INVESTIGATIONS LINKING THE PHILOSOPHY AND PSYCHOLOGY OF MATHEMATICS

BY

JAMES KELLER

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Philosophy
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2020

Urbana, Illinois

Doctoral Committee:

Assistant Professor Noel Saenz, Chair
Associate Professor Andrew Arana, Director of Research
Associate Professor Jonathan Livengood
Assistant Professor Benjamin Levinstein

ABSTRACT

Recent progress in the field of cognitive science, specifically with respect to mathematical cognition, along with the turn in the philosophy of mathematics to a focus on mathematical practice, make for a great opportunity for interdisciplinary work that brings together the cognitive science of mathematics and philosophy of mathematics. This dissertation seeks to add to recent examples of such interdisciplinary work. I discuss three somewhat self-contained topics. In chapter two, I discuss some recent work in cognitive science on the topic of mental representation and show how we can shed light on a puzzle in the philosophy of mathematics by applying it to the issue of the mental representation of mathematics. In chapter three, I discuss a couple theories of the cognitive psychology of deduction and why some recent work in the philosophy of mathematics on diagrammatic reasoning in Euclidian proofs shows those theories have some shortcomings. Finally, in chapter four, I discuss how we can get at the question of what makes a system of representation, e.g. numeral systems or formal languages, good for doing mathematics, by considering how the cognitive system uses such things in the course of doing mathematics.

TABLE OF CONTENTS

CHAPTER 1: ON MATHEMATICAL COGNITION AND MATHEMATICAL PRACTICE

Consider the following two psychological views on how mathematicians solve problems. One view of how mathematicians solve problems, which I will call the *aha view*, is that mathematicians put effort into contemplating the relevant mathematical ideas which produces some insight and through this insight they see what the correct answer is. Some tasks might be performed, such as performing calculations, drawing pictures, or surveying examples, in order to facilitate a clearer understanding of the relevant mathematical concepts. But ultimately, it is the mathematician's understanding of these concepts that will eventually enable them to see what the solution is to the problem they are trying to solve. A different view of how mathematicians solve problems, which I will call the *I/O view*, is that mathematicians utilize a complex system of cognitive processes, some of which are unconscious and automatic, to solve problems by inputting the relevant information to get the desired output.[1] I take the aha view to be in line with the experience of learning and developing mathematics whereas the I/O view is more in line with modern cognitive science.

I do not take these two views to conflict with each other, the apparent conflict can be resolved by considering them as views from different perspectives or as views rooted in different explanatory frameworks. But, even if there is not really a conflict here, the differences between the views will lead to differences in how we frame and investigate questions about mathematical problem-solving. Furthermore, we can observe that something like the aha view has been around for a long time, while the I/O view is relatively new. The point is that it is worth considering

[1] See (Searle 1990) for much discussion both against and for the idea that unconscious mental processes can play the sort of explanatory role cognitive scientists often use them for.

what implications the I/O view has for all sorts of questions about mathematical problem-solving, both practical and philosophical.

Much work has recently been done along these lines. For example, we can consider the broad question of what strategies are most effective and efficient for learning mathematics. The recent (2014) implementation of the common core math standards in US public schools includes teaching math problem solving strategies that are different than what children learned in the past (much to the frustration of parents who cannot easily help their children with their homework). These new teaching methods are aimed at improving the effectiveness of math education and are in part motivated by cognitive science research, see (Marchitello and Wilhelm 2014) for details and references.[2] It is too soon at this point to evaluate the impact of the new teaching methods, and if there are gains I would suspect they will be marginal. But, given the value that is placed on mathematics education, even small gains in effectiveness would be worth the effort. Anyway, the larger point is just that by applying cognitive psychology to the question of how to most effectively teach mathematics, researchers proposed different strategies than what had been in place. So, we see the newer approach yielding at least slightly different answers.

Another example of applying recent work in cognitive science to some old questions, this time in the field of philosophy, is Marcus Giaquinto's *Visual Thinking in Mathematics* (2007). Giaquinto investigates the epistemological role of visualization in mathematics. In contrast to the historically predominant view (especially in the first half of the 20th century), which was that visualization could not have any significant epistemic role to play in mathematics, Giaquinto

[2] While most of the impact on math education has come from cognitive psychology, there is some literature on cognitive neuroscience and math education. See (Susac and Braeutigam 2014) for an argument that cognitive neuroscience is applicable to mathematics education. See (Ansari and Lyons 2016) for an overview of some research and some issues that need to be addressed for cognitive neuroscience research to be applicable to mathematics education.

argues that it does. Relevantly, a crucial part of his argument relies on considerations based on the cognitive science of perception. So here again, we have an approach to a question rooted in something like the I/O view yielding different answers.

Giaquinto's work is one among many recent works in the philosophy of mathematics that utilizes the cognitive science of mathematics. The aim of this dissertation is to add to this growing body of work. More specifically I want to demonstrate how an interdisciplinary approach to the philosophy and cognitive science of mathematics has the potential to foster progress in both fields and perhaps benefit mathematical practice itself. The subsequent chapters of this dissertation will provide examples that illustrate the benefits of the interdisciplinary approach. Chapter two will illustrate how the cognitive science of mathematics can help develop some answers to questions in the philosophy of mathematics. Chapter three will illustrate how some work in the philosophy of mathematics exposes some shortcomings in current theories in cognitive science. Finally, chapter four will illustrate why some work from this interdisciplinary approach could be of interest to practicing mathematicians.

In this chapter, I want to give some general background which will help put the work of the later chapters in context and relate them to each other. To that end, I want to touch on each of the philosophy of mathematics, the cognitive science of mathematics, and the aspects of mathematical practice that are of interest. I will consider this chapter successful if the reader gains the appropriate background information to put what follows in context (and is also interested enough to continue reading). I will proceed as follows. I will first discuss a distinction between two ways of doing the philosophy of mathematics, which will show how this work fits into a recent trend in the field. Next, I will discuss mathematical practice a bit and highlight the aspects that will be of interest going forward. Then I will give some general background on cognitive science as well as focus in on the current state of the cognitive science of mathematics specifically. Lastly, I will examine a critique of a theory of advanced mathematical cognition that currently exists and how it illustrates that the development of new theories of mathematical cognition can benefit by being informed by work in the philosophy of mathematics.

The Philosophy of Mathematical Practice

In this section, I will talk about an explicit focus on mathematical practice by some philosophers of mathematics. Following (Mancuso 2008), we can make a rough distinction between two sorts of philosophy of mathematics. First, there is the philosophy of mathematics that is concerned with the more or less traditional philosophical questions, e.g. metaphysical and epistemological questions, aimed at mathematics. This philosophy of mathematics is closely tied to the effort to secure the foundation of mathematics in a way that can guarantee against the possibility of inconsistencies. Second, there is the philosophy of mathematics that is concerned with questions that mathematical practice itself raises. This latter approach is often called 'the philosophy of mathematical practice'.[3]

To illustrate the idea of letting mathematics as it is practiced guide the inquiry, we can give a rough sketch of how a philosophical question could develop. We could start by observing that mathematicians will often develop multiple and varied proofs of the same theorem, even after it is commonly accepted that a particular proof definitively proves the theorem to be true. This suggests that proofs have more than just the one purpose of providing definitive justification for theorems. Thus, the question of what other purposes proofs have is raised. One idea, provided by observing how some mathematicians talk about certain proofs, is that some proofs play an

[3] There is a lot more to the distinction as Mancuso draws it, I am just highlighting the part relevant for my purposes here.

explanatory role in mathematics. So, we might investigate, as some philosophers of mathematics are actively doing now, what the nature of explanation in mathematics is and whether there something about certain proofs that make them explanatory. This example highlights that the distinction between these two sorts of philosophy of mathematics is mainly methodological. We could certainly imagine an epistemologist being curious about the nature of understanding and explanation in general and then aiming their investigation at the case of mathematics specifically. The difference is the method that guides what is investigated.

While the main difference is that philosophers of mathematical practice use observations and case studies from mathematical practice to guide their inquiry, there can also be a sort of normative attitude that this methodology is superior to the alternative. My motivation for the interdisciplinary approach is similar in spirit, to do the philosophy of mathematics and the cognitive science of mathematics as well as can be done, each need to be informed by the other. This claim should not be taken too polemically or as too grandiose. The point is just that, similar to how the philosophy of mathematics as a field is enriched when some practitioners also investigate mathematical practice and use that to inform their philosophical theorizing, the philosophy of mathematics and the cognitive science of mathematics can be enriched by some of its practitioners studying the other and applying what is relevant.

By tying this project in with the philosophy of mathematical practice, I do not mean to say that the cognitive science of mathematics is unrelated to more traditional debates in the philosophy of mathematics. Indeed, in chapter two, the philosophical points at issue are traditional issues in metaphysics and epistemology aimed at mathematics. Furthermore, as briefly mentioned above, some of Giaquinto's work aims to get at traditional epistemological questions aimed at mathematics. But, with that point made, much of the work I see being done by philosophers of mathematics that is informed by the cognitive science of mathematics is likely best classified as within the philosophy of mathematical practice. Furthermore, the work in the philosophy of mathematics that I think is most applicable to the cognitive science of mathematics is work from the philosophy of mathematical practice. We will see how work from the philosophy of mathematical practice is relevant to the cognitive science of mathematics both toward the end of this chapter and in chapter three. Now that I have situated this dissertation within a broader trend in the philosophy of mathematics, let us turn to discussing mathematical practice.

Mathematical Practice

In this section, I will highlight a couple parts of mathematical practice that are especially relevant to this interdisciplinary approach.[4] There are a couple useful approaches we can take to getting acquainted with mathematical practice. One approach is to start by looking at examples of things mathematicians do as mathematicians. For example, we can observe that mathematicians make definitions, construct examples, make conjectures, search for counterexamples, make informal arguments, and make formal arguments, among other things. Once we have an idea of what sorts of things mathematicians do, we can consider how mathematicians do these things (e.g. what is the strategy used for making a mathematical definition that is useful?) and why they do the things they do in the way that they do them (e.g. why make a new definition and what makes a definition a useful definition?). We might call this the what, how, and why of mathematical practice. This approach is useful because we need examples to both base our theories on and test our theories with.

[4] I am going to set aside the problems for referring to '*the* current mathematical practice' raised by disagreements among mathematicians about various aspects of mathematical practice. There is certainly enough agreement that we can avoid any areas of disagreement.

Another approach is to come up with a framework we can use that enables us to distinguish between and characterize different aspects of mathematical practices. Philip Kitcher takes such an approach in his (Kitcher 1984). For his own project, Kitcher wants a framework in which he can think and talk about changes to mathematical practice over time. Consequently, he finds it useful to characterize *a* mathematical practice and talk about change in mathematical practice as a move from one mathematical practice to another. Kitcher characterizes a mathematical practice as consisting of five components: a language in which mathematics is done, a set of accepted sentences, a set of important unsolved problems, a set of acceptable forms of reasoning, and a set of views about how mathematics should be done. This approach to considering mathematical practice is particularly useful at the moment, because it helps us to focus in on the parts of a mathematical practice that are of interest here, specifically, I am mainly interested in the set of views about how mathematics should be done and the language in which mathematics is done. These aspects of mathematical practice are especially relevant to the interdisciplinary approach.

Let us consider briefly the language in which mathematics is done. I will actually use the term ‘external representations’ instead, ‘representations’ because there is a great variety in the sorts of representations mathematicians use and ‘external’ because when we are discussing the cognitive science of mathematics we need to make it clear when we are talking about mental representations and we are not. First, I will give some examples. Then, I will talk a bit about the roles external representations play in mathematical practice. Lastly, I will say why they are of such interest in the context of an interdisciplinary approach to the philosophy and cognitive science of mathematics.

The external representations used in mathematical practice are varied and can be highly specialized in that the representations are used for particular subfields of mathematics. For example, knot diagrams are used specifically for knot theory, which is a subfield of topology. Euclidean diagrams, which will be discussed in chapter three, are used specifically in Euclidean geometry. Even the language of mathematics is developed to suit the specific purpose of doing mathematics. As discussed in (Ganesalingam 2013), the language of mathematics has some similarities with natural language, but also important differences. The most obvious difference is the mixture of symbols that appear to be a cohesive part of natural language and symbols which do not. Ganesalingam gives the example '$\{(x,y)$ in $\mathbb{N}^2$ | x and y are coprime$\}$', which demonstrates how the natural language like symbols can be contained within statements starting and ending with non-natural language like symbols. Distinct from this textual language, as Ganesalingam calls it, there are also formal languages in mathematics that are designed to be more suited to syntactic computation, one tradeoff being that unlike the textual language, formal languages are much less suitable for communication. For example, the language consisting of the logical constants of first-order logic (with identity), one non-logical constant (given here by the symbol 0), one binary operator (given here by the symbol +) and one unary operator (given here by the symbol *S*), is a formal language that can be used to express the formal theory of arithmetic known as Presburger arithmetic.

Not only is there variety in the sorts of representations used, but also in the way these representations are used in mathematical practice. I will go into more detail on some of the roles external representations play in chapter four, but the main takeaway here is that they are not merely tools for communication. For example, representations play a role in discovering and proving theorems. An obvious illustration of this role is the use of diagrams in proofs in

Euclidean geometry. The diagrams are not merely there to help illustrate what is going on in the proof but are rather an integral part of the proof. Recent work on this use of diagrams, see (Manders 2008), demonstrates that this use of diagrams was rigorous in the sense that it conforms to contemporary views of what constitutes a proof. Another illustration of this role of external representations, from (Giaquinto 2007), is how one could come up with an idea for a proof of the Intermediate Value Theorem by visualizing graphs of continuous functions in the context of thinking about Bolzano's Theorem.

