pp. 20-39
240	Reasoning about knowledge in human-automation systems (preliminary report)	Moses, Y., Shamo, M.K.	2007	Collection of Technical Papers - 2007 AIAA InfoTech at Aerospace Conference 1 , pp. 72-87
241	Nuevos avances en el estudio científico de la comprensión de textos	[New advances in the scientific study of text comprehension]	Marmolejo-Ramos, F.	2007	Universitas Psychologica 6 (2) , pp. 331-343
242	Computational memory architectures for autobiographic agents interacting in a complex virtual environment: A working model	Ho, W.C., Dautenhahn, K., Nehaniv, C.L.	2008	Connection Science 20 (1) , pp. 21-65
243	Closed loop state space analysis and simulation for cognitive systems	Oberheid, H., Gamrad, D., Söffker, D.	2008	Proceedings - International Conference on Application of Concurrency to System Design, ACSD , art. no. 4574593 , pp. 39-44
244	Straight-party voting: What do voters think?	Campbell, B.A., Byrne, M.D.	2009	IEEE Transactions on Information Forensics and Security 4 (4) , art. no. 5256335 , pp. 718-728
245	Affect, behavioural schemas and the proving process	Selden, A., McKee, K., Selden, J.	2010	International Journal of Mathematical Education in Science and Technology 41 (2) , pp. 199-215
246	A social-cognitive framework of multidisciplinary team innovation	Paletz, S.B.F., Schunn, C.D.	2010	Topics in Cognitive Science 2 (1) , pp. 73-95
247	Scenistic methods in training: Definitions and theory grounding	Lyons, P.	2010	Journal of European Industrial Training 34 (5) , pp. 416-431
248	Merging internal and external processes: Examining the mental model convergence process through team communication	Kennedy, D.M., McComb, S.A.	2010	Theoretical Issues in Ergonomics Science 11 (4) , pp. 340-358
249	Naturalistic cognition: A research paradigm for human-centered design	Storkerson, P.	2010	Journal of Research Practice 6 (2)
250	Team creative processes: The importance of complementary and shared mental models	Kennedy, D.M.	2011	Proceedings of the Annual Hawaii International Conference on System Sciences , art. no. 5718448
251	The Influence of Connectives on Young Readers' Processing and Comprehension of Text	Cain, K., Nash, H.M.	2011	Journal of Educational Psychology 103 (2) , pp. 429-441
252	The effects of worked examples on CAD performance: An application of the four-component instructional design model to CAD instruction	Barnes, S., Wiebe, E.N., Branoff, T.J.	2011	ASEE Annual Conference and Exposition, Conference Proceedings
253	Subject-oriented elicitation of distributed business process knowledge	Oppl, S.	2011	Communications in Computer and Information Science 213 CCIS , pp. 16-33
254	Stakeholder-driven collaborative modeling of subject-oriented business processes	Wachholder, D., Oppl, S.	2012	Lecture Notes in Business Information Processing 104 LNBIP , pp. 145-162
255	Leveraging the affordances of Youtube: The role of pedagogical knowledge and mental models of technology functions for lesson planning with technology	Krauskopf, K., Zahn, C., Hesse, F.W.	2012	Computers and Education 58 (4) , pp. 1194-1206
256	Guiding decisions on authorization policies: A participatory approach to decision support	Bartsch, S., Sasse, M.A.	2012	Proceedings of the ACM Symposium on Applied Computing , pp. 1502-1507
257	Thinking Outside the Box? Applying Design Theory to Public Policy	Considine, M.	2012	Politics and Policy 40 (4) , pp. 704-724
258	Persuasive environments: Argumentation models in serious games	Pisan, Y., Tan, C.T.	2012	ICCSE 2012 - Proceedings of 2012 7th International Conference on Computer Science and Education , art. no. 6295308 , pp. 1319-1322
259	An ACT-R approach to reasoning about spatial relations with preferred and alternative mental models	Brüssow, S., Frorath, M., Ragni, M., Fangmeier, T.	2012	Proceedings of the 11th International Conference on Cognitive Modeling, ICCM 2012 , pp. 19-24
260	Multiple logics within argument: How defeasible and classical reasoning work together	Stenning, K.	2012	Frontiers in Artificial Intelligence and Applications 245 (1) , pp. 14-20
261	Diversity of reason: Equivalence relations over description logic explanations	Bail, S., Parsia, B., Sattler, U.	2012	CEUR Workshop Proceedings 846 , pp. 48-58
262	Extracting narrative timelines as temporal dependency structures	Kolomiyets, O., Bethard, S., Moens, M.-F.	2012	50th Annual Meeting of the Association for Computational Linguistics, ACL 2012 - Proceedings of the Conference 1 , pp. 88-97
263	Concept design as knowledge construction - The case of user-centered design	Liikkanen, L.A., Laakso, M.	2012	Proceedings of International Design Conference, DESIGN DS 70 , pp. 1341-1350
264	How a picture facilitates the process of learning from text: Evidence for scaffolding	Eitel, A., Scheiter, K., Schüler, A., Nyström, M., Holmqvist, K.	2013	Learning and Instruction 28 , pp. 48-63
265	Learning and open educational resources: A psychological perspective	Terras, M.M., Ramsay, J., Boyle, E.	2013	E-Learning and Digital Media 10 (2) , pp. 161-173

Abstracts

Below are the unique identifiers of the unique citations, along with an example reference to the relevant Johnson-Laird articles (so as to give an idea of the nature of the citation and the level of critical analysis found within it)

Table 3

Citation	Nature of abstract
1	How the design of presentation slides affects audience comprehension: A case for the assertion-evidence approach
2	Employing insights in visual communication and cognition science, this paper explores the possibilities for visually expressing negation and/or denial.
3	As work is an inherently cooperative phenomenon, it requires a common understanding of the nature of collaboration for all involved parties.
4	This paper explores how negation (e.g., the figure is not red) is understood using the visual world paradigm. Our hypothesis is that people will switch to the alternative affirmative (e.g., a green figure) whenever possible.
5	Not available
6	The present study investigated whether the negated state of affairs was represented when processing negative sentences.
7	Not available
8	The present study investigated whether the negated state of affairs was represented when processing negative sentences.
9	This study investigates why reasoning that involves negation is extremely difficult.
10	Not available
11	Not available
12	We compared learning outcomes in 110 engineering students who viewed a technical presentation in which the slides either integrated or violated six multimedia learning principles.
13	This study investigates the strengths and weaknesses of an alternative approach to design education, in which students were asked to develop a design idea through conceptual diagrams
14	The objective of this case study was to investigate the ability of 10th graders and pre-service teachers to solve logical–mathematical thinking challenges.
15	we present the first experiments on modal scope ambiguities in conditionals, with special attention paid to the scope ambiguities of the probability operator
16	Not available
17	Abstract: A new theory explains how people make hypothetical inferences from a premise consistent with several alternatives to a conclusion consistent with several alternatives.
18	Not available
19	Not available
20	However, limited knowledge is available regarding mental processes in interventional radiology. This research focuses on identifying mental model–related processes
21	We present the first experiments on modal scope ambiguities in conditionals
22	This chapter focuses on the role of team mental models in design and discusses the relevance of team sharedness in the process of designing in teams.
23	The purpose of this study was to (a) investigate the effectiveness of explicit nature of science (NOS) instruction in the context of controversial socioscientific issues and (b) explore whether the transfer of acquired NOS understandings, which were explicitly taught in the context of one socioscientific context, into other similar contexts (familiar and unfamiliar) was possible
24	Not available
25	This book describes the basic threats to the Internet (loss of trust, loss of advertising revenue, loss of security) and how they are related.
26	We propose a new approach to characterizing the timeline of a text: temporal dependency structures,
27	We present an approach to annotating timelines in stories where events are linked together by temporal relations into a temporal dependency tree
28	Abstract: We introduce in this paper a new theory of reasoning, the erotic theory of reasoning, based on the idea that the relationship between questions and answers is central to both our successes and failures of reasoning
29	Abstract: We report the results of an experiment to test the cognitive processes underlying inferences from negated conditionals
30	Abstract: The book Conceptual Structures: Information Processing in Mind and Machine surveyed the state of the art in artificial intelligence and cognitive science in the early 1980s and outlined a cognitive architecture as a foundation for further research and development[...] This paper reviews that architecture and compares it to four other cognitive architectures with their roots in the same era
31	Abstract: The book proves that the author created the logic of natural language representation by linear-tabular chart is most effective to address the priorities of logic and learning of their decision
32	Abstract: This article aims to present the findings of a research which investigated the effect of a difference in word formulation and mathematical characteristics of story problems on their successful solution by preservice mathematics teachers (students) and practising mathematics teachers.
33	Abstract: This paper examines the contemporary philosophical and cognitive relevance of Charles Peirce's diagrammatic logic of existential graphs (EGs)
34	Not available
35	Abstract: This paper deals with the ways of visualizing the process of conflict escalation, settlement, and resolution.
36	Abstract: We explored the influence of negation on cognitive dynamics, measured using mouse-movement trajectories
37	Abstract: In the context of comparing various theoretical accounts of sensorimotor embodiment in language comprehension, we discuss experimental results that demonstrate a bias toward the affirmative in memory
38	Abstract: Truth-table tasks probe for interpretations of if A then C conditionals by having people evaluate both true-antecedent cases (A and C, A and not-C) and false-antecedent cases, (not-A and C, not-A and not-C)[...] A reconsideration of the theory indicates that it accounts for all benchmark phenomena in truth-table tasks and even yields new predictions."
39	Abstract: Since the earliest NLP system, a variey of linguistic, philosophical, logical and computation issues have created controversies