As I will argue for a bit below (and even more substantially in chapter four), in many of the roles that external representations play in mathematical practice, they serve as cognitive tools. Consequently, an understanding of how the cognitive system uses external representations to enhance what it is capable of is crucial for a complete picture of mathematical cognition. So, it is obvious why the external representations would be of interest for the cognitive science of mathematics. The external representations of mathematics are of interest to the philosophy of mathematics just in virtue of the fact that they play an integral role in mathematical practice. For example, there are all sorts of interesting epistemic questions we can ask, one of the most basic being what role the external representations play in justifying mathematical beliefs. Note, we have both cognitive and epistemic issues intertwined here. If the external representations are playing a justificatory role, this could be very relevant to the cognitive scientist trying to understand the use of these external representations. Also, understanding how the cognitive system uses external representations could be very relevant to determining whether their use in a particular way really does provide justification. Due to how both cognitive and philosophical issues surrounding external representations in mathematics are intertwined, the topic is especially relevant to the interdisciplinary approach.

The second component of a mathematical practice I want to discuss is closest to what Kitcher calls 'the set of views about how mathematics should be done'. I will use the term 'heuristics' instead as I mainly want to discuss the strategies mathematicians employ to learn and develop mathematics. Of course, the interest goes beyond merely describing what heuristics mathematicians use, there is the related question of which heuristics are best. First, I will give a couple examples of heuristics and then discuss why they are of interest.

While the heuristics used in mathematical practice are largely left implicit (though some people have done some work toward making them explicit, e.g. Polya famously did so), it is evident that there are heuristics in mathematical practice and that knowing what the heuristics are is an important part of knowing how to practice mathematics. In philosophy, we similarly find that the heuristics of philosophical practice are largely left implicit (something to be picked up along the way throughout an effective philosophical education). Alan Hajek has recently done some work towards making the heuristics of philosophical practice explicit in both "Philosophical Heuristics and Philosophical Methodology" and "Philosophical Heuristics and Philosophical Creativity." One example he gives is the heuristic of checking extreme cases, which can be an efficient way of finding effective counterexamples to a claim. This heuristic is also used in mathematics for the same purpose.

Another heuristic that is sometimes used in solving some mathematical problem is to look for equivalent problems in other areas of mathematics and see if those equivalent problems are easier to solve. Giving an account of the conditions under which two problems are equivalent is too difficult a task to attempt here. But, to see how such a phenomenon can occur just consider that we can use some algebraic expressions to represent geometric objects and we can use set theoretic expressions as representations of numbers. So, we could turn a problem in geometry into a problem in algebra or we could turn a problem in number into a problem in set theory. A recent example of where this heuristic has been used is in solving the Kadison-Singer problem, which was an open problem for 54 years. The Kadison-Singer problem is a problem in functional analysis but had been shown to be equivalent to problems in several diverse areas of mathematics before one of these equivalent problems was solved (Casazza and Tremain 2006).

To see why the topic of the heuristics of mathematical practice is especially relevant to the interdisciplinary approach, consider the following example. In “Mathematics, Memory, and Mental Arithmetic,” W. T. Gowers asks the question, what features of mathematical statements and proofs make them easy to remember? The motivation for asking this question is the idea that mathematical statements and proofs that are easy to remember are better for practicing mathematicians to work with than statements and proofs that are harder to remember. Gowers goes on to make some hypotheses about what some of these features might be by reflecting on his own experience and considering some examples of proofs that seem harder to remember and proofs that seem easier to remember. He draws some distinctions between different types of memory and defines some features of proofs that he believes will be useful in investigating this issue. He does not come to any substantial conclusions but raises this issue precisely so that it can be studied further. There is a clear desire expressed for a heuristic for writing memorable proofs.

We already saw above that investigating why mathematicians often create multiple proofs of the same theorem is the sort of question that a philosopher of mathematical practice is interested in investigating. The fact that a prominent mathematician such as Gowers raises this question about proofs is of obvious interest to such an investigation, the idea behind Gowers question being that some proofs may have certain cognitive advantages that others do not. The question raised by Gowers also has obvious interest to a cognitive scientist of mathematics. How different features of proofs impact the ability to remember them is a question a cognitive scientist would investigate. So again, we have the intertwining of concerns both philosophical and cognitive. Philosophers of mathematics are concerned with what the various features of proofs are, e.g. whether a proof is explanatory and why, which could help inform the cognitive scientists' investigation. In turn, the cognitive scientists' investigation could help illuminate the value of certain proofs to the philosopher of mathematics. Furthermore, this example shows how the results of this sort of interdisciplinary investigation could be of interest to mathematicians, such an investigation could have implications for norms related to proof writing.

The topic of heuristics is also related to the topic of external representations and their use in mathematics. Generally, for a given task there are multiple external representations that could be used, which raises the question of which should be used. For example, some cognitive advantages of diagrams as compared to sentential representations are discussed in (Larkin and Simon 1987). One advantage of diagrams is that in some circumstances searching for some piece of information on a diagram can be much more efficient on average than searching in some equivalent sentential representation. So, there may be situations in which it is better to use a diagram than a sentential representation, and this is the sort of thing the interdisciplinary approach to the philosophy and cognitive science of mathematics would be well suited to investigate.

Now that we have an idea of what parts of mathematical practice are of particular interest to me for the interdisciplinary approach to the philosophy and cognitive science of mathematics, I will try to give an orientation of sorts to the cognitive science of mathematics as it currently stands. Obviously, I cannot hope to do more than to scratch the surface here. So, I will focus on what will be most useful for what follows after.

Current State of The Field of Mathematical Cognition

In this section, I aim to give some general comments about some of the main trends in cognitive theorizing and then briefly survey some work that has already been done on mathematical cognition. There are four different general frameworks, i.e. sets of basic assumptions about cognition and how it should be modelled, that will be useful to consider. I will call these the classical framework, the connectionist framework, the situated framework, and the Bayesian framework. These frameworks are not strictly speaking inconsistent with each other, but they do each represent a different approach to theorizing about cognition. To understand the differences between these frameworks I will briefly describe what the main idea of each framework is and how we would go about modelling a cognitive process in each framework.

What I call ‘the classical framework’ has as its central claim that the mind processes information by performing computations on mental representations. In this way the human mind processes information in a similar manner to a digital computer (historically speaking, the invention and development of digital computers provided some of the inspiration for this hypothesis). Various particular cognitive theories within this framework supplement this central claim with further claims about what the mental representations are and what computations are being performed on them. A classical model of a cognitive process will specify what is being computed and the representations and algorithm that carry out the computation.[5]

---

[5] It is not difficult to find a lot written about the computational theory of mind. One influential discussion of it is found in (Pylyshyn 1986).

A simple example of a classical model is a production system as introduced in (Newell and Simon 1972). A production system consists of a set of rules, where each rule specifies some action to be performed only if some condition obtains, and the system operates by recognizing what the current circumstance is and then implementing the appropriate rule. For example, in (Young and O'Shea 1981) a production system for written subtraction is given. The utility of this model is that it can be used to explain systematic errors in performance of subtraction by children, by showing that the errors occur as a result of missing rules. An example of a rule in their production system is as follows. "Rule CM says that an appropriate behavior when the two digits are known is to Compare them. Compare deposits in working memory an element indicating the relative sizes of the subtrahend and minuend digits, one of (S < M), (S > M), or (S = M)." [6] What is computed by the production system is the appropriate numeral to write down for a given problem. Various things are represented in the course of the computation, such as the relevant numbers, relations between numbers, etc. The rules implement an algorithm that is sufficient for doing the computation.

The situated framework takes on the basic assumption of the classical framework but views it as incomplete and so requires that something be added to it. The particular positions in this framework would then be a specification of what is being left out and how it is to be included in a model of a cognitive process. There are some common views on what is left out. For example, one such view, called 'embedded cognition', is that giving a full account of a cognitive process requires as part an account of the environment that the cognitive system is interacting with when this process occurs and what the interaction is like. As a result, much more emphasis is put on taking account of the situation and its impact on the cognitive system in

---

[6] To see an example of a computational model that is not a production system, see Marr's computational model of vision found in (Marr 1982).

developing an embedded model of a cognitive process (though the model itself may look very much like a classical model). Another example of a view in the situated framework, called ‘embodied cognition’, is that details about how human cognition works are in some way dependent on details about the human body (and not just the brain). As a result, when giving an account of a cognitive process, details about the body (and perhaps also how it is situated in and interacts with the world) will likely be important to consider in creating a model of the cognitive process.[7]

What I call ‘the connectionist framework’ has as its central claim that neural net models of cognitive processes are generally representative of how the mind actually processes information. A neural net consists of nodes, connections between the nodes, and weights associated with the connections. When a node in the network is activated, the activation will spread to other nodes depending on the connections and the associated weights. The weights associated to the connection can change resulting in a change in how activation propagates throughout the neural net. The processing of information by a neural net results from the relationships between the various nodes in the neural net. For a given pattern of activation of the input nodes in the network, activation will spread throughout the network and result in some pattern of activation of the output nodes. The structure of connectionist models is much more similar to the structure of the brain than classical models. Of course, cognitive scientists who work in the classical framework hold that their models are in some way implemented by the brain. But it is much clearer how this could work for a connectionist model.[8]

---

[7] See (Robbins and Aydede 2009), particularly the introductory chapter, for more about the situated framework (later chapters provide examples of work within the situated framework).

[8] Some cognitive scientists who would call themselves a ‘connectionist’, because they work with neural net models do not claim that neural net models necessarily give a complete account of a cognitive process and may think at a higher level, something like a classical model of cognition is appropriate. But other connectionists think that the

An example of a neural net model is a neural net for speech recognition, i.e. the neural net computes what word is being spoken based on some input sound. The input nodes correspond to different phonemes (perceptually distinct units of sound) and the output nodes correspond to words. Given some word, for example 'lamp' (which has four phonemes), if we activate the input nodes corresponding to the appropriate phonemes, the neural net will activate the output node corresponding to the word 'lamp'. Clearly, a neural net that could recognize any English word would have to be very large, since it would need one output node per word. The neural net is created by training it to produce the correct output given some input. The process of training the neural net involves changing the weights of the connections between the nodes in a way (e.g. the Hebbian learning rule) such that after training the neural net with a large number of various inputs and outputs, it will eventually start to produce the correct output for a given input.

What I call 'the Bayesian framework' can easily be distinguished from the other frameworks by its use of probability in modelling cognitive processes. Bayesian models have been created for inference, decision making, and learning, as well as language processing and vision. See (Chater et. al 2006) for more about the motivation for Bayesian models. The simplest Bayesian model will consist of some set of propositions and a joint probability distribution over this set. For example, consider a simple Bayesian model of visual object categorization. The computation to be performed is to determine what object caused some given visual experience. The set of propositions would include things such as aspects of visual experience and different objects. The joint probability distribution can be thought of in terms of a probability distribution over the set of possible objects (which we can think of as the prior distribution) and conditional probabilities which give the likelihood that a particular object is present given some possible

---

classical framework is totally misguided and that neural net models give us a complete account of cognitive processes.

aspect of visual experience is actually experienced. A new probability distribution is generated by conditionalization on whatever experience is had. The object with the highest posterior probability is what is deemed the cause of the visual experience. See (Kersten and Yuille 2003) for a more detailed Bayesian model of visual object categorization.

Given these approaches to cognitive science, we can see that depending on which approach we take to mathematical cognition we will get some rather different looking models of the cognitive processes involved. All of the cognitive science in the rest of the dissertation will be within the classical framework, with the exception of Lakoff and Nunez's theory of conceptual metaphor which I will discuss briefly below and Barsalou's theory of mental representation that I discuss in chapter two, both of which are within the situated framework.

I will now consider some of the work that has been done on mathematical cognition. Specifically, I will discuss the research on numerical cognition (exemplified by Stanislas Dehaene's *The Number Sense*) and Lakoff and Nunez's theory that mathematical concepts are conceptual metaphors. Dehaene's work focuses on the issue of mental number representations. In chapter two, I will go through the current theory of mental number representation in some detail. The point I would like to make here is just that cognitive scientists know quite a lot about mental number representation. But it is also important to note that most of the research on mathematical cognition is just about numerical cognition. For example, in the *Handbook of Mathematical Cognition* (Campbell 2005), the vast majority of articles are about numerical cognition. Another illustrative example is the following quote from a chapter titled "Mathematical Cognition" in the *Cambridge Handbook of Thinking and Reasoning*: "Mathematics is a system for representing and reasoning about quantities, with arithmetic as its foundation" (Gallistel and Gelman 2005). It would seem from this quote that the authors just take mathematical cognition to be numerical cognition. While numerical cognition is distinctive of mathematics, there is a lot more to consider in mathematical cognition beyond it. Even if arithmetic is the foundation of mathematics, mathematics quickly moves beyond arithmetic and different cognitive processes are involved in advanced mathematical thinking (also, arithmetic is arguably not the foundation of mathematics, if there is a singular foundation of mathematics at all). The most prominent work in mathematical cognition that is not focused on numerical cognition is the work of George Lakoff and Rafael Nunez.[9]

In "The Cognitive Foundations of Mathematics: The Role of Conceptual Metaphor," Lakoff and Nunez present a theory about cognition involving abstract concepts and then apply it to the case of mathematical cognition. The theory is essentially that the way we reason about abstract concepts is ultimately explained by the sensory motor system. According to their theory, the human sensory motor system causes humans to reason about spatial relations, e.g. above, on, supported by, etc., in particular ways and they refer to these stable inferential patterns involving particular spatial relations as image schemas. Through an unconscious and automatic process, which they call 'conceptual metaphor', the inferential patterns from a particular image schema are mapped onto some domain of abstract concepts. An example they give is the abstract concept of time which they show has the same inferential patterns as motion along a straight line. So, under this theory, the way in which mathematicians (informally) reason about mathematical concepts results from the process of conceptual metaphor.

To argue for this view of mathematical concepts, Lakoff and Nunez use a technique they call mathematical idea analysis in which a mathematical concept is shown to have the inferential

[9] Just to be clear this is not the only exception. Another example is research on the role of spatial concepts and reasoning in geometric thinking. A discussion of some of this research can be found in Marcus Giaquinto's *Visual Thinking in Mathematics*.

structure as some image schema. To some extent this involves looking at the informal language used by mathematicians when they describe how they reason about a particular concept. They also point out that metaphors can be pedagogically useful in mathematics to provide support for their view. One example of mathematical idea analysis they provide is an analysis of the concept of a set. They propose that our informal reasoning about sets is based on the 'a set is a container' conceptual metaphor.

One last takeaway from Lakoff and Nunez's theory is the observation that Nunez derives a philosophical implication from their theory in (Nunez 2008). Specifically, the implication is that mathematical truth is relative. This follows from their claim that the way in which a mathematician will reason is relative to the conceptual metaphor they use and their claim that we can have multiple conceptual metaphors for a mathematical concept with no mathematical basis to prefer one over the other.[10] Setting aside the merit of this view, it does illustrate why their theory would be of interest to philosophers of mathematics.

In this section, I pointed out that much of the work on mathematical cognition focuses just on numerical cognition, with one exception being Lakoff and Nunez's theory. But as we will see in the next section there is some reason to be concerned that Lakoff and Nunez's theory has significant problems. Also, even though Lakoff and Nunez's theory does go beyond numerical cognition, there are many other aspects of mathematical cognition to theorize about. Furthermore, as I pointed above, there are several different approaches to modelling cognitive processes, and it is worth exploring how they apply to mathematical cognition. So, while there is

[10] Nunez writes, "when imaginary entities are concerned, truth is always relative to the inferential organization of the mappings involved in the underlying conceptual metaphors" which I take to support my claim that they endorse the first conjunct. Nunez also writes, "... there is no ultimate truth regarding human imaginative structures" where human imaginative structures includes mathematics, which I take to support my claim that they endorse the second conjunct (Nunez 2008).

already plenty of great research and theorizing done in the field of mathematical cognition, there is still a need to develop new theories of mathematical cognition

Applying the Philosophy of Mathematics to the Cognitive Science of Mathematics

In this section, I want to illustrate how the philosophy of mathematics can help in developing new theories of mathematical cognition. The main idea is just that philosophers of mathematics have learned a lot from investigating mathematical practice, and this work can and should inform theorizing about mathematical cognition. To illustrate I will consider an objection that Dirk Schlimm, a philosopher of mathematics, has raised to Lakoff and Nunez's theory.

In "Conceptual Metaphors and Mathematical Practice" Dirk Schlimm points out that the observations used by Lakoff and Nunez to justify their theory are based on a distorted picture of mathematical practice. Specifically, Schlimm takes their view to be based on what he calls the textbook view of mathematics, i.e. the view of mathematics we get when we look only at mathematical notions that has been sufficiently developed by mathematicians that they can be put into a textbook. Building a theory on only some of the relevant observations is of course potentially problematic as the theory may not fit the relevant observations not considered. The cognition involved developing mathematical concepts is the sort of mathematical cognition Lakoff and Nunez are theorizing about and so needs to be explainable by Lakoff and Nunez's position. Schlimm demonstrates that if we look at the early conceptions of sets, we see not just one concept of sets, but many (even just Cantor alone seems to have had many distinct conceptions of a set throughout his work developing set theory). So, to get to the modern notion of a set, mathematicians had to deliberate about which conception of a set was best. So, the claim that the inferential patterns that involve our informal notion of a set arise from an unconscious and automatic cognitive process seems to ignore the history of how the mathematical concept of a set developed. The way in which Cantor developed his ideas about sets in a conscious and deliberate way is a counterexample to Lakoff and Nunez's theory. While it may be the case that our reasoning about some mathematical concepts result from conceptual metaphor, it is not obvious that conceptual metaphor plays as significant of a role in shaping mathematics as Lakoff and Nunez's theory implies.

The main takeaway from Schlimm's objection is that we need to base our theories on observations not just from how mathematical thought is presented in textbooks but also observations from the development of mature mathematics. Fortunately, with the recent focus on mathematical practice by philosophers of mathematics, there is much work that has been recently done and is currently being done on historical case studies of mathematical practice, which can provide us with relevant data.

While we are on the topic of what sorts of observations are important, I want to say a bit about the role of first-person observations. We should be wary of first-person observations.[11] One motivation for this is the recognition that some cognitive processes are unconscious and automatic in which case there would be no direct first-person experience of these cognitive processes. Consequently, first-person observations might give us an incomplete picture of what is going on. Another motivation for being wary of first-person observations is that first-person observations can be unreliable. For example, people tend to overestimate the scope of their visual acuity, see (Schwitzgebel 2008).

[11] I take this to be the standard position currently in cognitive science.

The following is an example of the sort of argument that I will take a skeptical position toward. In (Devlin 2006), Keith Devlin argues that "the nature of mathematical thought is not linguistic." He writes:

> What is the nature of mathematical thought? Although I have been a mathematician for forty years, I am still not clear exactly what the nature of the mathematical thought process is. I am sure it is not linguistic, or at least not totally so, and probably not mostly so. Mathematicians do not think in sentences; at least not most of the time. The precise logical prose you find in mathematical books and papers is an attempt to communicate the results of mathematical thought. It rarely resembles the thought process itself. I am in remarkably good company in having this view of mathematical thinking. For instance, in 1945, the distinguished French mathematician Jacques Hadamard published a book titled *The Psychology of Invention in the Mathematical Field*, in which he cited the views of many mathematicians on what it feels like to do mathematics. Many of them insist that they do not use language to think about mathematics… Of particular relevance to my thesis are the mathematicians' descriptions of the way they arrived at the solutions to problems they had been working on. Time and again, the solution came at a quite unexpected moment, when the person was engaged in some other activity and was not consciously thinking about the problem. Moreover, in that inspirational moment the whole solution suddenly fell into place, as if the pieces of a huge jigsaw puzzle had been dropped onto the floor and miraculously landed as a complete picture. The mathematician "saw" the solution and instinctively knew it was correct. No language is involved in this process. Indeed, with a problem for which the solution is fairly complex, it might take the mathematician weeks or even months to spell out (in linguistic form) the step-by-step logical argument that constitutes the official solution to the problem — the proof of the result.

Devlin wants to determine what the "nature of mathematical thought" is and I take this to mean that he wants to determine what sorts of mental representations are utilized in advanced mathematical thought. The importance of being skeptical toward such arguments is illustrated when we consider some work that has recently been done on the role of linguistic and symbolic representation in mathematical cognition. There is some evidence that some cognitive tasks are completed just using mental representations of linguistic and symbolic representations (as opposed to the mental representations of the referents of these linguistic and symbolic representations). A clear example of this is mental arithmetic, which some theorize is performed using the mental representations of numerals, see for example (Johansen 2010). But if we take Devlin's argument seriously then we would be pushed toward rejecting this as a possibility. I think that rejecting this possibility would be a mistake.

My point is not that Devlin is surely wrong, just that the first-person observations of mathematicians should not be taken to rule out a theory of mathematical cognition. If Schlimm's objection were just based on testimony from Cantor's experience, I do not think it would be a significant objection, but there is much more to the observations Schlimm brings to bear against Lakoff and Nunez's theory. First-person observations from mathematicians certainly have value. For example, in chapter three I will consider and to some extent rely on mathematicians' judgments about what proofs do and do not convey understanding. In this case the judgment is made by introspecting on whether one does or does not understand. The point is just that we need to mindful of the role these sorts of observations are playing in developing and evaluating our theories, especially because they have the potential to be very useful.

Conclusion

In this chapter, I discussed the idea of taking an interdisciplinary approach to the philosophy and cognitive psychology of mathematics. We considered that some philosophers of mathematics use observations of mathematical practice to guide their investigation. The cognitive science of mathematics can play a similar role in the philosophy of mathematics and vice versa, to both fields' benefit (as some existing work already demonstrates). I then focused on a couple aspects of mathematical practice that are especially relevant to this interdisciplinary approach, the heuristics of mathematical practice and the external representations used in mathematical practice. In both cases we find cognitive and philosophical issues intertwined with each other. Next, I gave some background on the current state of the cognitive science of mathematics, the main takeaway being that while there is a lot of great work already done, much of it focuses just on numerical cognition. Finally, I pointed out that recent work in the philosophy of mathematics that is focused on mathematical practice can be a great source of observations relevant to developing new theories of mathematical cognition.

To conclude this chapter, I will give a brief overview of the chapters to follow and say how they fit into the interdisciplinary approach to the philosophy and cognitive science of mathematics. In chapter two, I will do some work toward creating a general theory of mathematical cognition that could serve as an alternative to Lakoff and Nunez's theory. The goal will be to do some preliminary work on theorizing about the mental representations that mathematicians employ when doing mathematics. I will begin by considering the consensus theory of mental number representation. Then, to move beyond mental number representation, I will discuss the Language and Situated Simulation theory of mental representation and how we can apply it to mathematics. Philosophically the upshot will be that, unlike Lakoff and Nunez's theory, this theory of mental representation is ontologically neutral and so would be consistent with a view like Platonism.[12] This chapter will fit into the interdisciplinary approach in a couple ways. One way is that is considers the philosophical implications that theories of mental representation in mathematics have. Another way is that it works toward a novel theory of mathematical cognition.

In chapter three, I will argue that the psychological process of deduction plays an integral role in gaining explanatory understanding in mathematics. With that established, we can relate some recent work in the philosophy of mathematics on explanatory proofs with work in cognitive psychology on deduction. The upshot will be that there are relevant examples of deduction that these theories of the psychology of deduction cannot account for and so they are

[12] Arguably Lakoff and Nunez's theory of mental representation in mathematics should also be considered ontologically neutral. Along with their theory of conceptual metaphor, they also make the claim that there is no basis for distinguishing between correct or incorrect conceptual metaphors, which in turn implies mathematics is subjective (see footnote 11).

at best incomplete. This fits into the interdisciplinary approach in that it illustrates how we can bring the philosophy and cognitive science of mathematics together to the benefit of cognitive science.

In chapter four, I will discuss the external representations used in mathematical practice. I will consider some ways that they are used by the cognitive system in practicing mathematics. I will use some ideas from cognitive science to help define a couple features of external representations and then I will argue that these features are good-making features of external representations in mathematics. This fits into the interdisciplinary approach in that it uses observations on the roles of external representations, as is done in the philosophy of mathematical practice, and applies some ideas from cognitive science to make some normative claims about the external representations of mathematics.

CHAPTER 2: ON THE MENTAL REPRESENTATION OF MATHEMATICS

In this chapter, I will give an account of the mental representation of mathematical objects, operations, properties, and relations that is consistent with a Platonist ontology of mathematics while also being consistent with a sort of naturalism and a sort of empiricism. As we will see, the account is naturalistic in the sense that it is based on contemporary views in cognitive science and it is empiricist in the sense that the mental representations are acquired through perception. The account I will give is based on a theory of mental representation from (Ramsey 2007) and a theory of mental representation from (Barsalou et al. 2008), called the Language and Simulated Simulation theory (LASS). As we will see these theories cover different ground and so are not in competition. Roughly speaking, the central idea of the account I will give below is that the cognitive system uses what are originally perceptual representations in such a way that they come to represent mathematical objects, properties, etc. While showing that mental representations which satisfy the above conditions plausibly exist is of interest, using the two theories mentioned above, we can get some idea of what these mental representations are like, which gives some direction for future research.

There has been some philosophical work done recently similar to what I do in this chapter. Helen De Cruz's "Numerical Cognition and Mathematical Realism" (2016) is quite similar in aim. I will discuss her work and how it relates to mine below. Also similar is what is done in C.S. Jenkins' *Grounding Concepts* (2008) and in Markus Pantsar's "An empirically feasible approach to the epistemology of arithmetic" (2014). Jenkins' goal is to give an account of how we can have arithmetical knowledge that is empirical (in a stronger sense than I use above), a priori, and objective. Pantsar's project is aimed at showing that the current consensus view of mental number representation is consistent with arithmetical knowledge being in some sense a priori. The two projects are similar to mine in that they take on the challenge of showing how some seemingly inconsistent criteria for a theory of mathematical thought can be satisfied. The main difference is that these two projects are epistemological whereas mine falls more in the philosophy of mind.

The philosophical motivation for this project has to do with satisfying seemingly inconsistent intuitions about mathematics and mathematical cognition. Those intuitions are just the three mentioned above, Platonism, naturalism, and empiricism. Platonism has been objected to famously by Benacerraf (1973) on epistemological grounds, which has led to further objections based on a concern about cognitive access to abstract objects. There are further objections to Platonism based on theories of mathematical cognition, e.g. (Dehaene 1997) and (Lakoff and Nunez 2000). The objection to Platonism based on these theories is simply that these theories of mathematical cognition, along with a couple plausible philosophical assumptions, are not consistent with Platonism. My account will not engage directly with Benacerraf-type objections but could eventually serve as part of a counterargument to them. My account will give Platonists who have a sort of naturalistic outlook a compatible view of the mental representation of mathematical objects, unlike those of Dehaene and Lakoff and Nunez.

From the perspective of cognitive science, the motivation for this project is to increase the number of plausible accounts of the mental representation used in mathematical cognition. Lakoff and Nunez give the only general account of mathematical cognition, which they get by applying their theory of conceptual metaphor to mathematics (Dehaene's account is focused just on mental number representation). So, this can in part be viewed as part of a larger project to lay out the various possibilities. A narrower motivation is that mathematics is a source of difficulty for empiricist theories of mental representation, which claim that mental representations are acquired through perception. If we take mathematical objects, e.g. numbers, to be abstract objects (and so non-spatiotemporal), it is not obvious how such theories can account for our mental representations of them. For such theories of mental representation to be plausible, they need to be able to account for our mental representations of mathematical objects (of course the alternative route is to deny mathematical objects are abstract objects). LASS is one of these empiricist theories and so we will see how at least one of these theories can overcome this difficulty.

This chapter will proceed as follows. First, I will consider the consensus view on mental number representation in cognitive psychology. I will then show that there is a view of mental number representation based on the consensus view that is consistent with Platonism if we adopt a specific view about the content of mental representations. I will then point out a couple reasons why it is difficult to extend this account to mathematical operations, properties, and relations. I will be using ideas from LASS in order to overcome the difficulties posed. So, I will then explain what LASS is in detail, review some of the evidence for it, and show how it fits with the view of mental number representation already discussed. Finally, I will show how we can extend the account of the mental representation of mathematical numbers to operations, properties and relations using LASS.