40	Abstract: I will confine the present study in what I take to be the major challenges as well as opportunities this recently resurrected diagrammatic and iconic logical method faces from the points of view of contemporary philosophy of logic, reasoning and cognitive representation.
41	Abstract: The development approach is to empower a machine with a predictive psychological theory of human reasoning (specifically, Philip N. Johnson-Laird's mental models theory) and to use this theory as part of the machine's computational theory of mind, i.e., its ascriptive theory about the mental contents and operations of its audience
42	Abstract: Here, we explore whether constraining this number of possibilities provides more information about processing negation.
43	Abstract: This principle [of truth in mental model theory] is used in the model theory to describe the mechanism in both the correct and fallacies. Other sources of errors in reasoning may be: Too many models, too many relationships between the lements named in the models, too many elements that need to be integrated in a single model. This article will detail the various types of errors arising from these four reasons.
44	Abstract: I have developed a new framework, which is based on the mental models framework in psychology championed by P. N. Johnson-Laird. My new framework models how students search possibility space when thinking about conceptual physics
45	Abstract: The present research program contributes to the theoretical debate that has been installed around the reasoning with conditional statements in the form 'if then q '[...] his evidence supports the hypothesis proposed modulation within the theory of mental models and, together with the results of the tasks , allow advance refinements of the current theory of mental models .
46	This paper reviews some of the challenges and issues associated with providing operators with actionable state awareness through data fusion and argues for the overarching importance of integrating human factors as part of intelligent control systems design and implementation
47	Herein we report the tremendous ability of amphiphilic carbohydrate molecules to form complex three-dimensional architectures
48	Abstract: These findings suggest that standard quantifiers should be replaced with logically clarified quantifiers in teaching and in future research
49	Abstract: Here, we document, for the first time, the pull toward contrasting responses during evaluative thinking, reporting the results of a study examining the trajectory of participants' reaching movements toward different response options
50	Abstract: Humans come prepared to recognize two fundamental features of our surroundings: animate objects and agents[...] Our cultural evolution is knotted to an expanding sense of shared experiences.
51	Abstract: Cognitive systems should be demythologized away from Cartesian detachment, and towards transactions with others and with nature.
52	Not available
53	Abstract: Subject of the article is the investigation of pondering with the group-based dialogic introspection.
54	Abstract: This project argues that student analysts benefit from using articulated heuristics[...] with the communication to be analyzed
55	In this chapter some of the most influential psycholinguistic models of language representation and speech production in multilingual speakers are introduced
56	OBJECTIVE: To determine the health care resource utilization of adult heamophilia patients with and without HIV and HCV infection in a commercially insured population in the United States.
57	Not available
58	Not available
59	Not available
60	Not available
61	Abstract: Two Experiments demonstrate the existence of a "collapse illusion"
62	Abstract: In contrast, according to the mental model theory, a negated conditional is consistent only with the determinate state of affairs, p and not-q.
63	Abstract: I will outline an interpretation of structure preservation and naturalness within a philosophical framework. This leads to the claim that mental models are structures partially isomorphic to what they represent and that they contain exclusively perceptual relations
64	Abstract: his work proposes a rational but domain-specific synthesis aimed at integrating converging lines of research in the WST [Wason selection task] debate
65	Not available
66	Not available
67	Abstract: This study examines the role of "conceptual diagrams" in architectural design, specifically how they facilitate conceptual elaboration in problem-solving processes.
68	Abstract: This paper reviews some of the challenges and issues associated with providing operators with actionable state awareness through data fusion and argues for the overarching importance of integrating human factors as part of intelligent control systems design and implementation
69	Abstract: The judgements of validity by subjects untrained in logic may, for some inferences, systematically differ from those of the weakest logical systems. This paper is about a subset of these inferences
70	Not available
71	Abstract: "This thesis poses three questions: First, what is the nature ohe initial phases of design? Second, what are the roles of conceptual diagrams in the initial phases? Third, how to study the initial phases of design?
72	Not available
73	Abstract: We argue that a single case can never prove a general rule to be true, as philosophy of science has taught any researcher.
74	Three experiments were conducted to study on a more fine-grained level how processing a picture facilitates learning from text.
75	After examining what has been done in science learning with AR supports, several suggestions for future research are proposed.
76	In our work, we begin to investigate how investigating narrative forms of knowing is useful to understanding urbanized, cosmopolitan societies as well.
77	After discussing the conceptual differences between adaptive expertise and related team learning and innovation concepts, we examine longitudinal behavioural data on novelty, routine and adaptive expertise.
78	It has been claimed that older adults (60-90-year-olds) are less likely to encode and remember some kinds of information from texts than young adults. The experiment described here shows that they do make a particular kind of inference to the same extent that college-age adults do.
79	We argue that a similar circuit structure in all cerebellar areas may carry out various operations using a common computational scheme.
80	Not available
81	After a review of shared mental models, we discuss the shared mental models in design.
82	Thirty-one three- to five-year-old children from one public library, one preschool, and home schools were recruited to interpret images of two picture books in short interviews.
83	We hypothesized that conceptual change of a mental model requires change in the system of relations between the features of the prior model. To test this hypothesis, we compared instruction aimed at revising knowledge at the mental model level called holistic confrontation - in which the learner compares and contrasts a diagram of his or her flawed mental model - to an expert model to instruction aimed at revising knowledge at the false belief level - in which the learner is prompted to self-explain the expert model alone.

84	Not available
85	In this paper, we propose a measure of lexical entrainment for multi-party speaking situations.
86	This paper presents an approach for the collaborative modelling of subject-oriented business processes with the aid of an interactive, distributed platform and introduces concepts for information awareness and tool supported development of cooperative work aspects for effectively supporting modellers.
87	We show how simulated robots evolved for the ability to display a context-dependent periodic behaviour can spontaneously develop an internal model and rely on it to fulfil their task when sensory stimulation is temporarily unavailable.
88	In this paper I present a conceptual framework for understanding the effects of scaffolding.
89	In this article, we present an approach for characterizing evidence-the evidence map approach.
90	This manuscript ties together previously unrelated theories of the role of prediction in different aspects of human information processing to create an integrated framework for cognition.
91	The authors investigated young readers' knowledge, processing, and comprehension of temporal, causal, and adversative connectives using offline and online tasks.
92	This paper presents an individual-centered approach to cooperative elicitation of work process knowledge in co-located and distributed business settings.
93	This chapter explores the role and development of mental models in coordination.
94	The present article focuses on our attempt to provide a comprehensive, multi-level procedure for both structural and functional analysis of narrative discourse produced by speakers with brain damage.
95	Not available
96	Not available
97	We propose some necessary requirements for a complete model of inclusive cognitive interaction and establish four criteria for what would constitute a good framework for the purpose of developing a research approach that could be used to construct and test predictive tools for design.
98	In this paper, we discuss borderline examples of (mis)understanding where it is not clear whether or not a misunderstanding has occurred, whether or not communication was successful, and where the participants do not try to negotiate an understanding, even though different interpretations are very likely to exist.
99	This article aims to explore the nature and resolution of breakdowns in coordinated decision making in distributed safety-critical systems.
100	Five studies identify the conditions in which the two processes occur and outline the various mechanisms that might underlie these effects.
101	Not available
102	In this paper, we explored the use of low fidelity Synthetic Environments (SE; i.e., a combination of simulation techniques) for product design.
103	The authors replicated and extended previous research investigating cognitive processes involved in the refutation text effect.
104	Not available
105	In this chapter we explore the notion of iteration and distinguish two domains of iteration: iterations associated with cognitive processes that take place during design and iterations over representational artefacts about the design.
106	The main aim of this study was to describe the impairment of micro and macrolinguistic abilities in a group of twenty-nine schizophrenic patients during the phase of illness stability compared to forty-eight healthy participants matched for age, gender and educational level.
107	Through the Network for Computational Nanotechnology Web site, nanoHUB.org, tens of thousands of users from 172 countries collaborate, share resources, and solve real nanotechnology problems. The authors share their experiences in developing and using the site's unique cyberinfrastructure.
108	This article explores how engineering education can support acquisition of a wide range of knowledge and skills associated with comprehending and using STEM knowledge to accomplish real world problem solving through design, troubleshooting, and analysis activities.
109	This article argues that conducting experiments involving the ability to control and even manipulate the cognitive load in working memory (WM; storage and or processing load) should make it possible to identify the processes involved during the integration of information coming from multiple sources.
110	An enduring problem confronting design science is the question of how to distil design principles and propositions in contexts where only limited evidence has accrued directly in connection with the design problem at hand. This article illustrates how researchers can address this challenge by recourse to well-established bodies of basic theory and research in the wider social and organizational sciences that suggest robust design options.
111	Not available
112	Behavioural variables were entered into a principal component analysis.
113	The purpose of this chapter is to examine the process of mental model construction while reading science text.
114	This chapter is an introduction to current and contemporary work on mental representations. In particular, we emphasize theoretical and empirical views that have focused on links between perception and action, and what those links imply for learning.
115	The paper presents a novel technique for the integrated application of simulation and state space analysis within a single automated run of a Coloured Petri Net (CPN) model in CPN Tools.
116	The paper argues that the practice of thought experimenting enables scientists to follow through the implications of a way of representing nature by simulating an exemplary or representative situation that is feasible within that representation.
117	Mental models are described and discussed as a means of communicating the underlying aspects of instructional leadership.
118	Eleven patients with right hemisphere damage (RHD), 11 left hemisphere damaged (LHD) nonaphasic subjects, and 11 neurologically intact controls were given three story description tasks.
119	This introductory chapter delineates the basic premises underlying the inquiry that is currently carried out in a variety of fields on writing as sociocultural practice.
120	In this contribution, competing theories within the history of the psychology of language processing are discussed and analyzed. It is argued that it is, however, merely a pseudo-competition.
121	his paper proposes an easy and natural way to create ontologies through concept mapping.
122	The article deals with research on framing effects.
123	The OECD Halden Reactor Project performed two closely related experiments, which allowed assessments of whether the quality of human-automation cooperation would be promoted by a human-machine interface designed to increase the observability of the automatic system's activity using graphical and verbal feedback, as compared to a conventional human-machine interface.
124	In this paper I examine a cognitive mechanism of incommensurability.
125	This paper offers a solution to a problem in Herschel studies by drawing on the dynamic frame model for concept representation offered by cognitive psychology.