Mental Number Representations

Any account of mental number representation will be involved in explaining how our various numerical thoughts and abilities work. For example, consider the following task. You have to plan a lunch for an upcoming conference and in particular you are asked to come up with budget for the lunch. One way to come up with a budget is to first determine how many people will be eating lunch, then determine how much it will cost to feed each person, and finally multiply the cost per person by the number of people who will be eating lunch in order to get the total cost. In order to do this, you need to represent the various relevant quantities and then perform arithmetic operations on those quantities to get the desired answer. An adequate account of mental number representation will be able to provide part of an explanation of how it is we complete this task, by specifying what the mental representations are like that are used in the cognitive processes involved in this task. There are of course many other tasks that presumably involve mental number representations. So, an account of mental number representation should be useful in explaining our ability to perform a variety of tasks. As we will see, there are actually multiple types of mental number representation posited to account for our numerical abilities.

What follows will be an overview of the consensus view on mental representation and some of the research that supports it. While there are many unsettled questions about numerical cognition in general and about mental number representation specifically, there is a consensus that the three types of mental number representation I will consider below exist. The three types of mental number representation I will discuss are analog magnitude representations, parallel individuation representations, and mental numeral representations. After going through the three types of mental number representation I will summarize an account of mental representation from (Ramsey 2007) and argue that the vehicles of mental numeral representations can serve to represent numbers.

A brief terminological note: I will use sometimes use the term ‘number’ when I am referring to a quantity. Assessing and performing computations on quantities is one of the main numeric tasks that mental number representations are used for, so the literature on mental number representation sometimes uses the terms ‘number’ and ‘quantity’ interchangeably. The view that numbers are abstract objects, is not necessarily compatible with holding that numbers are quantities. Given that I want to show Platonism is consistent with naturalism and empiricism my using these terms interchangeably should not be taken as a tacit endorsement of the position that numbers and quantities are the same thing.

Analog Magnitude Representation

The first type of mental number representation I will consider is called an *analog magnitude representation*. Dehaene, who has done some of the main work showing that these mental representations exist, uses the following analogy to help explain what this type of representation is like. Dehaene introduces the idea of an accumulator. Suppose we have bucket under a water faucet. The bucket in this scenario is the accumulator. I can use the water level in the bucket to represent the quantity of members in a finite set as follows. For each new individual object in the set I turn on the water for 10 seconds. The water level in the bucket depends on just the quantity of objects in the set. In this way the water level can serve to represent quantity. The water level is an analog magnitude (analog because it is a continuous variable and magnitude because the water level depends on the amount of water). As the water level is a continuous variable, we should expect the water level representation of number to be somewhat imprecise. Every time I take stock of a new individual in the set, I am supposed to turn on the water for precisely 10 seconds, but I will inevitably fail to have the water on for precisely 10 seconds every time. If I represent two sets of 100 individuals with two different buckets of water, we should expect the water level in those buckets to differ slightly. There is a lot of evidence pointing to the existence of mental representations of number that are analog magnitude representations. It is plausible that there is some neural mechanism that is effectively an accumulator. Dehaene has created a neural net model that acts as an accumulator, which demonstrates that such a neural mechanism could exist.

The most impressive evidence for the existence of analog magnitude representations comes from number discrimination tasks. Number discrimination tasks are tasks in which a person is presented with two numbers (represented either with arrays of dots or sequences of tones) and tested on whether they can discriminate between the two numbers, i.e. to correctly detect whether the numbers are different or not. We can make the following prediction. If humans' mental number representations are analog magnitude representations, then we should expect to make the following observations in number discrimination tasks. Given pairs of numbers where the difference is small, say a difference of two (such as the pair 1 and 3 or the pair 95 and 97), we should observe higher rates of error in judgments for larger pairs of numbers than for smaller pairs of numbers, because analog magnitude representations of numbers are less precise for larger numbers than for smaller numbers. Also, given two pairs of numbers where one has a smaller difference than the other (such as the pair 1 and 2 and the pair 1 and 3), we should expect a higher rate of success in correctly judging that there is a difference in the pair with the greater difference than in the pair with the smaller difference. These predictions match the observations made by researchers.

Consider for example the study detailed in (Xu & Spelke 2000). In this study six-month infants were tested for the ability to discriminate between the numbers 8 and 16 and between the numbers 8 and 12. The evidence shows that infants of this age can reliably discriminate between 8 and 16 but cannot reliably discriminate between 8 and 12. This is demonstrated through a habituation experiment. The idea behind a habituation experiment is that infants' attention is focused on a novel stimulus for longer than a stimulus they have become accustomed to. Infants are habituated to arrays of 8 dots that vary in things like size of the dots and distance between the dots so we can be confident that the infant is habituated just to the number of dots. The infants are then presented with an array of 12 or 16 dots and we observe whether the array with more than eight dots is a novel stimulus for the infant (some infants were habituated to the larger number and then tested on whether eight was a novel stimulus as well). As it turns out infants do look longer when going from arrays of 8 dots to arrays of 16 dots but do not look longer when going from arrays of 8 dots to arrays of 12 dots. This shows that infants cannot discriminate between 8 and 12 but can discriminate between 8 and 16. This in turn is evidence that infants' representations of number are analog magnitude representations.

Parallel Individuation Representations

There are some observations from experiments involving infants that are not well explained by analog magnitude representations of number. These observations have led some cognitive scientists to posit a second type of mental number representation called a *parallel individuation representation (*also called an *object-file representation*). One such study, detailed in (Feigenson & Carey 2003), was aimed at testing infants' (12-14 months old) ability to discriminate between various small quantities. In the study, infants observed someone place different quantities of objects in box that the infant can reach into but not see into. The infant is then allowed to reach into the box to retrieve the objects placed into it. When infants see one, two or three objects placed in the box they regularly search with their hand in the box until the number of objects placed in the box are retrieved. This demonstrates that they can discriminate between sets of size one, two and three. However, when an infant sees four objects placed in the box, they regularly stop searching after they retrieve just one or two of the objects. This would seem to indicate that the infant cannot discriminate between sets of size one or two and sets of size four.

If infants had analog magnitude representations for small numbers (specifically one, two, and three), then we would expect it is harder to discriminate between sets of size two and three than between sets of size one and four. But this is not what we observe, infants can distinguish between sets of size two and three but not between sets of size one and four. So, it appears that infants do not have analog number representations of small numbers (infants can discriminate between sets of size four and eight as shown in (Xu 2003) so it does appear that infants do utilize analog magnitude representations for quantities larger than three). In light of this we need to figure out what representations are utilized in successfully completing the task of retrieving all of the objects placed in the box. The predominant theory is that infants simply represent the individual objects themselves in working memory. These are parallel individuation representations. The explanation for infants' behavior is as follows. The infants see the one, two, or three objects placed in the box and represents each object placed in the box in working memory. This enables infants to recognize when each object placed in the box has been retrieved and stop searching. It is theorized three is the upper limit of how many objects can be represented at once in working memory by infants of this age. So, when an infant sees four objects placed in the box, they cannot represent each object and so do not know when to stop searching for the objects and are often satisfied with retrieving only one or two of the objects. This plausible explanation provides support for the claim that parallel individuation representations are used by humans for representing the quantity of objects in small sets.

Mental Numeral Representations

A third type of mental number representation, and the most important for my purposes, is *mental numeral representation*. These mental representations are just mental representations of numerals. There are numerals of the verbal variety, e.g. in English there is 'one', 'two, 'three', etc. We also have numerals of the solely written variety. I will go through a few examples. The unary numeral system is perhaps the simplest to understand. It represents a number with a corresponding quantity of symbols, e.g. the number one is represented by '/'. Similarly, the number three would be represented by '///' and the number 10 would be represented by '//////////', etc. To determine the number represented by a string of symbols we just count up how many symbols there are in the string. The Roman Numeral system uses strings of the symbols 'I', 'V', 'X', 'L', 'C', 'D', and 'M' to represent numbers. To determine what number a string represents we need to know what number each individual symbol represents and then the value of the string is determined by adding (or subtracting depending on the position of the symbol in the string) the numbers corresponding to each individual in the string together. The binary numeral system uses strings of the symbols '1' and '0'to represent all the numbers. This numeral system, unlike the Roman numeral system is a *positional* numeral system, which means that the position of a symbol in the string gives the power of some base value (in this case the system has base two). So, take for example '101'. There is a '1' in the second place, a '0' in the first place, and a '1' in the zeroth place. This indicates we take the second power of the base and multiply it by one, we then take the first power of the base and multiply it be zero and finally then take the zeroth power of the base and multiply it by one and add these together giving us five. Each of these numeral systems have substantial differences in how they serve to represent numbers. There are many numeral systems that can be used to represent numbers and so there are many potential mental numeral representations. Though of course not all people will have mental numeral representations and most adult humans are only familiar the verbal variety and the Hindu-Arabic numeral system.

While the existence of mental numeral representations is not in question, what role they play in numerical cognition and how they play that role is unclear. There is evidence that mental numeral representations are crucial for completing numerical tasks that require precision, such as mental arithmetic. To see why the two types of mental number representation considered above are insufficient, consider again the example of making a budget for a conference lunch. A parallel individuation representation is insufficient for calculating the budget due to the limitations of working memory. I could not represent a sufficiently large quantity (unless the conference is really small) using a parallel individuation representation. An analog magnitude representation could represent a sufficiently large quantity, but analog magnitude representations would be too imprecise to accurately represent the number of conference attendees. Furthermore, there is no apparent way these two types of representation could be used together to accomplish the task. Since human adults can accomplish such tasks there must be something other sort of representation used. One theory is that mental numeral representations are what is used. We can observe that only people who have mental numeral representations (i.e. have learned a numeral system of some kind) are able to perform precise numeric calculations. For example, people raised in cultures, such as the Pirahã culture, that have no numeral system cannot perform precise numeric tasks. This is evidence for the claim that mental numeral representations play an integral role in precise numeric calculations.

It is an open question exactly how mental numeral representations are used in performing mental arithmetic. I will go through a few possible ways that mental numeral representations could be used in performing mental arithmetic, because it is an issue that will be relevant below.

One possibility is that mental numeral representations are tied to analog magnitude representations in a way that when I go to add two numbers, the mental numeral representations activate the appropriate analog magnitude representation. Though this would presumably only be effective for small numbers (less than say 20) given the imprecise nature of analog magnitude representations. This theory is suggested by evidence that analog magnitude representations, while not sufficient for doing mental arithmetic are necessary for it. It was observed, as detailed in (Dehaene and Cohen 1997), that a couple people who had mental numeral representations but did not have analog magnitude representations (due to brain damage of the relevant area) were unable to perform simple mental arithmetic operations such as addition. A second possibility is that mental numeral representations are utilized in doing mental arithmetic similar to how a calculator does arithmetic. Electronic calculators use binary numeral representations in performing calculations, where a binary numeral is represented by an array of transistors (a transistor being on corresponds to '1' and its being off corresponds to '0'). The electronic calculator is designed in a way that when a calculation is given as an input the calculator manipulates the numeral representations using a series of logic gates to produce an output that is the solution of the calculation. Mental numeral representations could be utilized in numeric cognition in a similar way. A third possible way in which mental numeral representations could be used in mental arithmetic is that they could be used for encoding sums, products, etc. in memory. Given some input, e.g. '5 + 7', long term memory could be searched for the appropriate string of numerals. The solution, '12', could then be retrieved from the string of numerals in long term memory. Interestingly, this process does not seem to rely on the relationship between the numeral system and the set of numbers it represents.

While there are clearly important open questions remaining, I want to now move on to the issue of how it is that the theory of mental number representations just considered can be consistent with the view that numbers are abstract objects. I will argue that the vehicles of mental *numeral* representations can serve as mental *number* representations even when numbers are taken to be abstract objects. To do this, we need to make sense of how these mental numeral representation vehicles can properly be viewed as representing numbers in the relevant circumstances. Making sense of how mental numeral representation vehicles can effectively become mental number representations must be done through an account of what determines the content of a mental representation. We need an explanation of how in one circumstance, a vehicle of a mental representation can have a numeral content and in a different circumstance the same vehicle can have a number content. Not all views of how the content of mental representations is fixed allow for this possibility. For example, an account of content based on causal relations, such as that in (Dretske 1981) or that in (Fodor 1987), would not allow for this. If what determines the content of a mental representation is some causal relations between the vehicle of the representation and its content, then, a vehicle of representation could not have two distinct contents (and furthermore if numbers are abstract this sort of account of mental content would not be compatible anyway).

Ramsey on Mental Representation

One view of content that can provide us with what we need is an account given in (Ramsey 2007). Ramsey argues that any computational view of cognition is committed to two notions of representation, I/O-representation and S-representation. A computational view of cognition views cognitive processes as computations. Under such a view, specifying a cognitive process involves specifying what is being computed, and explaining how this computation is done involves giving the algorithm that carries out the computation. I/O-representation is a notion of representation that arises from giving an account of how a cognitive process is done via an algorithm. The algorithms that underlie cognitive processes will tend to be complex (not just a simple mapping of inputs to outputs). Complex algorithms can be decomposed into parts. In order to view the entire algorithm as carrying out its computation, we must interpret the inputs and outputs of the various parts as representations (Ramsey 2007, pp. 68-77).

For example, suppose we want to explain how someone chooses between two options in a decision under certainty. A simple version of the algorithm we put forward might be as follows. For each option input, a function is applied which returns the value of that option, the option-value pairs are stored memory. Then a function is applied to the values that compares them and outputs the higher value. The option associated with this value is then retrieved from memory and is output as the chosen option. We must view whatever brain states and processes instantiate this algorithm as containing representations of options and values and operating on these things. Otherwise, the explanation would not make sense. For example, we would have to hold that none of the brain states or their parts involved represent the values of the options but that the values associated with the options play a role in determining which option is chosen. Of course, it might be that decisions under certainty are computed using some other algorithm. The point is that if we explain the computation as being carried out by an algorithm involving things like values and options, then it is a requirement of our explanatory strategy that we view the inputs and outputs throughout the various parts of the process as representations.

We can consider how this notion of I/O-representation relates to the ideas for how mental numeral representations could be used in performing arithmetic computation. One of the proposed ways was that numeral representations are used to encode the appropriate relations between numeral representations of arithmetic computations in memory. So, if ‘7 + 5’ is the input, then ‘12’ will be associated in long term memory and can be retrieved and output. In this case, the algorithm does not involve representations of numbers, just representations of numerals, and so we do not need to posit that arithmetic computation involves mental representation of numbers. If the algorithm that the mind uses to carry out mental arithmetic is similar to the algorithm that a digital calculator uses, we would need to posit that it involves mental representations of numbers, because the explanation for how the algorithm works involves numbers. The states of the digital calculator do not represent numerals, but rather instantiate numerals. So similarly, some brain states or parts of them would have to themselves be numerals, which is to say representations of numbers. The important thing to note is that a representation of a numeral can itself be a numeral. For example, we could represent the Hindu-Arabic numerals with binary numerals, but binary numerals are themselves numerals.

S-representation is a notion of representation that arises from explaining how a cognitive process works by appealing to features of a posited representation vehicle or system of representation vehicles that make it effective as a representation. The idea is that representation vehicles can serve as models or simulations of what is represented. Ramsey draws an analogy with a map. A map is useful as a representation of a region when the distances between points on the map are structurally similar to the distances between the locations the points on the map represent. In explaining why the map can be used as a map, we need to posit that the points on the map represent locations (Ramsey 2007, pp. 77-92). Another apt example is numeral systems. Suppose for example that we uncover some clay tablets from a previously unknown ancient civilization and that we suspect these clay tablets to have arithmetic calculations on them. Part of the argument for why we should interpret the strings of symbols as representing arithmetic calculations will need to involve some account of how the symbol system enables one to carry out arithmetic calculations. In explaining how it is that the symbols system can be used for arithmetic computation, we need to posit that the symbols are numerals, i.e. representations of numbers. Similarly, in many computational theories of cognition we find that in explaining how some brain state or processes can effectively be used in some cognitive process requires us to view it as the vehicle of some representation. We saw this above is Dehaene's account of how a neural accumulator can serve to represent quantity.

Ramsey notes that theories of representation that hold a vehicle has content in virtue of some similarity or resemblance between the vehicle and the content has a long history in which significant objections have been raised. There are two objections commonly raised and it will be useful to see how Ramsey's account avoids them. The first objection is that similarity is a symmetric relation while representation is an asymmetric relation. So, it cannot be that something is a representation just in virtue of its being similar to what it represents. One way to motivate this objection is to just consider examples. For example, a numeral system is structurally similar to the number system that it represents and vice versa. But while we consider the numeral system to be a representation of the number system, we do not take the number system to be a representation of the numeral system. Ramsey's account does not face this objection because S-representations are representations in part in virtue of playing a functional role in a cognitive system. The S-representation relation is a three-place relation that holds between a cognitive system, a representation vehicle, and a representation content, and so the question of symmetry does not arise.

Even though the question of symmetry does not arise for Ramsey, we can try to recreate the objection. Consider the following example. If I have a blueprint of my house, presumably it would be possible to use the house as a representation of the blueprint of it, due to their structural similarity. I could make inferences say about the number of lines on the blueprint by walking through my house and counting the number of walls. It is consistent with Ramsey's view that the blueprint of my house is a representation of my house and my house is a representation of the blueprint of my house. But, my intuition in this case is that my house does indeed represent the blueprint if used as such. I suspect the reason why we may have the intuition that it is not when presented with the example originally is that typically I would not use my house as a representation of the blueprint of my house.

The second objection to the representation relation being based on similarity is that similarities abound, and any representational vehicle will be structurally similar to many things. So, unless we are willing to accept that a representation vehicle represents many things, similarity will not be able to ground representation. Ramsey's account avoids this problem because S-representations are viewed as representations because of the explanatory role the representation vehicle is playing. Suppose for example we have a system of representation vehicles that are similar to two structures, *A* and *B*. We can be justified in claiming that in a given situation the representation vehicles represent parts of structure *A* and do not represent parts of structure *B* if we are positing the representation vehicles to explain how the cognitive system carries out some cognitive process whose target is *A*. (Ramsey 2007, pp. 93-96).

Mental Number Representation

The point of introducing Ramsey's theory of mental representation is to see how the vehicles of our mental numeral representations could serve to represent numbers. We saw with I/O-representations that there are circumstances in which we would need to posit mental number representations, but not necessarily in a way where the same vehicles represent both numerals and numbers. But, the best candidates for S-representations of numbers are the vehicles of our mental numeral representations.[13] The vehicles of mental numeral representations will serve as the vehicles of S-representations of numbers given the following two conditions are satisfied. The first condition is that the mental numeral representations must be S-representations of the numeral system. Since the numeral system will be structurally similar to the number system it represents, this condition ensures that these representation vehicles are structurally similar to the relevant number system. The second condition is that the explanation for why mental numeral representations are useful for completing numeric tasks, such as mental arithmetic, is that they are structurally similar to the relevant set of numbers. The question now is whether these conditions are satisfied. As to the first condition, it seems plausible that for a subset of the natural numbers, say the first 100 natural numbers, we have some of the arithmetic structure stored in long term memory, based just on the observation that memorizing the arithmetic relationships for the first several natural numbers is something most people go through in school. Similarly, the ordering structure is plausibly stored in memory. Obviously in both cases, how much is stored in memory will depend on the person. As for the second condition, if indeed part of the arithmetic structure of the natural numbers is stored in long term memory and some mental arithmetic is done by retrieving the solution from memory, then part of the explanation for why this structure in memory is effective for doing mental arithmetic will appeal to its similarity to the arithmetic structure of the natural numbers.

While we cannot say definitively that the vehicles of our mental numeral representations are S-representations of numbers, it is at least plausible. Furthermore, these proposed mental

[13] See Giaquinto's "Cognition of Structure" section 2.4 for a proposal, based on the idea of visual category specifications, that we mentally represent the natural numbers using a sort of mental number line representation. This could be another possible S-representation of numbers.

number representations satisfy the conditions of being naturalistic, empiricist, and compatible with Platonism about numbers. The proposal is naturalistic in that it is based on an account of mental representation implicit in a computational view of cognition and consequently it is consistent with accepted views in cognitive science. It is empiricist in that the posited S-representations of numbers are not innate but acquired. It is consistent with Platonism in that if the explanations of the cognitive processes are correct and numbers are abstract objects, it really will be the case that the mind has mental representations of abstract objects.

It is worth briefly contrasting this proposal with that found in (de Cruz 2016). Helen de Cruz aims to show that debunking style objections to mathematical realism fail. In the course of doing this, she gives a proposal for how a version of mathematical realism, in particular an ante rem structural realism like that advocated for by Stewart Shapiro, is compatible with the consensus view of mental number representation. Her proposal is roughly as follows.[14] Children learn to associate the first few counting numerals with the parallel individuation representations of the first three numbers. From doing this, children are essentially able to learn the successor function and in doing so make a one-to-one association between larger numerals and larger cardinalities. The analog magnitude representations supply the needed mental representations of cardinalities greater than three.

---

[14] The part of her essay I am describing is quoted in full here: "How then do humans learn to represent a natural number like 54? Through their object-file system, young children have an innate capacity to distinguish 1-patterns, 2-patterns, and 3-patterns. As they learn to count, they realize that these patterns correspond to the linguistic utterances 'one', 'two', and 'three'. Remarkably, preschoolers always learn the numbers 1, 2, and 3 in that order (i. e., children learn that 'one' represents a unique cardinal value, then 'two', then only 'three'). While one would expect the next step is four, children make a crucial induction: they make an analogy between next in the numeral list and next in series of object-files: if n is followed by n + 1 in the counting sequence, adding an individual to a set with cardinal value n results in a set with cardinal value n + 1. Children generalize this to higher magnitudes, which helps them to understand the successor function (Sarnecka in press). Next to object-files, the approximate-magnitude system continues to play a critical role in arithmetical skills in adults, as it helps them to gain semantic access to symbolic representations of numerosities > 3" (de Cruz 2016, p. 10).

One contrast between the proposal given here and de Cruz's work is that de Cruz takes a developmental perspective by focusing on how children learn to represent numbers. My approach is focused on adults that have a thorough familiarity with a numeral system and an ability to do mental arithmetic. Since I argue for the existence of mental number representations based in part on the ability to do mental arithmetic, it leaves the question of whether young children who cannot do mental arithmetic have mental representations of numbers unanswered. While I do not take this to be problematic, clearly it is a question worth investigating because it may be controversial if it came out on my view that children did not have mental representations of numbers. Another difference is that, whereas I have suggested mental numeral representation might serve as S-representations of numbers, on de Cruz's account, while mental numeral representation play a crucial role, it is the parallel individuation and analog magnitude representations that represent numbers. One concern with this approach, raised in (Rips 2008), is that analog magnitude representations are not precise enough to represent numbers. My proposal has the advantage of not tying the semantics for mental numeral representations to analog magnitude representations.

While we have some idea of what the mental representations of numbers are like given Ramsey's account of mental representation and the research that has been done on numerical cognition, ultimately I would like to determine what the mental representations for all mental representations of mathematical objects, operations, properties, and relations are like. There are some difficulties in figuring this out though. In the case of numerical cognition, there are several cognitive processing that likely involve mental representations of numbers, and for at least some of them, e.g. mental arithmetic, we have some idea of what computational processes might underlie them. This is not the case when we consider other mathematical objects, operations, properties, and relations. For example, consider a mathematical operation such as addition. It is not clear the mental process of adding two numbers requires an explicit representation of addition. Answering questions about addition such as, is addition a binary operation, does seem like it would require a representation of addition. But this raises the question of what the computational process that underlies answering this and similar questions is like. Another issue is that, while in the case of numbers there is a clear candidate for what the vehicles of mental S-representations of number are, it is less clear what the vehicles for the S-representations of mathematical operations, properties, and relations might be. It is not obvious how the symbol '+' or the word 'addition' could serve as an S-representation of addition. The rest of this chapter will be aimed at solving this latter issue about what the vehicles of S-representation might be. The answer I will propose is based on LASS and so we will turn to LASS now.

The Language and Situated Simulation Theory

In this section I will explain what LASS is and consider some of the motivation for it. LASS is a theory that essentially brings three claims together. The first is that a conceptual system is central to cognition. [15] The second is that (most if not all) mental representations are modal.[16] The third is that there are two functionally distinct cognitive systems, one that is used in

---

[15] I will make use of the term 'system' throughout this section and to avoid confusion I will clarify my usage of this term. A 'system' will denote something that consists of representations and processes involving those representations that serves some purpose. What may be confusing is that I will pre-fix the term 'system' in various ways. Some pre-fixes will be used to distinguish systems by the representations (and/or processes) that in part make up that system. Other pre-fixes will be used to distinguish systems by their purpose (and one system could serve more than one purpose and so the same system could be referred to by multiple prefixes). I will mostly leave it up to the reader to figure out what how the pre-fix is acting to determine the reference.

[16] The hedging of this claim to most is due to two factors. The first is that, as we will see below, it is not clear that the representations used by the language system are modal. The second is that the proponents of LASS state that they are open to the possibility that amodal representations are used by the cognitive system (Barsalou et al 2008, p. 245).

tasks dealing with language and the other used in tasks dealing with perceptual information. This claim is a version of what is commonly known as the *dual-coding theory*. To understand LASS, we need to understand these claims and how they fit together. I will go through each in turn, give some of the motivation for them, and then explain how they fit together.

The Conceptual System

A *conceptual system* is a system in which categories and the relations between them are represented. So, the claim that concepts are central to cognition is just the claim that category representations are central to cognition (a claim shared by many other theories of cognition). Generally speaking, concepts are theorized to be central to many cognitive processes. The motivation for such a view comes from a few different considerations. In particular, I will point to three different things that concepts get used to explain. The first is that we seem to be capable of having an unbounded number of thoughts. This points to the existence of some base set of mental representations from which we can generate others. Related to this is the apparent compositionality of thought. Just like different sentences may be composed of the same words it seems that many thoughts are made up of the same constituents. Concepts are posited to be what thoughts are composed of. The second thing concepts are used to explain is our ability to bring information learned from past experiences to bear on the present. For example, when I recognize an object in front of me as a dog I am able to use my past experiences with dogs (and information acquired in other ways such as by testimony) to interact with the dog, discuss the dog with others, etc. The ability to categorize the dog as a dog and access stored information about the category of dogs is to be explained by concepts. Given a computational account of cognitive processes, whatever mental representation is central to my ability to categorize the dog is a concept (or at least part of the concept). The third thing concepts are used to explain is language use and comprehension. The thought is that in order to understand, e.g. a sentence, my cognitive system builds up a representation of the proposition represented by the sentence out of concepts. A theory of concepts can be judged in part on how well it can account for these three things. We will see briefly how LASS does this below.

Mental Representation is Modal

The distinction between modal and amodal representation is a distinction about the form (as opposed to the content) of the representation. What makes a representation modal is that the form of the representation is particular to a specific perceptual system such as the visual system or auditory system. An analogy will help to make distinction clearer. Consider for example the way in which digital computers encode visual information and auditory information. A digital computer essentially uses a string of binary numbers to represent both visual and auditory information. But importantly the information is encoded differently as it corresponds to the structure of the input. Roughly speaking, to represent 2D visual information, a computer will use an array of binary numerals where each numeral represents the color of the pixel in that location. To represent audio information, a computer will use a series of binary numerals where each numeral represents the amplitude of the sound wave. Even though binary numerals make up each representation, the representations as a whole will have different structural properties that make them structurally similar to what they represent. Even though auditory information and visual information will ultimately be represented by neural states, these neural states will encode the information in a particular way, and this makes them modal representations.