126	We propose that the manner in which a purchase decision is initiated has an important influence on subsequent product evaluation and choice.
127	The dual-task paradigm was used to show how visuospatial working memory and the phonological loop are involved in processing scientific texts and illustrations presented via computer.
128	Four experiments investigated the conditions in which people use a prototypic event sequence to comprehend a situation-specific sequence of events.
129	Not available
130	The central argument of this article is that the initial phases of the consumer decision process have a major effect on product evaluation and choice.
131	Not available
132	Not available
133	This paper presents an approach to the description and analysis of complex Man-Machine Systems (MMSs) called Cognitive Systems Engineering (CSE).
134	It is argued here that the organizational perspective has produced research that complements and extends past research on groups and teams,
135	In this report, the concept of mental models of human operators is proposed. It is to be used in software design in order improve safety by continuing human control of complex, networked systems in case of chaotic behaviour.
136	This paper presents the structure and contents of a "operator's mental model" and it can deal with flexible operations under any situation.
137	This paper reports a study that examined two types of exploratory computer learning methods: self-discovery vs. co-discovery, the latter of which involves two users working together to learn a system.
138	The purpose of this paper is to show that mental model theory can provide an account for the beneficial effect of illustrations in text comprehension.
139	A theory is proposed for designing decision support systems (DSS) so that the confidence a decision maker has in a decision made using the aid equals the quality of that decision.
140	A theoretical model of partial occlusion drawing is presented, along with three experiments.
141	A theory of mental models (structures of analogue knowledge) is described. A methodology incorporating this theory has been used to analyse the user's mental models in work settings.
142	We present a cognitive model for the understanding of speech and vision sensory data.
143	This is a special issue that is concerned with considerations and results that relate to the fact that spatial cognition and spatial communication involve points of view. This contribution serves as an introduction and as a synopsis.
144	Not available
145	Stories were analysed in terms of a number of general features
146	In this paper, the theory of syllogistic reasoning proposed by Johnson-Laird (1983, 1986; Johnson-Laird and Bara, 1984; Johnson-Laird and Byrne, 1991) is shown to be inadequate and an alternative theory is put forward.
147	The issue of our paper is whether the way humans experience a specific spatial layout (= route of experience) becomes a part of their spatial knowledge representation.
148	In this paper, we draw upon assimilation theory of learning to propose and test a two-stage model of mental model development.
149	We explored the problem of achieving in-depth understanding of natural language sentences from narrative technical reports through knowledge-based free text understanding.
150	This article presents the results of an experiment which investigated elementary school children's explanations of the day/night cycle.
151	In this paper, we address the question of whether a representation of the surface form of a text is directly implicated in the interpretation of definite pronouns in that text.
152	Not available
153	This experiment investigates whether or not process highlighters, incorporated in a computer simulation, facilitate the knowledge representation of early adolescent children.
154	This article reviews theories of concept structure proposed since the mid- 1970s, when the discovery of typicality effects led to the rejection of the view that instances of a concept share necessary and sufficient attributes.
155	Two levels of representation in text comprehension were postulated: a propositional representation and a mental (or situation) model of the content of the text.
156	We investigated the role of global discourse organization on the comprehension of temporal order in narrative.
157	In this article, an alternative view is proposed, labelled the minimalist hypothesis. According to this hypothesis, the only inferences that are encoded automatically during reading are those that are based on easily available information, either from explicit statements in the text or from general knowledge, and those that are required to make statements in the text locally coherent.
158	The experiments reported in this article examine young children's knowledge-acquisition abilities, relative to older children, with the goal of identifying how young children learn and then using this understanding to design situations in which age differences in learning are eliminated.
159	Not available
160	The aim of this paper is to study causal knowledge and demonstrate how it can be used to support the knowledge acquisition process.
161	Not available
162	This paper presents a methodology to acquire general knowledge from a single example (episode) by the use of a repertoire of prior cases based on a hybrid architecture of EBL and SBL.
163	Not available
164	Not available
165	Not available
166	Not available
167	Not available
168	Not available
169	Not available
170	Not available
171	Not available
172	Not available
173	Not available
174	This paper discusses the on-line interpretation of texts within the framework of the mental models theory of discourse comprehension.
175	Not available
176	Not available

177	Not available
178	Not available
179	Not available
180	Not available
181	Not available
182	Not available
183	Not available
184	Not available
185	Not available
186	Not available
187	Not available
188	Not available
189	Not available
190	Not available
191	Not available
192	Not available
193	Not available
194	Not available
195	Not available
196	Not available
197	Not available
198	Not available
199	Not available
200	Not available
201	Not available
202	Not available
203	Not available
204	Not available
205	Not available
206	Not available
207	Not available
208	Not available
209	Not available
210	Not available
211	Not available
212	Not available
213	Not available
214	Not available
215	Not available
216	Not available
217	Not available
218	Not available
219	Not available
220	Not available
221	Not available
222	A theoretical model of partial occlusion drawing is presented, along with three experiments
223	The study described in this article is aimed at an investigation of this gerontological research literature to find out what is already known about age deficiencies in cognitive processes which might adversely influence instructional text processing.
224	Not available
225	This paper presents the structure and contents of a "operator's mental model" and it can deal with flexible operations under any situation.
226	With this contribution to the discussion about mental models we present a cognitive model for the understanding of speech and vision sensor data. The main point of this article is to explain the interaction between visual perception and language.
227	This paper analyzes recent work in psychology on the nature of the representation of complex forms of knowledge with the goal of understanding how theories are represented. The analysis suggests that, as a psychological form of representation, theories are mental structures that include theoretical entities (usually nonobservable), relationships among the theoretical entities, and relationships of the theoretical entities to the phenomena of some domain.
228	A critical review of the literature on the theory of mind is presented. Consistent with the "early onset" view, it is suggested that important precursors of a theory of mind are found much earlier than the age of 4.

229	There is a growing body of evidence that effective formats of instructional presentation can facilitate learning.
230	They propose that unresolved loss involves intrusion and avoidance phenomena similar to those of PTSD.
231	It was shown that the main role in formation of the verbal consciousness in the 4-8-year old children was played by the right hemispheric mechanisms of cognition
232	The main purpose of this research is to specify which strategies allow reading beginners to understand sentences with one relative clause.
233	Scientific models represent aspects of the empirical world. I explore to what extent this representational relationship, given the specific properties of models, can be analysed in terms of propositions to which truth or falsity can be attributed.
234	We examine the ontology of digital identification in the wider context of privacy.
235	In this contribution, competing theories within the history of the psychology of language processing are discussed and analyzed. It is argued that it is, however, merely a pseudo-competition.
236	To help answer questions about the behaviour of participants in human-robot systems, we propose the Cognitive Evaluation of Human-Robot Systems (CEHRS) method based on our work with the Personal Exploration Rover (PER)
237	In this paper, we discuss about content-based design research. By means of it, we have separated four different types of thought processes, which occur during engineering design.
238	We hypothesized that a process verb refers to the representation of a situation. We examined this hypothesis within the framework of verb polysemy for 12 French verbs.
239	This work was done based on the Historical-Cultural Psychology perspective. It emphasizes the Problem Solving as a High Psychological Function linked with other psychological functions such us: language, thinking, reasoning etc. It presents ociocultural reflection on Cognitive Problem Solving
240	In a supervisory control system the human agent's knowledge of past, current, and future system behaviour is critical for system performance.
241	This paper reviews some of the classical and, particularly, current works in cognitive science and new advances in the neuroscience of text comprehension.
242	In this paper, we discuss the concept of autobiographic agent and how memory may extend an agent's temporal horizon and increase its adaptability.
243	The paper presents a novel technique for the integrated application of simulation and state space analysis within a single automated run of a Coloured Petri Net (CPN) model in CPN Tools.
244	We presented survey respondents with multiple SPV scenarios, testing both their understanding of SPV-marked ballots and their likelihood of using SPV to make multiple selections on a ballot.
245	In this largely theoretical article, we discuss the relation between a kind of affect, behavioural schemas and aspects of the proving process.
246	We present a new framework of team innovation in multidisciplinary science and engineering groups that ties factors from both literatures together.
247	The aim of this article is to describe the scenistic approach to training with corresponding activities and the theory bases that support the approach.
248	We present a framework for examining mental model convergence via communicated mental model content. Our framework is theoretically based on research in cognitive science and communication, as well as more recent team mental model research.
249	This article makes a case for the explicit, formal study of implicit, naturalistic thinking within the fields of design. It develops a framework for defining and studying naturalistic thinking and knowledge
250	Herein, two types of mental models, complementary and shared mental models, are related to team creative processes.
251	The authors investigated young readers' knowledge, processing, and comprehension of temporal, causal, and adversative connectives using offline and online tasks.
252	This presentation discusses enhancements to current instructional practices for engineering graphics and constraint-based modelling courses taught at the collegiate level, and introduces a novel application of an instructional design framework,
253	This paper presents an individual-centered approach to cooperative elicitation of work process knowledge in co-located and distributed business settings.
254	We compared instruction aimed at revising knowledge at the mental model level called holistic confrontation - in which the learner compares and contrasts a diagram of his or her flawed mental model to an expert model - to instruction aimed at revising knowledge at the false belief level - in which the learner is prompted to self-explain the expert model alone.
255	This paper presents an approach for the collaborative modelling of subject-oriented business processes with the aid of an interactive, distributed platform and introduces concepts for information awareness and tool supported development of cooperative work aspects for effectively supporting modellers.
256	Web-based digital video tools enable learners to access video sources in constructive ways. To leverage these affordances teachers need to integrate their knowledge of a technology with their professional knowledge about teaching. We suggest that this is a cognitive process, which is strongly connected to a teacher's mental model of the tool's affordances.
257	In this paper, we present a participatory approach to collecting this information, which not only reduces cost, but increases effectiveness because it ensures that specific local knowledge and downstream risks are represented and visible to decision-makers.
258	By linking some of the key attributes of a design model of decision making to an account of expertise, it is possible to formulate a stronger model of public policy design expertise.
259	The aim of this project is to extend and adapt current computational models of argumentation to support the range of user actions and diverse scenarios in serious games; test and evaluate the effectiveness of the argumentation model in changing people's behaviour and decision through gameplay; and develop tools for game designers to explicitly model player reasoning enabling richer and more adaptive interactive environments.
260	Not available
261	In the present paper, we introduce and explore several equivalence relations over justifications for entailments of OWL ontologies.
262	We compare two parsing models for temporal dependency structures, and show that a deterministic non-projective dependency parser outperforms a graph-based maximum spanning tree parser, achieving labelled attachment accuracy of 0.647 and labelled tree edit distance of 0.596.
263	Not available
264	Three experiments were conducted to study on a more fine-grained level how processing a picture facilitates learning from text.
265	Not available

References

Below are the unique identifiers of the unique citations, along with an example reference to the relevant Johnson-Laird articles (so as to give an idea of the nature of the citation and the level of critical analysis found within it)

Table 4

Citation	Reference
1	Not available
2	In the examples (1), taken from Khemlani et al. (2012), (1a) exemplifies broad scope or sentential negation, whereas (1b) exemplifies narrow scope or VP- negation
3	Not available
4	According to the model theory, individuals understand negation by simulating either the alternative affirmative or the negated argument while applying a symbol that represents negation (Khemlani, Orenes, & Johnson-Laird, 2012)
5	Not available
6	Not available
7	Not available
8	Not available
9	Khemlani, Orenes, and Johnson-Laird (2012a, 2012b) presented an extension of the Mental Models Theory (MMT) of human thought (Byrne & Johnson-Laird, 2009; Johnson-Laird, 1983, 2008) to account for the specific case of negation in sentences. The aim of this study was to test two hypotheses derived from such an account
10	Not available
11	Not available
13	Not available
13	Not available
14	Not available
15	Not available
16	Abstract: We argue that a single case can never prove a general rule to be true, as philosophy of science has taught any researcher.
17	"The inference made by the heuristic is one of possibility rather than necessity (Barres & Johnson-Laird, 2003;[)]", p104;
18	Not available
19	Not available
20	There is a debate about the similarities and differences between mental models and visual images
21	Not available
22	An interest in mental models was reignited with the pioneers works of... Johnson-Laird
23	Multiple diverse experiences targeting a similar theme would promote the construction of a generalized mental model
24	Not available
25	Not available
26	The annotators were instructed to link each event in the story to a single nearby event, similar to what has been observed in reading comprehension studies
27	We ask the annotators to link each event in the story to a single nearby event, similar to what has been observed in reading comprehension studies
28	Ref: "For example, people find it difficult to list all the possibilities compatible with the negation of a conjunction (Khemlani et al, 2012)", p26
29	Not available
30	Ref: Peirce called EGs his "chef d'oeuvre" and claimed that the rules of inference for EGs represent "a moving picture of the mind in thought." After a detailed comparison of Peirce's EGs to current theories about mental models, the psychologist Johnson-Laird (2002) agreed
31	Not available
32	Not available
33	Ref: It would nevertheless be a gross error to mistake the theories of EGs for the class of theories such as mental models (Lakoff & Turner 1989, Johnson Laird 2002
34	Not available
35	Not available
36	Two theoretical perspectives on negation have produced their own empirical literatures investigating it. The first may be termed the ''propositional'' account of negation, exemplified in studies of thinking and reasoning that explicitly take into account the discrete influences negation has on processing, such as the construction of mental models (e.g., Barres & Johnson-Laird, 2002), p984
37	Psycholinguists have examined the comprehension of negated statements. One of the main findings is that people are slower to read negated sentences than they are to read affirmative sentences because they are more complex (see Barres & Johnson-Laird 2003), pp3-4

38	Not available
39	Ref: copy of Cognitive Architectures for Conceptual Structures
40	Ref: Mental models, model-theoretic semantics, and the psychosemantic conception of truth
41	Not available
42	Ref: One of the earliest and most reliable findings about negation is that people are slower to read negated sentences than they are to read affirmative sentences, due to their increased complexity (see Barres & Johnson, 2003;[]), p1192
43	Ref: Results of experiments conducted by Patricia Barres and Johnson-Laird (2003) confirm that in order to realize what would happen if any statement was false, subjects first form a model, assuming that the statement is true.", p73
44	Ref: "the process of deduction that students use when solving physics problems has not been thoroughly stud- ied. In an effort to better understand that reasoning process, I have developed a new framework, which is based on the mental models framework in psychology championed by P. N. Johnson-Laird." p1
45	Ref: The inspection of this model identifies relationships between elements that were not listed in assumptions (see for a similar argument, Johnson-Laird, 2002; Peirce, 1931-1958):
46	Ref: For a review of the history of mental models see Johnson-Laird [13-15].
47	Not available
48	Finally, consistent with other findings (Barres & Johnson-Laird, 2003), we observed that reasoners had difficulty working out the implications of false premises", p225
49	Ref:"Our results are consistent with previous claims regarding proposition verification (Barres & Johnson-Laird, 2003)2
50	Not available
51	Not available
52	Not available
53	Dass Dispositions disjunktionen auf aussagenlogische Wahrheitstabellen umstandlos projiziert werden, kommt auch vor (s. das Beispiel von Barres und Johnson-Laird, 2003,[])" (disjunctive disopossessions that are projected on propositional truth tables, also occurs (see the example of Barres and Johnson-Laird, 2003)), p98
54	Ref: "research in logic studies and mental models generally advocate for diagrams as a special kind of problem representation that eases solution
55	Not available
56	Not available
57	Not available
58	Not available
59	Not available
60	Not available
61	Not available
62	Ref: "Since there is nothing resembling negation, there cannot be iconic representations of negation[...] There are several difficulties with that view" p270
63	Not available
64	The idea to combine a Bayesian approach and different structural models may not only be relevant for the WST debate, but may also shed light on related research on the truth table tasks, or on a-, b-, c- or d-cell biases and different strategies in the evaluation of contingency tables (see e.g. [...] Barres & Johnson-Laird, 2003)", p150
65	Not available
66	Not available
67	Ref: "Johnson-Laird also points to an important similarity between mental models and diagrams in that they both maintain structural correspondences with their represented domains. We propose that some diagrams can provide external representations of structural correspondences of a mental model of a core design conceptualization.
68	Ref: [see Johnson-Laird 2002] for a recent discussion of Pierce as a precursor of mental model theory in reasoning
69	Ref: Barres & Johnson-Laird (2003) point out an unexpected phenomenon", p8
70	Not available
71	Abstract: "This thesis poses three questions: First, what is the nature of the initial phases of design? Second, what are the roles of conceptual diagrams in the initial phases? Third, how to study the initial phases of design?
72	Not available
73	The verification task also suffers from the conceptual impossibility of making a rule true[...] Yet other truth-table task studies do not ask people to reason about the truth of conditionals. They ask people to reason from conditionals—that is, they ask people assume the conditionals are true and to evaluate (or construct: Barres & Johnson-Laird, 2003) the truth contingencies as possible or impossible given the truth of the conditionals[...] These studies do not suffer from the induction problem either.
74	Not available
75	Not available
76	Not available
77	Not available
78	Not available
79	Not available
80	Not available

81	Not available
82	Not available
83	Not available
84	This suggestion is sometimes implicit in logical approaches to language since Boole[...] and bears some resemblance to the psycholinguistic notion that reasoning follows and builds upon interpretation [johnson-Laird 1980[]]
85	Not available
86	Not available
87	Not available
88	Such schemas contain mental representations of categories of concepts and general processes by which individuals make sense of the world and perform actions (Johnson-Laird 1980).
89	Not available
90	Not available
91	Not available
92	Not available
93	Not available
94	Not available
95	Other terms for schemata include scripts, mental models.
96	Not available
97	Not available
98	Methodologically, we will follow a theory (van Dijk and Kintsch, 1983) that builds on mental models, a concept from cognitive science (e.g., Johnson-Laird, 1980).
99	Not available
100	Not available
101	Not available
102	Not available
103	Not available
104	Once in place, these cognitive structures are used to reason about the world and eventually motivate action/s in it (Johnson-Laird 1980[)]
105	Not available
106	Not available
107	Not available
108	Not available
109	Not available
110	Not available
111	Not available
112	Not available
113	Not available
114	Not available
115	Not available
116	Not available
117	Not available
118	Not available
119	Not available
120	Not available
121	Not available
122	Not available
123	Not available
124	Not available
125	Not available
126	Not available
127	Not available
128	On one hand, they may simply store separate representations of their personal experiences in memory as event exemplars, or situation models (Johnson- Laird
129	Not available
130	Not available
131	Not available
132	Not available

133	Not available
134	Not available
135	Not available
136	Not available
137	Not available
138	Not available
139	Not available
140	Not available
141	Not available
142	The theory of mental models assumes that they can be constructed on the basis of either verbal or perceptual information
143	Not available
144	Not available
145	Not available
146	In this paper, the theory of syllogistic reasoning proposed by Johnson-Laird (1983, 1986; Johnson-Laird & Bara, 1984; Johnson-Laird & Byrne, 1991) is shown to be inadequate and an alternative theory is put forward.
147	Not available
148	Not available
149	Not available
150	Not available
151	Not available
152	Not available
153	Not available
154	Not available
155	Not available
156	Not available
157	Not available
158	Not available
159	Not available
160	Not available
161	Not available
162	Not available
163	Not available
164	Not available
165	Not available
166	Not available
167	Not available
168	Not available
169	Not available
170	Not available
171	Not available
172	Not available
173	Not available
174	Not available
175	Not available
176	Not available
177	Not available
178	Not available
179	Not available
180	Not available
181	Not available
182	Not available
183	Not available
184	Not available

185	Not available
186	Not available
187	Not available
188	Not available
189	Not available
190	Not available
191	Many excellent sources are available on the topics raised in this chapter[...] (Johnson-Laird[]).
192	Not available
193	Not available
194	Not available
195	Not available
196	Not available
197	Not available
198	Not available
199	Not available
200	Not available
201	Both classical and fuzzy set theory require that large or infinite sets of objects be examined in order to evaluate quantificational statements, thus making these theories of questionable psychological interest (see Johnson-Laird, 1980).
202	Not available
203	Not available
204	Not available
205	Not available
206	Not available
207	Not available
208	Not available
209	Not available
210	Not available
211	Not available
212	Not available
213	Not available
214	Not available
215	Not available
216	Not available
217	We cannot, however, share Johnson-Laird's conclusion
218	Not available
219	Not available
220	Not available
221	Not available
222	Not available
223	Not available
224	Not available
225	Not available
226	Not available
227	Not available
228	In other words, these children would not be able to construct two different models of the world (see Johnson-Laird, 1980)
229	Not available
230	Not available
231	Not available
232	Not available
233	Not available
234	Not available
235	Not available
236	Not available

237	Not available
238	Not available
239	Not available
240	Not available
241	Therefore, states that infer is to understand and be understood is to build a model of the real or fictional situation to which the text refers, rather than the text itself (Johnson- Laird , 1980)
242	Not available
243	Not available
244	Not available
245	Not available
246	Not available
247	Not available
248	Not available
249	The logic of the proposition leads to the conclusion that the fifth is on the first's right, but what if they are in a circle? (Johnson-Laird, 1980, p.87). In short, mental models encode unrecognized assumptions (people in a line), so they break in ways that nnot be systematically predicted and may escape detection.
250	Not available
251	Not available
252	Not available
253	Not available
254	Not available
255	Not available
256	Not available
257	Not available
258	Not available
259	Not available
260	Not available
261	Not available
262	Not available
263	Not available
264	Not available
265	Not available

Bibliography

Abramson, D., 2011, "Philosophy of mind is (in part) philosophy of computer science" in *Minds & Machines*, No. 21, pp203-219.

Agarwal, A.; Lang, J.H., 2005, *Foundations of analog and digital electronic circuits*, New York: Morgan Kaufman Publishers.

Anscombe, G.E.M., 2001, *An introduction to Wittgenstein's Tractatus: Themes in the philosophy of Wittgenstein*, London: St. Augustine's Press.

Aristotle, 2006a, "On interpretation" in *Categories, on interpretation and on sophistical refutation*, translated from Latin by E.M. Edghill and W.A. Pickard-Cambridge, Stilwell: Digireads Publishing, pp27-43.

Aristotle, 2006b, *On the soul*, translated from Latin by J.A. Smith, Stilwell: Digireads Publishing.