There are three positions to consider: there are only modal representations (which is roughly the position of LASS), there are only amodal representations, or there are both modal and amodal representations. In order to convey what LASS is committed to, I will consider these positions and some of the evidence for them following the discussion of these issues in (Prinz 2002). Molyneux's problem is useful to consider here. The problem is to determine whether someone blind since birth that has tactile experience of a sphere would recognize a sphere visually as a sphere if they were to gain the ability to see. If the somatosensory system uses the same sort of representation as the visual system, then it would seem to follow that a comparison could be made between the representation resulting from touching the sphere and looking at the sphere. This would enable the person to recognize visually that it is the same object they had experienced tacitly. If we assume that a comparison of the representations (presumably requiring a comparison of the vehicles of representation) is necessary to answer in the affirmative to Molyneux's problem, then this affirmative solution to Molyneux's problem would appear to imply the existence of amodal representations.

The following empirical observation seems to be evidence that we should answer affirmatively to Molyneux's problem (and so appears to be evidence that amodal representations exist). Suppose that we give an infant a pacifier with a tip that has a particular shape and we then show the infant pictures of various shapes, one of which matches the shape of the tip of the pacifier we gave them. What we would see is that the infant will stare longer at the picture of the shape of the tip of the pacifier they were given as compared to the pictures of the other shapes. (Kaye and Bower 1994) shows infants that are just 12 hours old display this behavior. This seems to be evidence that the infant recognizes the that the picture is of the same shape as the tip of the pacifier that they experienced somatosensorily. This in turn seems to indicate that the tactile representation of the shape of the tip of the pacifier is comparable to the visual representation of the same shape. That they are able to be compared seems to show that the same sort of representation is generated by each experience.

Does this show that only amodal representations exist? Only if it a direct comparison of representations is necessary, but there is an alternative explanation. We can suppose that there is a modal tactile representation of the shape and a modal visual representation of the shape but also an amodal representation of the shape which would be activated by both modal representations. This would enable the infants to recognize that the shapes are the same in each case. Indeed, there is good evidence indicating that there are modal representations and so the view that there are both modal and amodal representations is better supported than the view that there are just amodal representations.

One consideration in support of the existence of modal representations, found in (Kellman and Arterberry 1998), stems from the above observation of infant behavior. We saw when discussing the evidence for analog magnitude representations that infants will stare longer at novel stimuli. In the case of shape above, if there are only amodal representations then it would seem to follow that the picture of the shape would not be a novel stimulus and our observation that the infant stares longer at the picture of the shape corresponding to the shape of the tip of the pacifier they interacted with somatosensorily requires an explanation. One explanation is that there are modal representations, in which case the picture of the shape would be a novel stimulus.

We could even explain these observations of infant behavior without positing the existence of amodal representations as shown in (Prinz 2002). Consider the analogy between different forms of mental representations and different languages again. We can translate between the various languages directly (as opposed to translating them through some third sort of representation). Prinz suggests that it is possible that there may be mental processes that in effect translate visual representations into somatosensory representations. This would enable us to explain the above observations without requiring us to posit any amodal representations. A proponent of LASS would presumably need to adopt this strategy.

It is noteworthy that Dehaene claims analog magnitude representations are amodal representations of number and so here we have a case in which a proponent of LASS would use Prinz's strategy. Dehaene writes, "The simplest explanation is that the child really perceives numbers rather than auditory patterns or geometrical configurations of objects. The very same representation of number '3' seems to fire in its brain whether it sees three objects or hears three sounds. This internal, abstract and amodal representation enables the child to notice the correspondence between the number of objects on one slide and the number of sounds that are simultaneously heard" (Dehaene 1997, p. 52). Why does Dehaene think analog magnitude representations are amodal? Consider the following observation. As detailed in (Starkey, Spelke, and Gelman 1990), infants were presented with two arrays of dots after hearing a sequence of tones. The number of dots in one of the arrays was the same as the number of tones in the sequence. The infants stared longer at the array of dots that had the matching number. Dehaene takes this to be evidence that the representation is amodal. But this seems to be essentially the same sort of observation made above. So, as for the above observation, we can explain these observations as resulting from some sort of translation between modal representations of number. There does not need to be some third amodal type of representation into which the modal representations are both translated. So, LASS can point to this alternative explanation as a way of arguing that it is at least plausible that analog magnitude representations are modal representations.

We have seen that proponents of LASS can account for observations that seem to point to the existence of amodal representations. But there is still the question of why LASS is committed to there being only (or at least mostly just) modal representations. The primary motivation is to give a parsimonious account of cognition. As we will see, the account of cognition we get from LASS is based on the perceptual system. The way in which LASS is more parsimonious than some other theories is that instead of positing the conceptual system as distinct from the perceptual system, LASS claims the perceptual system is a conceptual system. In (Barsalou 1999), Barsalou argues that, theoretically speaking, the perceptual system can function as the conceptual system. There is also observational evidence that the areas of the brain associated with perception are active during conceptual tasks. While there are other possible explanations for this observation, the explanation that these brain areas are active because they are executing the conceptual task has the advantage of being more parsimonious in that we can explain both how perception and cognition works by just posting one system. There is some sense in which this would be a more simple and elegant design.

Dual-coding Theory

LASS gets its name from the two functionally distinct cognitive systems it posits, the situated simulation system and the language system. The language system handles some cognitive tasks involving language and the situated simulation system does everything else. Since we will see the details of how these systems work below, I will just briefly mention the motivation for dual-coding theory. There are two sources of evidence for dual-coding theory. There is neurophysiological evidence for dual-coding theory. Roughly the idea is that during different sorts of tasks, different areas of the brain are active, which points to distinct functional systems. Along with this, there is psychological evidence that is based on response times in various tasks. As dual-coding theory makes predictions about how certain tasks will be done, we can make predictions about the relative time tasks should take based on what the tasks are like, e.g. the form of the input or output may vary. Observations are consistent with the predictions of dual-coding theory. See (Paivio 1990) for an extensive treatment of dual-coding theory.

Positing two systems here instead of one may seem to run counter to the motivation of parsimony mentioned above. But there are two points to consider. The first is just that judging parsimony in terms of number of systems is difficult because systems are distinguished in various ways. The second point to consider is that there is a tradeoff here in terms of lessening the cognitive processes we need to posit. Since information is entering in both image-like and language-like formats, if we posit a single system to handle both then we need to posit processes which convert the information from one format to another, whereas if we just posit one system for each type of format then we do not need to posit translation-type processes.

How These Claims Fit Together

The claim that (most of if not all) mental representation are modal serves as a constraint on what representations we can use to explain how cognition works. So, it must be that the concepts in some way consist of perceptual representations. The dual-coding claim fits with the claim that there is a conceptual system as follows. According to LASS, the situated simulation system is the conceptual system. So now we need to know what the relationship between the language system and the conceptual system is. The language system to some extent operates independently of the conceptual system, but the importance it has in the context of understanding cognition is the way it interacts with the conceptual system. The language system activates concepts in a manner that mirrors the structure of linguistic representations. For example, upon hearing a sentence a particular combination of concepts will be activated, and in this sense the language system acts as a sort of control mechanism for the conceptual system. The language system, as we will see, also stores information about word associations and so hearing, seeing, or touching words will activate concepts related through word associations as well. I will go into more detail about these functions below.

Putting these ideas together we get the following statement of LASS:

(1) All cognition is accomplished by two systems, the language system and the situated simulation system,

(2) The situated simulation system is a conceptual system

(3) The conceptual system consists of representations generated through perception

(4) A distinct language system adds to the power of the conceptual system

The Cognitive System According to LASS

This section will go through the details of how the situated simulation system and language system work and interact. I will then go through an argument that the situated simulation system is a conceptual system.

The Situated Simulation System

According to LASS, the situated simulation system uses representations of perceptual information to make a conceptual system. Lawrence Barsalou describes how the mechanisms underlying the processing and storage of perceptual information can function as a conceptual system in (Barsalou 1999) and I will give a brief recapitulation of his ideas here. To understand how the perceptual system can represent categories I will focus mostly on the visual system but things work analogously for the other senses. Based on our common scientific understanding of visual perception, vision works roughly as follows. The visual system transduces physical signals into nervous signals which go through extensive processing from the retina to the visual cortex. Since visual information goes through extensive processing there are actually many different stages where information is represented in different ways. For example, it has been demonstrated in research on vision that there are neurons and groups of neurons that serve as feature detectors which detect visual features such as edges. These feature detectors represent perceptual information and eventually in the visual cortex some pattern of neural activation serves to represent the visual information being taken in. Similarly, in other parts of cerebral cortex neural patterns serve to represent perceptual information from other senses.

Barsalou calls the patterns of neural activity that represent perceptual information (in the cerebral cortex) *perceptual symbols.* Barsalou then postulates the existence of various neural mechanisms which utilize these perceptual symbols to represent categories and enable us to categorize. One of these mechanisms serves to select portions of the overall perceptual representation for further processing (like identifying what is represented) and storage. Barsalou suggests that this mechanism is just the system that governs attention. For example, if I perceive a car driving down a road in the rain, then attention might select signals that represent the tire of the car for perhaps some further processing and storage. Furthermore, these perceptual symbols may be distributed across various sensory modalities. So, a perceptual symbol could be constituted by a neural state that represents both visual and auditory information such as the appearance and sound of car driving past you.

According to Barsalou, the second mechanism that is involved in creating category representations is a mechanism that aggregates perceptual symbols that represent similar things and enables the reactivation of perceptual symbols for use in various cognitive processes. In the association areas of the cortex, neurons capture the activation patterns of perceptual symbols and store the patterns. But perceptual symbols are not stored in memory independently of one another. For example, after attention selects my representation of the tire to be stored in memory it is stored in association with other perceptual symbols of tires. This gives rise to what Barsalou calls a *simulator*. These patterns can then be reactivated (at least partially) which Barsalou calls a *simulation*. Simulators, as I discuss below, are what play the role of concepts according to LASS and the simulator's ability to create simulations is what enables them to play a role in categorization.

The Language System

The language system is involved in completing cognitive tasks involving language. Some tasks it completes on its own while others are completed by the language system in conjunction with the situated simulation system. The language system uses representations of linguistic information to complete some language-based tasks independently. The linguistic information being referred to includes things like the form of letters, words, and sentences, as well as more complex information about language, which for example could include statistical information about the distribution of words in texts that an individual has read. What is not included by the term ‘linguistic information’ here is information about the meaning of words. Any tasks that require information about what words mean are completed by the language system in conjunction with the situated simulation system. According to LASS when an instance of a word or sentence is perceived it is categorized and then associated words or sentences in memory are activated. Depending on the task, if a semantically superficial linguistic processing is sufficient to complete the task then this is carried out and if there is no such strategy then simulations serve to provide the information for further processing to complete the task.

To understand how the language system could complete linguistic tasks without relying on semantic information it is useful to consider Searle’s Chinese Room thought experiment. In this thought experiment we imagine that someone is in a room and is passed slips of paper with strings of Chinese symbols through the door. The strings of Chinese symbols are meaningful questions that someone fluent in Chinese would be able to read and comprehend. The person in the room does not understand the Chinese language. But the person in the room does have books of strings of symbols organized such that, if the person matches the string of symbols given to them to a string in the book and then writes down the corresponding string of symbols, they will have a meaningful and appropriate response to the question. Instead of a book that simply has all these pairs of strings of symbols written out we could imagine there is a computer program that uses statistical information about the distribution of words and sentences in numerous Chinese texts to construct a string of symbols in response. Effectively we have updated the Chinese room metaphor of a book to reflect recent research in natural language processing. For example, latent semantic analysis (LSA) is a method for analyzing and representing the meaning of words using just statistical information about the co-occurrence of words in large numbers of texts. LSA representations can be used to perform various tasks, for example finding synonyms. LASS does not claim that LSA representations are the linguistic representations actually used in cognition, but LSA does demonstrate the ability for linguistic representations of some sort to be sufficient for the performance of some linguistic tasks. LASS leaves it open what sort of linguistic information might be represented to accommodate whatever future research will indicate. But the language system posited by LASS essentially operates like a sophisticated natural language processing computer program.

Interaction Between These Two Systems of Representation

LASS posits that in general the linguistic and perceptual representation systems interact extensively during many cognitive tasks. Words and sentences are assumed to be associated with perceptual symbols. The categorization of a word can lead to the activation of an associated perceptual symbol and similarly the categorization of a perceived object, property, or relation can lead to the activation of associated linguistic representations. So, for example if I am making plans with some friends for the next day, this may give rise to simulations in the perceptual system of what I want to do which then get verbalized through the linguistic system. Alternatively, my friend might make a suggestion for what we should do using language which upon my hearing the words might give rise to a simulation of the suggested scenario.

Concepts Under LASS

As stated above, LASS takes simulators to play the role of concepts. In this section I will review Barsalou's argument that concepts are simulators. To determine the plausibility of whether simulators are concepts we need to assess whether simulators can account for the sort of cognitive processes that concepts are supposed to account for. So, the argument essentially consists of demonstrating that simulators can be used to account for our ability to perform the relevant cognitive tasks.

Concepts are primarily used to account for categorization. So first I will explain how simulators are supposed to enable categorization. Barsalou proposes the following. When someone perceives an object to be categorized, we will have some perceptual symbol generated by perceiving the object. Various simulators create simulations and a comparison is made between the representation of the object to be categorized and the various simulations. The simulator that generated the most similar representation is identified and this serves to indicate what category the object belongs to. For example, suppose I am in front of a poodle and someone asks me what kind of animal it is. My dog simulator would then create a simulation of the poodle (which would be more similar to my representation of the object I am experiencing right now than any of the other simulations generated) and this would in turn activate the associated word in the linguistic system which would enable me to report that it is a dog. Obviously, many details about the process of categorization are left unspecified. The claim is simply that theoretically it is plausible that simulators could in some way, like the way just outlined, enable us to categorize things.