Audi, P., 2012, "A clarification and defense of the notion of grounding" in in F. Correia and B. Schnieder (eds.) *Metaphysical grounding: Understanding the structure of reality*, Cambridge: Cambridge University Press, pp101-121.

Barres, P.; Johnson-Laird, P.N., 2003, "On imagining what is true (and what is false)" in *Thinking and Reasoning*, Vol. 9, No. 1, pp1-42.

Baggio, G.; van Lambalgen, M.; Hagoort, P., 2012, "Language, linguistics and cognition" in R. Kempson, N. Asher and T. Fernando (eds.) *Handbook of the philosophy of science: Philosophy of linguistics*, Oxford: Elsevier, pp325-355.

Baggini, J.; Fosl, P., 2010, *The philosopher's toolkit: a compendium of philosophical concepts and methods*, Oxford: Wiley-Blackwell.

Baumgartner, M., 2010, "Causal slingshots" in *Erkenntnis*, Vol. 72, pp111-133

Baz, A., 2012, *When words are called for: A defense of ordinary language philosophy*, London: Harvard University Press.

Bechtel, W.; Abrahamsen, A.; Graham, G., 2001, "Cognitive science: History" in *International Encyclopedia of the Social and Behavioral Sciences*, New York: Elsevier, pp2154-2158.

Beebe, J.R., 2004, "The Generality Problem, Statistical Relevance and the Tri-Level Hypothesis" in *Noûs*, Vol. 38, No. 1, pp177-195.

Belletti, A.; Rizzi, L., 2002, "Editors' introduction: Some concepts and issues in linguistic theory" in N. Chomsy *On nature and language*, Cambridge: Cambridge University Press, pp1-44.

Bennett, M.R.; Hacker, P.M.S., 2003, *Philosophical foundations of neuroscience*, Oxford: Blackwell.

Bermúdez, J.L., 2010, *Cognitive science: An introduction to the science of mind*, Cambridge; Cambridge University Press.

Block, N., 1978, "Troubles with functionalism", in C. Wade Savage (ed.) *Minnesota studies in the philosophy of science*, Vol. 9, Minneapolis: University of Minnesota Press, pp268-305.

Block, N., 2002, "Twenty-one years in the Chinese room" in J. Preston and M. Bishop (eds.) *Views into the Chinese Room: New essays on Searle and artificial intelligence*, Oxford: Oxford University Press, pp70-79.

Bobrow, D.G.; Collins, A., 1975, *Representation and understanding: Studies in cognitive science*, Michigan: Elsevier.

Bowerman, M., 1977, "The acquisition of word meaning: An investigation of some current concepts" in P. Johnson-Laird and P.C. Wason (eds.) *Thinking: Readings in cognitive science*, Cambridge: Cambridge University Press, pp234-243.

Bradley, F.H., 1916, *Appearance and reality: A metaphysical essay*, London: Allen & Unwin Ltd.

Brentano, F., 2009, *Psychology from an empirical standpoint*, translated by A.C. Rancurello, D.B. Terrell & L.L. Mcalister, London: Routledge.

Brook, A., 2009, "Introduction: Philosophy in and philosophy of cognitive science" in *Topics in Cognitive Science*, Vol. 1, No. 2, pp216-230.

Brooks, R.A., 1990, "Elephants don't play chess" in *Robotics and automated systems*, Vol. 6, pp3-15.

Burks, A.W., 1949, "Icon, Index, and Symbol" in *Philosophy and Phenomenological Research*, Vol. 9, No. 4, pp. 673-689.

Byrge, L.; Sporns, O.; Smith, L.B., 2014, "Developmental process emerges from extended brain–body–behavior networks" in *Trends in Cognitive Sciences*, Vol. 18, No.8, pp395-403.

Cadinu, M.R., Kiesner, J., 2000, "Children's development of a theory of mind" in *European Journal of Psychology of Education*, Vol. 15, No. 2, pp93-111.

Carnap, R., 1959, "The Elimination of Metaphysics through Logical Analysis of Language", in A.J. Ayer (ed.), *Logical positivism*, New York: Free Press, pp60-81.

Chisholm, R.M., 1955, "Sentences about believing" in *Proceedings of the Aristotelian society*, Vol. 56, pp125-148.

Chomsky, N., 1955, "Review of Skinner's *Verbal behaviour*" in *Language*, Vol. 35, pp26-58.

Chomsky, N., 2000, *New horizons in the study of language and mind*, Cambridge: Cambridge University Press.

Chomsky, N., 2002, *On nature and language,* Cambridge: Cambridge University Press.

Church, A., 1936, "An unsolvable problem of elementary number theory" in *American journal of mathematics*, Vol. 58, pp345-363.

Church, A., 1943, "Carnap's Introduction to Semantics" in The Philosophical Review, Vol. 52, No. 3, pp298-304.

Church, A., 1956, *Introduction to mathematical logic*, New Jersey: Princeton University Press.

Churchland, P., 1981, "Eliminative materialism and the propositional attitudes" in *The journal of philosophy*, Vol. 78, No. 2, pp67-90.

Cole, D., 2009, "Jerry Fodor, LOT 2: The Language of Thought Revisited, New York: Oxford University Press, 2008, x+228, $37.95, ISBN 978-0-119-954877-4", in *Minds and Machines*, Vol. 19, No. 3, pp165-181.

Colyvan, M., 2001, *The indispensability of mathematics*, New York: Oxford University Press.

Conant, J., 2000, "Elucidation and nonsense in Frege and Early Wittgenstein" in A. Crary and R. Read (eds.) *The New Wittgenstein*, London: Routledge, pp174-217.

Copeland, J., 2002, "The Chinese room from a logical point of view" in J. Preston and M. Bishop (eds.) *Views into the Chinese Room: New essays on Searle and artificial intelligence*, Oxford: Oxford University Press, pp109-122

Correia, F.; Schnieder, B., 2012 "Grounding: an opinionated introduction" in F. Correia and B. Schnieder (eds.) *Metaphysical grounding: Understanding the structure of reality*, Cambridge: Cambridge University Press, pp1-36.

Cowie, F., 2003, *What's Within?: Nativism Reconsidered*. New York: Oxford University Press US.

Clark, A., 2003, *Natural born cyborgs: Minds, technologies, and the future of human intelligence*, Oxford: Oxford University Press.

Craik, K., 1966, "The mechanism of human action" in *The nature of psychology*, Cambridge: Cambridge University Press, pp7-90.

Craik, K., 1967, *The nature of explanation*, Cambridge: Cambridge University Press.

Cuda, T., 1985, "Against Neural Chauvinism" in *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, Vol. 48, No. 1, pp. 111-127.

Daitz, E., 1953, "The Picture Theory of Meaning" in *Mind*, Vol. 62, No. 246, pp184-201.

Daly, C., 2010, *An introduction to philosophical methods*, London: Broadview Press.

Davidson, D., 1963, "Actions, Reasons, and Causes" in *The Journal of Philosophy*, Vol. 60, No. 23, pp. 685-700.

Davidson, D., 1969, "True to the facts" in *The Journal of Philosophy*, Vol. 66, No. 21, pp748-764.

Davidson, D., 1984, *Inquiries into truth and interpretation*, Oxford: Clarendon Press.

Davies, B.D., 2007, *Doing a successful research project: Using qualitative or quantitative methods*, Hampshire; Palgrave Macmilan.

Damasio, A.R., 2005, *Descartes' error: Emotion, reason, and the human brain*, New York: Penguin.

Dennett, D., 1980, "The milk of human intentionality" in in *Behavioural and Brain Sciences*, Vol. 3, pp428–430.

Dennett, D., 1989, *The intentional stance*, London: MIT Press.

Dennett, D., 1991, *Consciousness explained*, New York: Back Bay Books.

Dennett, D., 2000, "True believers: The intentional strategy and why it works" in J. Haugeland (ed.) *Mind design II: Philosophy, psychology, artificial intelligence*, London: MIT Press, pp57-80.

Dennett, D., 2009, "The part of cognitive science that is philosophy" in *Topics in Cognitive Science*, Vol. 1, No. 2, pp231-236.

Denscombe, M., 2010, *The good research guide: For small-scale social research projects*, 4th edition, New York; Open University Press.

Devitt, M., 2014, "Linguistic intuitions are not "the voice of competence"" in M.C. Haug (ed.) *Philosophical methodology: The armchair or the laboratory*, London: Routledge, pp268-293.

Diamond, C., 1981, "What nonsense might be" in *Philosophy*, Vol. 56, No 215, pp5-22.

Dinsmore, J., 1991, *Partitioned representations: a study in mental representation, language understanding, and linguistic structure*. New York: Springer

Dretske, F., 1986, "Misrepresentation" in R.J. Bogdan (ed.) *Belief, form, content, and function*, Oxford, Clarendon Press, pp17-36.

Du, P.; Liu, D.; Zhang, L.; Hitchma, G.; Lin,C., 2014, "The processing of contradictory and non-contradictory negative sentences" in *Journal of Cognitive Psychology*, Vol.26, No.4, pp461-472.

Eco, U., 1976, "Peirce's Notion of Interpretant" in *Comparative Literature*, Vol. 91, No. 6, pp1457-1472.

Edwards, K., 2014, "Keeping (direct) reference in mind" in *Noûs*, Vol. 48, No. 2, pp342-367.

Eitel, A.; Scheiter, K.; Schüler, A.; Nyström, M.; Holmqvist, K., 2013, "How a picture facilitates the process of learning from text: Evidence for scaffolding" in *Learning and Instruction*, Vol. 28, pp48-63.

Elqayam, S., 2005, "Mental models, model-theoretic semantics, and the psychosemantic conception of truth" in *Philosophia Scientiae*, Vol. 9, No. 2, pp259-278.

Evans, G., 1982, *Varieties of reference*, Oxford: Oxford University Press.

Evans, J.; Handley, S.J.; Over, D.E., 2003, "Conditionals and conditional probability" in *Journal of experimental psychology: learning, memory and cognition*, Vol. 29, No. 2, pp321-335.

Feyerabend, P.K, 2011*, The tyranny of Science*, Cambridge: Polity Press.

Fine, K., 2012, "Guide to ground" in F. Correia and B. Schnieder (eds.) *Metaphysical grounding: Understanding the structure of reality*, Cambridge: Cambridge University Press, pp37-80.

Finocchiaro, M.A., 2005, *Arguments about arguments: Systematic, critical and historical essays in logical theory*, Cambridge: Cambridge University Press.

Fodor, J., 1975, *The Language of Thought*. Hassocks: Harvester Press.

Fodor, J., 1981, *RePresentations: philosophical essays on the foundations of cognitive science*, Brighton: Harvester.

Fodor, J., 1987, *Psychosemantics: The problem of meaning in the philosophy of mind*, London: The MIT Press.

Fodor, J., 1990, *A theory of content and other essays*, London: MIT Press.

Fodor, J., 1991, "A modal argument for narrow content" in *The journal of philosophy*, Vol. 88, No. 1, pp5-26.

Fodor, J., 1994, *The elm and the expert: Mentalese and its semantics*, London: The MIT Press.

Fodor, J., 1997, "The representational theory of mind" in *American behavioural scientist*, Vol. 40, No. 6, pp829-841.

Fodor, J., 2000, *In critical condition: Polemical essays on cognitive science and the philosophy of mind,* Massachusetts: MIT Press.

Fodor, J., 2001, *the mind doesn't work that way*, Massachusetts: MIT Press.

Fodor, J., 2007, "Semantics: An interview with Jerry Fodor" in *ReVEL*, Vo. 5, No. 8, pp1-12.

Fodor, J., 2010, *LOT 2: The language of thought revisited*, Oxford, Oxford University Press.

Fodor, J.A.; Fodor, J.D, 1980, "Remarks and replies: Functional structure, quantifiers and meaning postulates" in *Linguistic Inquiry*, Vol. 11, No. 4, pp759-770.

Ford, J., 2011, "Hellen Keller was never in a Chinese room" in *Minds and Machines*, Vol. 21, No. 1, pp57-72.

Ford, M., 1995, "Two modes of mental representation and problem solution in syllogistic reasoning" in *Cognition*, Vol. 54, No. 1, pp1-71.

Frege, G., 1948, "Sense and Reference" in *The Philosophical Review*, Vol. 57, No. 3, pp. 209-230

Frege, G., 1952, *Translations from the philosophical writings of Gottlob Frege*, edited by P. Geach and M. Black, Blackwell: Oxford.

Fresco, N., 2010, "Explaining computation without semantics: Keeping it simple" in *Minds and Machines*, Vol. 20, pp439-443.

Friedenberg, J.; Silverman, G., 2006, *Cognitive science: An introduction to the study of mind*, London, SAGE Publications.

Gallagher, S., 2005, *How the body shapes the mind*, Oxford: Oxford University Press.

Garnham, A., 1987, *Mental models as representations of discourse and text*, Chichester: Ellis Horwood.

Garnham, A., 1991, *The mind in action*, London: Routledge.

Garnham, A., 1996, "The other side of mental models: Theories of language comprehension" in *Mental models in cognitive science: Essays in honour of Phil Johnson-Laird*, Hove: Psychology Press.

Gendler, T.S., 2010, *Intuition, imagination and philosophical methodology*, Oxford: Oxford University Press.

Gettier, E. L., 1963, "Is Justified True Belief Knowledge?" in *Analysis*, Vol. 23, pp121-123.

Goldfarb, W., 2011, "Das uberwinden: Anti-metaphysical readings of the *Tractatus*" in *Beyond the Tractatus wars: The new Wittgenstein debate*, London: Routledge, pp6-22.

Goldin, D.; Wegner, D., 2008, "The interactive nature of computing: Refuting the strong Church-Turing thesis" in *Minds and Machines*, Vol. 18, No. 1, pp17-38.

Goldman, A.I., 1989, "Interpretation psychologized" in *Mind and Language*, Vol. 4, No. 3, pp161-185.

Goldvarg, E.; Johnson-Laird, P.N., 2001, "Naïve causality: a mental model theory of causal meaning and reasoning" in *Cognitive Science*, Vol. 25, No.4, pp565-610.

Goodwin, J.W.; Hein, U., 1982, "Artificial-intelligence and the study of language" in *Journal of Pragmatics*, vol. 6, no. 3-4, pp241-280.

Gray, D.E., 2009, *Doing research in the real world*, 2nd edition, London; SAGE Publications.

Griffin, J., 1963, *Wittgenstein's logical atomism*, London: University of Wahsington Press.

Gumperz, J.J.; Levinson, S.C., 1996, *Rethinking linguistic relativity: Studies in the social and cultural foundations of language*, Cambridge: Cambridge University Press.

Hacker, P.M.S., 1997, *Insight and illusion: themes in the philosophy of Wittgenstein*, Bristol: Thoemmes Press.

Hacker, P.M.S., 2001, "When the whistling had to Stop", in D.O.M. Charles and T.W. Child (eds.) *Wittgensteinian themes: Essays in honour of David Pears*, Oxford: Clarendon Press, pp13-48.

Hacker, P.M.S., 2013, *The intellectual powers: A study of human nature*, Oxford: Wiley.

Harnard, S., 1990a, "Against computational hermeneutics" in *Social epistemology*, Vol. 4., pp167-172.

Harnard, S., 1990b, "The symbol grounding problem" in *Physica D: Nonlinear phenomena*, Vol. 42, pp335-346.

Harnard, S., 1991, "Other bodies, other minds: A machine incarnation of an old philosophical problem" in *Minds and Machines*, Vol. 1, No. 1, pp43-54.

Harré, R., 2002, *Cognitive science: A philosophical introduction*, London: SAGE Publications.

Haug, M.C., 2014, ""Introduction – debates about methods: From linguistic philosophy to philosophical naturalism" in M.C. Haug (ed.) *Philosophical methodology: The armchair or the laboratory*, London: Routledge, pp1-26.

Haugeland, J., 2000, *Having thought: Essays in the metaphysics of mind*, London: Harvard University Press.

Haugeland, J., 2002, "Syntax, semantics, physics" in J. Preston and M. Bishop (eds.) *Views into the Chinese Room: New essays on Searle and artificial intelligence*, Oxford: Oxford University Press, pp379-392.

Hauser, L., 1997, "Searle's Chinese box: Debunking the Chinese room argument" in *Minds and Machines*, Vol. 7, pp199-226.

Heil, J., 2004, *Philosophy of Mind: a contemporary introduction,* London: Routledge.

Hughes, J.; Sharrock, W., 1997, *The philosophy of social research*, Harlow: Longman.

Hutchinson, P., 2007, "What's the point of elucidation?" in *Metaphilosophy*, Vol. 38, No. 5, pp691–713.

Hutchinson, P., 2008, *Shame and philosophy: An investigation in the philosophy of emotions and ethics*, Basingstoke: Palgrave.

Hutto, D., 2008, *Folk psychological narratives: The socio-cultural basis of understanding reasons*, London: MIT Press.

Ifenthaler, D; Seel, N, 2011. "A longitudinal perspective on inductive reasoning tasks: Illuminating the probability of change" in *Learning and Instruction*, Vol. 21, No. 4, pp538-549.

Ishiguro, H., 1969, "Use and reference of names" in P. Winch (ed.) *Studies in the Philosophy of Wittgenstein*, Oxon: Routledge, pp20-51.

James, W., 1892, "A Plea for Psychology as a 'Natural Science'" in *The Philosophical Review*, Vol. 1, No. 2, pp146-153.

Johnson-Laird, P.N., 1976, "To define true meaning" in *Nature*, Vol 264, pp635-636.

Johnson-Laird, P.N., 1977, "Procedural Semantics" in *Cognition*, Vol. 5, pp189-214.

Johnson-Laird, P.N., 1978a, "Meaning, amount of processing, and memory for words" in *Memory & Cognition*, Vol. 6, No., 4, pp372-375.

Johnson-Laird, P.N., 1978b, "What's wrong with Grandma's guide to procedural semantics: A reply to Jerry Fodor" in *Cognition*, Vol. 6, pp249-261

Johnson-Laird, P.N., 1980, "Mental models in cognitive science" in *Cognitive science 4,* pp71-115

Johnson-Laird, P.N., 1983, *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness,* Cambridge: Cambridge University Press.

Johnson-Laird, P.N., 1987, "The mental representation of the meaning of words" in *Cognition*, Vol. 25, pp189-211.

Johnson-Laird, P.N., 1988, "How is meaning mentally represented?" in U. Eco, M. Santambrogio and P. Violi (eds.) *Meaning and mental representations*, Bloomington: Indianna University Press, pp99-118.

Johnson-Laird, P.N., 1993, *The Computer and the Mind: An introduction to cognitive science*, London: Fortana Press.

Johnson-Laird, P.N., 1994a, "A model theory of induction" in *International Studies in the Philosophy of Science*, Vol. 1, No. 5, pp5-29.

Johnson-Laird, P.N., 1994b, "Mental models and probabilistic thinking" in *cognition*, Vol. 50, pp189-209.

Johnson-Laird, P.N., 1999, "Deductive reasoning" in *Annual Review of Psychology*, Vol. 50: 109-135

Johnson-Laird, P.N., 2002, "Peirce, logic diagrams, and the elementary operations of reasoning" in *Thinking and Reasoning*, Vol. 8, No. 1, pp69-95.

Johnson-Laird, P.N., 2005, "Mental models and thought" in K.J. Holyoak and R.G. Morrison (eds.) *The Cambridge Handbook of Thinking and Reasoning*, New York: Cambridge University Press, pp185-208.

Johnson-Laird, P.N., 2006, *How we reason*, Oxford: Oxford University Press.

Johnson-Laird, P.N., 2008a, "Mental models and deductive reasoning", in L. Rips and J. Adler (eds.) *Reasoning: studies in human inference and its foundations*, Cambridge: Cambridge University Press, pp206-222.

Johnson-Laird, P.N., 2008b, "How we reason: A view from psychology" in *The Reasoner*, Vol. 2, pp4-5.

Johnson-Laird, P.N., 2010, "Mental models and human reasoning" in *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 107, No. 43, pp18243-18250.

Johnson-Laird, P.N., 2011, "Mental models and language" in P.C. Hogan (ed.) *The Cambridge encyclopedia of language sciences*, Cambridge: Cambridge University Press, pp479-481.

Johnson-Laird, P.N., 2013, "Mental models and cognitive change" in *Journal of Cognitive Psychology*, Vol. 25, No. 2, pp131-138.

Johnson-Laird, P.N.; Khemlani, S.S., 2013, "Toward a unified theory of reasoning" in *Psychology of learnind and motivation*, Vol. 59, pp1-42.

Johnson-Laird, P.N.; Legrenzi, P.; Girotto, V., 2004, "How we detect logical inconsistencies" in *Current Directions in Psychological Science*, Col. 13, No. 2, pp41-45.

Johnson-Laird, P.N.; Yang, Y., 2008, "Mental logic, mental models, and simulations of human reasoning" in R. Sun (ed.) *Cambridge Handbook of Computational Psychology*, Cambridge: Cambridge University Press, pp339-358.