One of the important differences between this proposal and some other prominent theories of category representation is the flexibility of a simulator to create a large variety of simulations that have not necessarily been experienced. This flexibility also helps to solve the problem of the productivity of the conceptual system, i.e. how to account for the ability of the conceptual system to create concepts of things not experienced (or imaginary). Simulators are productive as they can combine to produce many novel simulations. Simulators can also account for knowledge effects, unlike for example the exemplar and prototype theories of category representation. See (Margolis and Laurence 1999) or (Murphy 2004) for a discussion of issues pertaining to concepts and categorization. The point is, not only can LASS account for categorization but can account for categorization in some ways better than other theories.

As I pointed out in the background section on concepts, concepts are also used to account for language comprehension. Since simulators in the perceptual system are hypothesized to be associated with words and sentences in the linguistic system. We can explain language comprehension as working by activating simulations corresponding to the perceived words and sentences. As for the compositionality of thought, simulators work together to create simulations so while simulators do not themselves combine to compose thought, there is a clear sense in which they do compose complex thoughts. While these are not detailed explanations, given that we can plausibly use simulators to account for the cognitive process that concepts are supposed to account for, within the context of LASS it makes sense to consider simulators concepts.

Abstract Concepts Under LASS

In discussing the situated simulation system, Barsalou shows how abstract objects, etc. can be represented by perceptual symbols. The main insight is that abstract objects, etc. are represented by perceptual symbols of introspective states. I will go through one of the examples he gives and discuss the general method he employs for determining what perceptual symbols represent a given abstract object, which he claims should work for any abstract object. Since mathematical objects, etc. are abstract the method he gives should work for them.

Barsalou's claim that we have perceptual symbols of introspective states may seem odd because it is unusual to view introspection as a sort of perception. Whether or not this is a legitimate use of the term 'perception' is inconsequential though. Barsalou is just claiming that the same mechanism which enable us to represent categories through the sensory system also enable us to represent categories of introspective states. Barsalou says that there are at least three important types of introspective states: representational states, cognitive operations, and emotional states. To see how the same mechanisms suffice to represent introspective states, consider as a case the emotional state of anxiety. Barsalou is proposing that the neural activity that underlies the emotional state could serve as a perceptual symbol of the feeling of anxiety if the attentional system isolates the appropriate neural activity. Over time experiences of anxiety would generate a simulator for anxiety.

Barsalou gives the following example of the concept truth. One caveat he gives is that terms denoting abstract objects are highly polysemous and so he points out that he is trying only to show how a particular conception of truth can be represented using perceptual symbols. The conception of truth that he wants to analyze is truth in the sense of a representation matching reality as it is directly perceived, something like truth as verification. Barsalou describes a sequence of events. First a person constructs a simulation. That person then attends to the current perceptual state. Finally, the person attempts to map the simulation onto the current perceptual state and assesses introspectively whether the mapping was successful or not. Over time the perception of success resulting from this sort of sequence of events results in a simulator that represents this sense of truth as representation matching reality.

In this example we can see the following strategy is employed for showing how an abstract concept is represented with perceptual symbols. First, given the concept, we need to identify the sequence of events that frame the concept. In other words, we need to find a sequence of event-types that we can point to as the things that over time will give rise to a simulator. Second, we need to determine what the relevant information is in this sequence of events that will be represented. As mentioned above, Barsalou claims that this will always involve some representation of introspective information. Third, we need to specify what precisely the content is that the abstract concept refers to. In the case of truth above truth is just the success of the matching.

Why does Barsalou think that abstract objects need to be represented by perceptual symbols of introspective states? The argument that he gives basically eliminates the other possibilities. Since abstract objects and properties are non-spatiotemporal it seems to follow that perceptual symbols that result from our perception of the external world could not be representations of them. Similarly as the linguistic representations are representations of information about language they are not representations of abstract objects and properties (perhaps insofar as information and language are abstract we might take these to be representations of something abstract, but this would still not get us all abstract objects and properties). So, the only option left for LASS is to hold that perceptual symbols of introspective states in some way represent abstract objects and properties.

It is notable that the mental representations of number that are consistent with numbers being abstract objects I discussed above do not involve perceptual symbols of introspective states. So, my account of mental number representation conflicts with the claim that abstract objects are represented by perceptual symbols of abstract states. But I think it is clear that Barsalou's elimination argument fails if we adopt Ramsey's account of mental representation. As I argued above, Ramsey's theory of representation content is consistent with abstract objects being represented by what we can view as perceptual symbols of numerals. So, while it may be the case that in some cases representations of abstract objects, etc. include perceptual symbols of introspective states, this need not always be the case. But as we will see below, I think Barsalou's idea can give us a solution to the problem of determining what the vehicles of representation for mathematical operations could be.

LASS and Representations of Operation, Properties, and Relations

In this section, I will consider what the vehicles of S-representations of mathematical operations, properties, and relations could be in the context of LASS. I will focus just on S-representations because arguing for the existence of I/O representations would require specifying the algorithm that underlies a relevant cognitive process and I am not in a position to do this. Also, the ideas in LASS are relevant to the notion of S-representation. There are two ways we can view perceptual symbols as S-representations. The first way is to consider what the vehicles of the perceptual symbols are like. As we saw above, LASS claims that perceptual symbols are modal. Modal representations, due to the way in which they are encoded, are structurally similar to what they represent. Consider again the analogy with how digital computers encode visual information. The string of binary numerals that represents a 2D image is structurally similar to the image it represents. Similarly, perceptual symbols are structurally similar to what they represent. The second way is that the cognitive system uses perceptual symbols as a stand in for what they represent in simulations. So, if I am trying to plan a route to get from my house to my office, my simulations of various routes serve to represent the different possible courses of action I am considering. The goal below will be to show that there are perceptual symbols that can serve as S-representations of mathematical operations, properties, and relations. The section will be laid out as follows. First, I will consider mathematical operations, then I will consider mathematical properties, and lastly, I will consider mathematical relations. For each I will go through an example or two.

Operations

As we saw above, it is unclear what the vehicles of representation are that can serve as S-representations of operations such as addition. Certainly, the symbol '+' is not structurally similar to addition and it is not clear how the symbol could serve as a model or simulation of addition. One possibility worth considering is to think about the ways in which operations are represented externally by mathematicians. Operations can be treated as particular type of relation and relations can be represented as sets of ordered tuples. It does seem possible that we could have an S-representation of an operation consisting of perceptual symbols that is structurally similar to such a set. We could then appeal to the structural properties of this mental representation in explaining how the mind comes up with answers to questions about addition. While I think this is a possibility worth exploring, one issue is that it is seems unlikely that non-mathematicians would acquire such a representation and yet plausibly non-mathematicians have a representation of addition.

The way in which LASS tries to account for abstract concepts provides another promising possibility. LASS claims that our representations of abstract things involve perceptual symbols of introspective states. Using this idea, we can ask whether there are perceptual symbols of any introspective states that could serve as an S-representation of addition. Consider the perceptual symbols that could arise from the process of doing mental addition. When I add two numbers "in my head" there is some pattern of neural activity that, if we move to a cognitive level of description, constitutes doing mental addition. Perceptual symbol of this neural process could serve as an S-representation for addition. Furthermore, using the idea of simulator that we get from LASS. We can imagine that these perceptual symbols might be aggregated into a simulator for addition, giving us the concept of addition.

We now need to evaluate whether the mind could use such perceptual symbols as S-representations of addition in cognitive tasks. This is somewhat difficult to do given it is unclear what cognitive processes would need to make use of a representation of addition. A plausible candidate is comprehending statements involving the term 'addition' such as in the statement 'addition is commutative'. Young adults (if not children) can presumably understand this statement. Since a conceptual approach to cognition typically uses concepts to give an account of language comprehension, we can ask, could the simulator proposed for addition be used in such a way? What I think we can confidently say here is that it seems as plausible that the addition simulator could be used in comprehending the statement 'addition is commutative' as the dog simulator could be used in comprehending the statement 'the dog is hungry'. So essentially, if LASS is correct then it seems highly plausible that the posited addition simulator would be a mental representation of addition. Furthermore, there is nothing in the above account specific to addition so presumably a similar account can be given for other mathematical operations for example composition of functions, rotations and other geometric operations, set operations, etc.

Properties

One possibility for the mental representation of properties is to treat them in a similar way to that of operations. The idea is that part of learning about a mathematical property involves learning, in addition to what the conditions under which something has the property are, how to check if these conditions are satisfied by something. So, if we consider a property like prime-ness, it seems that the property could be represented by a perceptual symbol of the mental process of verifying a number is prime. An issue with this approach is that there are relatively few instances in which someone confirms a number is prime. While there are not many details on the conditions under which a simulator forms, presumably it takes many perceptual symbols to form a simulator. This is of course not to say that these perceptual symbols cannot play any role in representing prime-ness.

A second possibility is that we could view the concept of prime-ness as in a way consisting of its definition. In this case we would be departing from LASS to some extent. We would not need to deny that simulators are concepts, just that concepts are only simulators. The classical view that complex concepts are composed of simple concepts in a definition-like way seems appropriate here. We could posit that concepts are sometimes simulators and sometimes composed of multiple simulators in a definition-like way. Furthermore, drawing on the idea from LASS that language can serve as a control for the simulation system, the natural language definition of a property, such as prime-ness, should play an important role in our ability to represent prime-ness. While I think this proposal has some intuitive appeal, one drawback is that it does not seem like we really have an S-representation of prime-ness but rather an S-representation of its definition. We might find though that while there are cognitive processes that are best accounted for by positing a representation of the definition of prime-ness rather that a representation of prime-ness itself. Perhaps the cognitive system has no need to represent prime-ness itself. Clearly more work needs to be done on determining what cognitive processes might use a representation of prime-ness or the definition of prime-ness and how these cognitive processes might work.

Relations

In order to see how perceptual symbols can represent relations we need to focus on the functional aspect of S-representations. A perceptual symbol of objects that instantiate a relation can serve to represent the relation. For example, consider the relation in-between. When I observe three objects in a row in front of me, I can consciously note that one is in-between the others. My ability to recognize this, thereby categorizing the relation, indicates a simulator for the relation in-between according to LASS. Even though it is a bit unclear how such a simulator arises, presumably it consists of perceptual symbols of groups of objects where one is in-between the others. So, these perceptual symbols serve to represent the relation in-between. We can give a similar account of the relation less-than that holds between two quantities. If I have a box in front of me with two apples in it and another box in front of me with three apples in it I can note that the box with two apples in it has less apples in it than the box with three apples in it. This would seem to indicate a simulator for the relationship less-than consisting of perceptual symbols of two quantities where one is less than the other. Of course, the less-than relation that holds between two quantities is not the same as the less-than relation that orders the natural numbers where natural numbers are taken to be abstract objects. But here we can appeal to the structural similarity between the two relations. Given the structural similarity, the vehicles of the relation less-than, that holds between quantities, could serve as vehicles of the relation less-than, that holds between numbers. Of course, it remains to be understood what cognitive processes these representations would be used in and whether, given the cognitive processes, they can contribute in an appropriate way to an explanation of how the cognitive process is accomplished. The point here is just that these perceptual symbols and simulator seem like a good candidate as representations of this relation and something to keep in mind as we investigate the issue further.

Conclusion

In this chapter, I have worked to provide the beginning of an account of the mental representation of mathematical objects, operations, properties, and relations that is consistent with Platonism, naturalism, and empiricism. Clearly the ideas presented here are in need of further development. But, importantly, I take them to be worthy of further development and I hope I have provided a compelling case that they are. In order to develop these ideas further, more research needs to be done about what cognitive processes the mathematical concepts are being utilized for. Once we have a better grasp on what role a particular concept plays in doing mathematics, it will be clearer what potential algorithms could underlie these cognitive processes and in turn what the mental representations could be like.

In the previous chapter I discussed why external representations are a topic particularly relevant to an interdisciplinary approach to the philosophy and cognitive science of mathematics. As we have seen in the account given above, external representations play a significant role in the mental representation of mathematics. External representations and interactions with them are theorized to be appropriated by the cognitive system in some cases, such as mental number representation, to serve as internal representations of mathematical objects, etc. If external representations are integral to mathematical cognition, then it raises the question of how they can be best designed to work with the cognitive system. I will come back to this point in chapter four.

CHAPTER 3: THE PSYCHOLOGY OF DEDUCTION AND EXPLANATORY UNDERSTANDING IN MATHEMATICS

In this chapter, I want to begin working on the project of determining how the goal of gaining mathematical understanding relates to the cognitive system, as described from the perspective of cognitive psychology (as opposed to folk psychology). The motivation for this project is, both, to use what we learn about mathematical understanding to inform mathematical practices aimed at gaining understanding, and, to use what we can observe about mathematical understanding and mathematical practice to inform our theorizing about mathematical cognition.

To get a sense of what this project involves, let us consider what the project would involve if it were carried out for the goal of gaining mathematical knowledge, as it is a somewhat simpler case. Mathematicians achieve the goal of gaining mathematical knowledge, psychologically speaking, by forming the appropriate (i.e. true and justified) mathematical beliefs. Furthermore, there are many different things they do to form these beliefs, for example, learn or create proofs, read textbooks, attend lectures and conferences, work with diagrams, perform computations, etc. Giving an account of how the goal of gaining mathematical knowledge relates to the cognitive system requires figuring out how to put belief and the processes that generate the appropriate mathematical beliefs in terms of some psychological theory. Cognitive psychology is of course an active field with several competing theories of human cognition, there are different ways of understanding what belief is and how the various activities done by mathematicians generate beliefs, so it would be important to take stock of the various approaches.[17]

[17] It would be impossible to carry out this project for some psychological theories because beliefs are eliminated in some theories of human cognition, e.g. accounts of cognition on which there are no mental representations.

Carrying out this project for understanding rather than knowledge will be similar but is more complicated for a few reasons. One reason is that it is much less clear what the psychological component of understanding consists in, and, whatever it is, it may well be more complex than belief. Related to this issue is the general lack of consensus on what understanding consists in, not just psychologically, but also epistemically. Furthermore, there are various types of understanding and so it is important to be careful in distinguishing between them and to be clear which type of understanding is being considered (arguably there are also types of knowledge, but propositional knowledge gets by far the most attention).