Katz, J.; Fodor, J., 1963, "The structure of a semantic theory" in *Language*, Vol. 39, No. 2, pp170-210.

Katz, J.; Postal, P., 1964, *An integrated theory of linguistic descriptions*, Cambridge, Mass.: MIT Press.

Kenny, A., 1975, *Wittgenstein*, Hamondsworth: Penguin.

Khemlani, S.; Orenes, I.; Johnson-Laird, P.N., 2012, "Negation: A theory of its meaning, representation, and use" in *Journal of Cognitive Psychology*, Vol. 24, No. 5, pp541-559.

Khemlani, S.; Orenes, I.; Johnson-Laird, P.N., 2014, "The negations of conjunctions, conditionals, and disjunctions" in *Acta Psychologica*, No. 141, pp1-7.

Kirkham, R.L., 1997, *Theories of truth: A critical introduction*, Mass.: MIT Press.

Kosslyn, S.M., 1980, *Image and mind*. Harvard University Press.

Kripke, S., 1972, *Naming and necessity*, Blackwell Publishers: Oxford.

Kripke, S.A., 1982, *Wittgenstein on Rules and Private Language: An Elementary Exposition*. London: Harvard University Press.

Kuhn, T., 1970, *The structure of scientific revolutions*, London: University of Chicago Press.

Kumar, R., 1996, *Research Methodology: A step by step guide*, London: SAGE Publications.

Kuusela, O., 2011 "The dialectic of interpretations: Reading Wittgenstein's *Tractatus*" in *Beyond the Tractatus wars: The new Wittgenstein debate*, London: Routledge, pp121-148.

Lakatos, I., 1984, *The methodology of scientific research programmes: Philosophical papers volume 1*, London: Cambridge University Press.

Lewis, D., 1970, "General semantics" in *Synthese*, Vol. 22, No. 1-2, pp18-67.

Locke, J., 1997, *An essay concerning human understanding*, London: Penguin.

Lugg, A., 2003, "Wittgenstein's Tractatus: True thoughts and nonsensical propositions" in *Philosophical Investigations*, Vol. 26, No. 4, pp332-347.

Macbeth, G. et al, 2014, "Models for the negation of conjunctions and disjunctions" in Europe's journal of psychology, Vol.10, No.1, pp135-149.

Marmolejo-Ramos, F., 2007, "New advances in the scientific study of text comprehension" in *Universitas Psychologica*, Vol. 6, No. 2, pp331-343.

Marr, D., 1982, *Vision*, New York: W.H. Freeman and Company.

Maxwell, J., 2013, *Qualitative research design: An interactive approach*, California: SAGE Publications.

McDowell, J., 1977, "On the sense and reference of a proper name" in *Mind*, Vol. 86, No. 342, pp159-185.

McDowell, J., 1984, "*De Re* senses" in C. Wright (ed.) *Frege: Tradition & Influence*, Oxford: Blackwell, pp98-109.

McDowell, J., 2009, *The engaged intellect: Philosophical essays*, London: Harvard University Press.

McGinn, C., 1989, *Mental Content*. Oxford: Basil Blackwell.

McGinn, M., 2014, "Liberal naturalism: Wittgenstein and McDowell" in M.C. Haug (ed.) *Philosophical methodology: The armchair or the laboratory*, London: Routledge, pp62-81.

Miller, A., 2007, *Philosophy of Language*, Ithaca: McGill-Queen's University Press.

Miller, G.A., 1979, "A very personal history", *Occasional paper no. 1*, Cambridge, Mass.: Center for Cognitive Science.

Moran, D., 1996, "The Inaugural Address: Brentano's Thesis" *in Proceedings of the Aristotelian society: Supplementary volumes*, Vol. 70, pp.1-27.

Morton, A., 2004, *Philosophy in Practice: An Introduction to the main questions*, Blackwell: Oxford.

Mounce, H.O., 1997, *The two pragmatisms: From Peirce to Rorty*, New York: Routledge.

Mounce, H.O, 2003, "Reply to Read and Deans" in *Philosophical Investigations*, Vol. 26, No. 3, pp269-270.

Müller, V.C., 2007, "Is there a future for AI without representation?" in *Minds and machines*, vol. 17, pp101-115.

Nagel, T., 1974, "What is it like to be a bat" in *The Philosophical Review*, Vol. 83, No. 4, pp435-450.

Nissini, M., 1997, "Ten cheers for interdisciplinarity: The case for interdisciplinary knowledge and research" in *The Social Science Journal*, Vol. 34, No. 2, pp201-216.

Norman, D.A.; Rumelhart, D.E., 1975, *Explorations in cognition*, California: W.H. Freeman and company.

Orenes, I.; Beltran, D.; Santamaria, C., 2014, "How negation is understood: Evidence from the visual world paradigm" in *Journal of memory and language*, Vol.74, pp36-45.

Osherson, D. N.; Lasnik, H., 1990, "The study of cognition" in D. Osherson and H. Lasnik (eds.) *An Invitation to cognitive science: Vol.1, Language* London: MIT Press, ppxi-xix

Oversteegen, E; Schilperoord, J., 2014, "Can pictures say no or not? Negation and denial in the visual mode" in *Journal of pragmatics*, Vol.67, pp89-106.

Papineau, D., 2001, "The rise of Physicalism" in C. Gillett & B.M. Loewer (eds.) *Physicalism and its discontents*. Cambridge: Cambridge University Press, pp3-36.

Papineau, D., 2004, *Thinking about consciousness*, Oxford: Clarendon.

Papineau, D., 2012, *Philosophical devices: Proofs, probabilities, possibilities and sets*, Oxford: Oxford University Press.

Papineau, D., 2014, "The poverty of conceptual analysis" in *Philosophical methodology: The armchair or the laboratory*, London: Routledge, pp166-194.

Patton, M., 2002, *Qualitative research & evaluation methods*, London: SAGE Publications.

Peirce, C.S., 1906, "Prolegomena to an apology for pragmaticism" in *The Monist*, Vol. 16, No. 4, pp492-546.

Peirce, C.S. 1931, *The collected papers of Charles Sanders Peirce, Volumes I-II*, edited by C. Hartshorne and P. Weiss, Cambridge, MA: Harvard University Press.

Peirce, C.S. 1933, *The collected papers of Charles Sanders Peirce, Volumes III-IV*, edited by C. Hartshorne and P. Weiss, Cambridge, MA: Harvard University Press.

Peirce, C.S. 1935, *The collected papers of Charles Sanders Peirce, Volumes V-VI*, edited by C. Hartshorne and P. Weiss, Cambridge, MA: Harvard University Press.

Peirce, C.S. 1958, *The collected papers of Charles Sanders Peirce, Volumes VII-VIII*, edited by A.W. Burks, Cambridge, MA: Harvard University Press.

Pezzulo, G; Calvi, G, 2011. "Computational explorations of perceptual symbol systems theory" in *New Ideas in Psychology,* Vol. 29, No. 3, pp275-297.

Pfeifer, R.; Fumiya I.; Max, L., 2014, "Cognition from the bottom up: On biological inspiration, body morphology, and soft materials" in *Trends in Cognitive Science*, Vol. 18, No. 8, pp404-413.

Pietarinen, A.-V., 2010, "Challenges and opportunities for existential graphs" in B.M. Paavola, A.-V. Pietarinen and H. Rydenfelt (eds.) *Ideas in action: Proceedings of the* Applying Peirce *conference*, Helsinki: Nordic pragmatism network, pp288-303.

Pietroski, P.M., 2005, "Meaning before truth" in G. Preyer and G. Peter (eds.) *Contextualism in philosophy: Knowledge, meaning, and truth*, London: Oxford University Press, pp253-300.

Pinker, S., 1999, *How the mind works*, New York: Penguin.

Plotkin, H., 2000, *Evolution in mind: An introduction to evolutionary psychology*, London: Harvard University Press.

Polger, T.W., 2007, "Realization and the metaphysics of mind" in *Australasian Journal of Philosophy*, Vol. 85, No. 2, pp233-259

Polger, T.W., 2012, "Functionalism as a philosophical theory of the cognitive sciences" in *WIREs Cognitive Science,* Vol. 2, No. 3, pp337-348.

Popper, K., 1957, *The poverty of historicism*, London: Routledge & Kegan Paul.

Popper, K., 1968, *The logic of scientific discovery*, London: Hutchinson.

Popper, K., 1974, *Conjectures and refutations: The growth of scientific knowledge*, London: Routledge.

Popper, K., 1994, *The myth of the framework: In defence of science and rationality*, London: Routledge

Popper, K.; Miller, D., 1983, *A pocket Popper*, London: Fontana.

Preston, J., 2002, "Introduction" in J. Preston and M. Bishop (eds.) *Views into the Chinese Room: New essays on Searle and artificial intelligence*, Oxford: Oxford University Press, pp1-50.

Putnam, H., 1960, "Minds and machines" in S. Hook (ed.) *Dimension of mind*, New York: New York University press, pp146-179.

Putnam, H., 1975, *Philosophical papers volume 2: Mind, language, and reality*, Cambridge: Cambridge University Press.

Putnam, H., 1977, "Realism and reason" in *Proceedings and addresses of the American philosophical association*, Vol. 50, No. 5, pp483-498.

Putnam, H., 1980, "The nature of mental states" in N. Block (ed.) *Readings in philosophy of psychology: Volume 1,* Massachusetts: Harvard University Press, pp223-231.

Putnam, H., 1981, *Reason, truth, and history,* Cambridge: Cambridge University Press.

Putnam, H., 1988, *Representation and reality*, MIT Press: London.

Pylyshyn, Z., 1984, *Computation and cognition: Toward a foundation for cognitive science*, Cambridge, Mass.: The MIT Press.

Pylyshyn, Z., 1989, "Computing in Cognitive Science" in M. Posner (ed.) *Foundations of Cognitive Science*, Cambridge: MIT Press, pp49-92.

Pylyshyn, Z., 1999, "What's in your mind?" in E. Lepore and Z. Pylyshyn (eds.) *What is cognitive science?* Malden, Mass.; Blackwell Publishers Ltd, pp1-26.

Quine, W.V.O., 1953, "Two dogmas of empiricism" in *From a logical point of view*, New York: Harvard University Press, pp20–46.

Quine, W.V.O., 1960, *Word and Object*, Cambridge, Mass.: Technology Press of the Massachusetts Institute of Technology

Quine, W.V.O., 1969, "Epistemology naturalized" in *Ontological relativity and other essays*, New York: Columbia University Press, pp69-90.

Quine, W.V.O., 1987, *Quiddities: An intermittently philosophical dictionary*, Cambridge, Mass.: Harvard University Press.

Quine, W.V.O., 1999, *Theories and things*, Cambridge, Mass.: Harvard University Press

Quine, W.V.O., 2008, "On empirically equivalent systems of the world" in D. Føllesdal and D.B. Quine (eds.) *Confessions of a confirmed existentialist and other essays*, Cambridge, Mass.: Harvard University Press, pp228-243.

Rapaport, W.J., 2007, "How Helen Keller used syntactic semantics to escape from a Chinese room" in *Minds and machines*, Vol. 16, pp381-436.

Rapaport, W.J., 2011, "Yes she was!: Reply to Ford's "Helen Keller was never in a Chinese room"" in *Minds and machines*, Vol. 21, No. 1, pp3-17.

Rapp, D.N.; Kurby, C.A., 2008, "The 'Ins' and 'Outs' of Learning: Internal Representations and External Visualizations" in J.K. Gilbert, M. Reiner and M. Nakhleh (eds.) *Visualization: theory and practice in science education,* Springer, pp29-52.

Read, R.; Deans, R., 2003, ""Nothing is shown": A 'Resolute' response to Mounce, Emiliani, Koethe and Vilhaur" in *Philosophical Investigations*, Vol. 26, No. 3, pp239-268.

Rickets, T., 1996, "Pictures, logic, and the limits of sense in Wittgenstein's *Tractatus*" in H.D. Sulga and D.G. Stern (eds.) *The Cambridge companion to Wittgenstein*, pp59-99.

Rodríguez, D., 2011, "Meaning in artificial agents: The symbol grounding problem revisited" in *Minds and machines*, Vol. 22, No. 1, pp25-34.

Roland, J.W., 2014, "On naturalism in the Quinean tradition" in *Philosophical methodology: The armchair or the laboratory*, London: Routledge, pp43-61.

Russell, B., 1918, "The philosophy of logical atomism" in *The Monist*, Vol. 28, No.4, pp. 495-527.

Russell, B., 2010, "Logical atomism" in *The philosophy of logical atomism*, Oxon: Fontana, pp126-150.

Russell, B.; Whitehead, A. N., 1999, *Principia mathematica to *57*, Cambridge: Cambridge University Press.

Schroyens, W., 2010, "Mistaking the instance for the rule: A critical analysis of the truth-table evaluation paradigm" in *Quarterly Journal of Experimental Psychology*, Vol. 63, No. 2, pp246-259.

Schweizer, P., 2012, "The externalist foundations of a truly total Turing test" in *Minds and machines*, Vol. 22, pp191-212.

Searle, J., 1980, "Minds, Brains and Programs" in *Behavioural and Brain Sciences*, Vol. 3, pp417–57.

Searle, J., 1984, *Mind, brains and science: The 1984 Reith lectures*, London: Pitman Press.

Searle, J., 1987, "Turing in the Chinese room" in *T.D.* Sing and R. Gomatam (eds.) *Synthesis of science and religion: Critical essays and dialogues*, Bombay: The Bhaktivedanta Book Trust, pp 295-301.

Searle, J., 1989, "Reply to Jacquette" in *Philosophy and phenomenological research*, Vol. 49, No.4, pp701-708.

Searle, J., 1990, "Is the brain's mind a computer program" in *Scientific American*, Vol. 262, pp20-25.

Searle, J., 2002, "Twenty-one years in the Chinese room" in in J. Preston and M. Bishop (eds.) *Views into the Chinese Room: New essays on Searle and artificial intelligence*, Oxford: Oxford University Press, pp51-69.

Siegel, H., 1984, "Empirical psychology, naturalized epistemology, and first philosophy" in *Philosophy of Science*, Vol. 51, No. 4, pp. 667-676.

Sellers, W., 1963, *Science, perception and reality*, London: Routledge & K.Paul.

Shaffer, M.J., 2009, "A logical hole in the Chinese Room" in *Minds and machines*, Vol. 19, pp229-235.

Shapiro, L.A., 2008, "Functionalism and mental boundaries" in *Cognitive Systems Research*, Vol. 9, Nos. 1–2, pp5–14.

Sharrock, W.; Read, R., 2002, *Kuhn: Philosopher of scientific revolution*, Malden, Mass.: Pollity

Shepard, R.N., 1978, "The mental image" in *American Psychologist*, Vol. 33, No. 2, pp125-137.

Shramko, Y., Wansing, H., 2009, "The slingshot argument and sentential identity" in *Studia Logica*, Vol. 91, pp429-455.

Simon, H. A.; Newell, A., 1976, "Computer science as empirical inquiry: Symbols and search", in *Communications of the ACM*, Vol. 19, No. 3, pp113–126.

Smith, M., 1987, "The Humean theory of motivation" in *Mind*, Vol. 96, No.381, pp36-61.

Sowa, J.F., 2011, "Cognitive architectures for conceptual structures" in S. Andrews, S. Polovina, R. Hill and B. Akhgar (eds.) *Proc. 19th International Conference on Conceptual Structures*, Heidelberg: Springer, pp. 35-49.

Starmans, C.; Friedman, O., 2012, "The folk conception of knowledge" In *Cognition,* Vol. 124, pp272-283.

Stern, D., 1995, *Wittgenstein on mind and language*, New York: Oxford University Press.

Straus, A.; Corbin, J. M, 2008, *Basics of qualitative research: Techniques and procedures for developing grounded theory*, Thousand Oaks, California: SAGE Publications.

Strawson, P.F., 1950, "On Referring" in *Mind*, Vol. 59, No. 235, pp320-344.

Strijbos, D.W.; de Bruin, L.C., 2012 "Making folk psychology explicit: The relevance of Robert Brandom's philosophy for the debate of social cognition" in *Philosophia*, vol. 40, pp139-163.

Swiatek, K., 2011, "Metaphysics and semantics of folk psychology" in *International Journal of the Humanities*, Vol. 8, No. 12, pp.35-43.

Thagard, P., 2009, "Why cognitive science needs philosophy and vice versa" in *Topics in Cognitive Science*, Vol. 1, No. 2, pp237-254.

Travis, C., 1989, *The Uses of Sense: Wittgenstein's Philosophy of Language*, Oxford: Oxford University Press.

Turing, A., 1936, "On Computable Numbers, with an Application to the Entscheidungsproblem" in *Proceedings of the London Mathematical Society*, Vol. 2, No. 42, pp230–65.

Turing, A., 1950, "Computing machinery and intelligence" in *Mind*, Vol. 59, No. 236, pp433-460.

Turing, A., 1969, "Intelligent machinery" in B. Meltxer and D, Michie (eds.) *Machine intelligence 5*, Edinburgh: Edinburgh University Press, pp3-26.

Verdonik, D., 2010, "Between understanding and misunderstanding" in *Journal of Pragmatics*, Vol. 42, No. 5, pp1364-1379.

Vogt, W. P.; Gardner, D.; Haeffele, L., 2012, *When to use what research design*, New York: The Guilford Press.

Vosgerau, G., 2006, "The perceptual nature of mental models" in C. Held, M. Knauff and G.Vosgerau (eds.) *Mental Models & the Mind: Current developments in Cognitive Psychology, neuroscience, and philosophy of mind*, Oxford: Elsevier, pp255-276.

Vosgerau, G., 2010, "Memory and content" in *Consciousness and cognition*, Vol. 19, pp839-846.

Walsh, C.R.; Johnson-Laird, P.N, 2004, "Co-reference and reasoning" in *Memory and Cognition*, Vol. 32, No. 1, pp96-106.

Waskan, J.A., 2006, *Models and cognition*, London: MIT Press.

Wason, P.C, 1966, "Reasoning" in B.M. Foss (ed.) *New horizons in psychology*, Harmondsworth; Penguin.

Wason, P.C, 1996, "Foreword: Early impressions of a research worker (1963-1980)" in *Mental models in cognitive science: Essays in honour of Phil Johnson-Laird*, Hove: Psychology Press.

Wason, P.C; Johnson-Laird, P.N., 1972, *The psychology of deduction: Structure and content*, Cambridge, Mass.: Harvard University Press.

Wason, P.C; Shapiro, D., 1971, "Natural and contrived experience in a reasoning problem" in *Quarterly journal of experimental psychology*, Vol. 23, pp63-71.

Williamson, K., 2000, *Research methods for students and professionals: information management and systems*, New South Wales: Centre for Information Studies.

Williamson, T., 2014a, "What is naturalism" in M.C. Haug (ed.) *Philosophical methodology: The armchair or the laboratory*, London: Routledge, pp29-31.

Williamson, T., 2014b, "The unclarity of naturalism" in M.C. Haug (ed.) *Philosophical methodology: The armchair or the laboratory*, London: Routledge, pp36-38.

Wittgenstein, L., 1929, "Remarks on logical form" in *Proceedings of the Aristotelian Society*, Supplementary Volumes, Vol. 9, pp162-171.

Wittgenstein, L., 1961, *Notebooks 1914-1916*, translated from German by G.E.M. Anscombe, New York: Harper and Brothers.

Wittgenstein, L., 1969, *The Blue and Brown books: Preliminary studies for the 'Philosophical investigations'*, Oxford: Blackwell Publishing.

Wittgenstein, L., 1974, *Philosophical grammar*, translated from German by A. Kenny, Oxford: Blackwell.

Wittgenstein, L., 1977, *Remarks on colour*, G.E.M. Anscombe (ed.), translated from German by L.L. McAlister and M. Schättle, Oxford: Blackwell.

Wittgenstein, L., 1990, *Tractatus Logico-Philosophicus*, translated from German by C.K. Ogden, New York: Routledge.

Wittgenstein, L., 1994, *Philosophical investigations*, translated from German by G.E.M. Anscombe, Blackwell: Oxford.

Wittgenstein, L., 2001, *Tractatus Logico-Philosophicus*, translated from German by D.F. Pears and B.F. McGuiness, New York: Routledge.

Wittgenstein, L., 2012, *Wittgenstein in Cambridge: Letters and documents 1911-1951*, edited by B. McGuiness, New York: Blackwell.

Wheeler, M., 2005, *Reconstructing the cognitive world: The next step*, Boston: MIT Press.

White, G., 2011, "Descartes among the robots: Computational science and the inner/outer distinction" in *Minds and machines*, Vol. 21, pp179-202.

von Wright, G.H., 1955, "Ludwig Wittgenstein: A biographical sketch" in *Philosophical Review*, Vol. 64, No. 4, pp527-545.

Zangwill, O.L., 1980, "Kenneth Craik: The man and his work" in *British journal of psychology* no. 71, pp1-16.