This chapter will begin the work involved in this project by focusing in on one type of understanding and just one activity in mathematics that generates it. Specifically, I am going to focus on explanatory understanding and how the activity of learning proofs generates explanatory understanding. I am choosing to focus on explanatory understanding and how learning proofs generates it, because, as I will argue, observations about which proofs are effective at generating explanatory understanding can be used in evaluating the psychological theories of deduction. There are two cognitive psychological accounts of deduction I will consider, specifically the mental models account and the mental logic account. After arguing that the observations are relevant, I will raise concerns for both theories based on the observational evidence.

The chapter will proceed as follows. The first section of this chapter will aim to get clear on what *explanatory understanding in mathematics* (henceforth EUM) is. I will only be giving a rough account of the psychological component of explanatory understanding. The focus from the psychological side of things will be on how the process of learning proofs generates explanatory understanding. The second section will introduce the mental logic and the mental models accounts of deduction. The third section will consider what implications each theory has for gaining explanatory understanding through learning proofs. The fourth section will consider some observations about proofs that generate explanatory understanding and show that these observations raise difficulties for both the mental models and the mental logic theories of deduction.

Before getting underway I want to make one terminological note. Mainly I will be using the term ‘deduction’ to refer to a mental process. But there is also a normative sense of the term ‘deduction’ that relates to the notion of a deductively valid argument. I will be sure to make it clear if I am using the term ‘deduction’ in the normative sense.

Explanatory Understanding in Mathematics

In this section, I will argue for two claims. The first claim is that some proofs are explanations. I do not take this to be very controversial, so I will just go through some of the arguments for this claim given in the literature on mathematical explanation. The second claim I will argue for is that having deduced the relevant conclusion from the relevant premises is a necessary condition for having explanatory understanding of why a given mathematical theorem obtains. Before I argue for these claims, I will give some context by saying a bit about what explanatory understanding is. I will consider two different approaches to giving a theory of explanatory understanding found in the literature. One approach takes understanding to be fundamental and defines explanation in terms of understanding, while the other approach takes explanation to be fundamental and defines understanding in terms of explanation. I will not be arguing for one approach over the other here. The approach that takes explanation to be fundamental is the predominant approach taken in the literature on understanding and explanation. Neither of my arguments depend on one of these approaches being the correct approach. The difference in approaches is relevant for what sort of arguments make sense on matters such as whether some proofs are explanations.

An Explanation-based Approach to EUM

An explanation-based approach to explanatory understanding holds that a person has explanatory understanding when they bear a certain psychological relationship to an explanation. So, giving an account of explanatory understanding involves specifying what psychological relationship must hold between a person and an explanation for them to have explanatory understanding (though some of the details may depend on what explanations are like). There are two main candidates discussed in the literature for what the psychological relation must be, belief and grasping. As we will see, belief is a weaker requirement than grasping and so the debate is over whether a stronger requirement than belief is needed and if so, what needs to be added.

The claim that the psychological relation that constitutes explanatory understanding consists of beliefs is most closely associated with the view that understanding is a form of knowledge, see for example (Greco 2014) and (Khalifa 2012).[18] If understanding is a form of knowledge and the psychological component of knowledge is belief, then it follows that the psychological component of understanding will also be belief (Baumberger et al. 2017, p. 11). For example, we might hold that explanatory understanding just consists in knowing the propositions that make up an explanation and knowing that these propositions collectively

[18] (Kelp 2015) and (Grimm 2006) also hold that understanding is a type of knowledge. But Kelp focuses on objectual understanding, not explanatory understanding. Also, (Grimm 2006) is a bit of an odd case because Grimm seems to accept that belief is not sufficient for understanding and that something like grasping is required. But Grimm proposes that we think of grasping as a species of belief. So, while he does claim understanding is a type of knowledge and that the psychological component of understanding is belief in (Grimm 2006), the notion of belief he is employing is nonstandard.

constitute an explanation. On such a view, given that the psychological component of knowledge is belief, the psychological component of explanatory understanding consists in believing the propositions that make up an explanation and believing that the propositions that are the object of these beliefs constitute an explanation.

The claim that understanding is a sort of knowledge is controversial.[19] We can motivate the idea that something more than knowledge is needed for understanding as follows. Consider the following situation. I am told by a math professor a proof of the Pythagorean theorem and also told that the proof explains why the Pythagorean theorem obtains (and suppose it really is an explanatory proof of the Pythagorean theorem). On the basis of this testimony, I come to know the propositions that constitute the explanatory proof and that these propositions together constitute an explanatory proof. But I only know the propositions constitute an explanatory proof in virtue of the expert testimony, I myself do not see how they even constitute a proof and so do not see how they constitute an explanation. The point is that we can distinguish between cases in which I see how something is a proof (or an explanation) and cases in which I cannot. But, plausibly, one can only have understanding if they do see how something constitutes an explanation. Since in both cases one has knowledge, it cannot be that understanding just consists in knowing (Baumberger 2011).[20] While I used the term 'see' in this example, this something more that is needed is typically cashed out in terms of 'grasping'. Seeing and grasping are of course metaphors and need explication.

---

[19] The debate is a bit difficult to get a handle on because there are different views on what constitutes knowledge. Some of the objections to the claim that understanding consists of knowledge are clearly directed at theories of knowledge where the justificatory requirement of knowledge is externalist.

[20] There are other reasons some have argued that understanding is not a type of knowledge. For example, it is argued in (Kvanvig 2003) that understanding, unlike knowledge, can be lucky (gettierized) Also, in (Elgin 2007) it is argued that understanding, unlike knowledge, need not be factive. But, for both of these potential differences, presumably the psychological component of understanding could be belief (though both Kvanvig and Elgin use the term 'grasping' in their discussions of understanding).

There are multiple ways grasping gets explicated in the literature. For example, (Strevens 2013) proposes a view on which explanatory understanding consists in grasping an explanation. Specifically, what needs to be grasped is each proposition that constitutes the explanation as well as the structure of the explanation. Strevens then cashes out grasping a proposition in terms of understanding that the proposition obtains (which may seem circular but is not, because we are now dealing with a different sort of understanding). He does not go into detail about what it is to understand that a proposition obtains. But presumably it involves representing the world as being such that the proposition obtains (or perhaps the ability to construct such a representation). For example, to understand that the proposition expressed by 'the cat is on the mat' obtains, I need to be able to mentally represent the proposition obtaining, i.e. represent the cat on the mat. The details of how this representing works would depend to a large extent on which theory of the mental representations of propositions we hold. But we might think that in order to understand that the cat is on the mat I need to have concepts of cat, mat, and on that are sufficiently accurate representations.[21]

Often, grasping is discussed with respect to other types of understanding, so it makes it a bit difficult to apply what is said directly to the case of explanations. For example, some philosophers cash out grasping in terms of abilities, such as (Grimm 2011). Grimm talks about grasping an abstract structure, such as a causal system, in the context of discussing scientific understanding. He claims that, in order to grasp a structure, one must have the ability to predict how some elements of the structure will change in response to changes of other parts of the structure. If there is some way we can conceive of the referents in an explanation as constituting a structure, then his idea can be extended to explanations. For example, if an explanation for why

[21] See (Bourget 2015) for more on this issue of understanding a proposition. Bourget argues for the view that phenomenal consciousness plays an integral role in understanding a proposition.

the glass in a window broke is that a baseball went through it, then we can treat the situation as a causal system consisting of the relevant variables, such as the velocity of the ball and the thickness of the glass.

The debate over what the psychological relation is that constitutes explanatory understanding is made a bit unclear by the usage of terminology that means different things to different people in the debate. But once we look past the commonly used terminology it is evident that there is a wide variety of positions. Fortunately, we can be agnostic about what the psychological component of understanding is in general for what follows. Below, we will just focus on one aspect of what the psychological component of having EUM is from learning a proof.

An Understanding-based Approach to EUM

Here I am just going to focus on how an understanding-based approach may differ from an explanation-based approach.[22] An understanding-based approach to EUM differs from an explanation-based approach in the methodology used in developing the theory and in how the theory is justified. As we just saw, an explanation-based approach starts with a theory of explanation, which is developed by looking at examples of explanations and considering the nature of explanation. In contrast, on an understanding-based approach, we make observations about the circumstances in which understanding is attributed and consider the nature of understanding to develop a theory of it. The theory itself will specify what epistemic and psychological conditions a person needs to satisfy in order to have understanding. Explanation can then be defined as what conveys explanatory understanding. We should expect that the theory of understanding and explanation we arrive at on an understanding based approach will be

[22] For an example of such an account, see Wilkenfeld's account of understanding, given in his (2013), and see (2014, p. 3384) for how it relates to explanation.

similar to the theory we arrive at on an explanation based approach in terms of what counts as instances of understanding and explanation, since regardless of the approach we are theorizing about the same things.

The superiority of the understanding-based approach to theorizing about understanding is argued for most extensively in (Wilkenfeld 2014).[23] The gist of his argument is just that the theory is more extensionally adequate than other theories. In the philosophy of mathematics literature, there also seems to be (at least implicit) support for this way of conceiving of explanatory understanding. The methodology for theorizing about explanation advocated for in (Hafner and Mancuso 2005, p. 221) is consistent with an understanding first approach, though they do not explicitly speak about this issue. In (Giaquinto 2016, p. 64), Giaquinto defines explanation in terms of understanding why, which is a clear endorsement of this approach to understanding.

One of the main differences between an understanding-based approach and an explanation-based approach, psychologically speaking, has to do with the content of the mental representation involved in having explanatory understanding. In the case of an explanation first account, on most theories, the content of the relevant mental representation is an explanation. In contrast, on an understanding first account, the content of the representation is not necessarily a representation of an explanation. Rather, it could be a representation of something else that merely enable one to do the sorts of things associated with understanding.

Another significant difference is that an understanding-based approach can result in a greater variety of what counts as an explanation. On an explanation-based approach, we start with a theory of explanation. A major component of the theory will consist of a specification of

[23] See also (Lipton 2009) and (de Regt 2017, ch. 3).

the form or forms that an explanation can have (and since simplicity is a virtue of a good theory, the fewer the better). Anything that does not have the form of an explanation of course fails to be an explanation on such a theory. There may be things which in some way can be effectively put into the form of an explanation, but that are not in themselves explanations. For example, we might imagine that on a theory of explanation, one aspect of the form of explanations is that all explanations consist of sentences. Given a picture or diagram, we might be able to come up with sentences that constitute an explanation, but the picture or diagram is not itself an explanation. In contrast, on an understanding-based approach, since explanation is defined in terms of what conveys explanatory understanding, we might just hold that the picture or diagram is itself an explanation.

Proof and Explanation

I will now argue that some proofs are explanations. But first, let us consider why we might think otherwise. One could motivate the claim that there are no explanations in mathematics by claiming that mathematical truths are necessary truths and that there is no sense in asking why a necessary truth obtains (it just has to). If there is no sense to asking why questions in mathematics, then there is nothing to explain.[24] Alternatively, it might follow from our theory of explanation that no proofs are explanations. For example, we could hold that all explanations fit the causal mechanical model of explanation given in (Salmon 1984). Since there are no causal relations in mathematics, it follows that there are no explanations in mathematics on this view.

---

[24] We might hold that there is some sense in asking why questions in mathematics, but that the explanation will be trivial. For example, we can ask why three is an odd number or why the sum of the angles of a triangle is always 180 degrees. But the explanation will essentially just be that it is part of the nature of three and part of the nature of triangles that they have these properties.

It is an open question whether explanation in mathematics is different than explanation in other domains such as physics. We might think that Salmon's causal mechanical model of explanation only fits scientific explanations (where science here does not include mathematics), in which case it would be possible to also hold that there are explanations in mathematics. There are accounts of mathematical explanation that at least leave it open that it is distinct from other types of explanation. Hafner and Mancuso's "The Varieties of Mathematical Explanation" treats mathematical explanation directly and implies that the advantage of such an approach is that mathematical explanation is distinct. Philip Kitcher's account of explanation, given in "Explanatory Unification" (1981), is a clear example of an account of explanation for which mathematical explanation is not distinct from scientific explanation. Which approach is better will depend on whether there is actually anything distinctive about explanation in mathematics. The important point is that while there are some views of explanation that are incompatible with holding some proofs are explanations, there are others that are compatible.

The argument I will give for the claim that some proofs are explanations just consists of giving some observations that are best explained by holding that some proofs are explanations. This is basically the standard argument in the literature on mathematical explanation. There are two observations for which an explanation needs to be given. The first observation is that mathematicians talk as if some proofs are explanations. This is pointed out in both (Steiner 1978) and (Hafner and Mancuso 2005), and a few examples of such talk can be found in each. Giaquinto gives a vivid example, which is a quote from the mathematician Michael Atiyah:

> I remember one theorem that I proved and yet I really could not see why it was true. It worried me for years . . . I kept worrying about it, and five or six years later I understood why it had to be true. Then I got an entirely different proof . . . Using quite different techniques, it was quite clear why it had to be true. (Giaquinto 2016)

Atiyah obviously does not use the term 'explanation'. But as the new proof shows why the theorem has to be true, it is not a stretch for us to claim that he is calling the proof an explanation, especially in contrast to a proof that apparently did not show why the theorem was true. Since, the obvious explanation for why mathematicians sometimes talk as if some proofs are explanations is that some proofs are explanations, given this quote and the existence of other such talk by mathematicians, we should infer that some proofs are explanations.[25]

The second observation to be accounted for is that mathematicians often create new proofs for theorems already established to be true. One way to explain this behavior is to posit that mathematicians have goals beyond gaining mathematical knowledge, one plausible candidate goal being the goal of gaining understanding. There may be several reasons mathematicians create new proofs. For example, one purpose may be to develop new techniques for proving other theorems. But, some of this activity can be accounted for by positing explanatory understanding as a goal of mathematicians that is fulfilled by creating explanatory proofs. Atiyah's quote above also supports this idea. So, even if mathematician's are not explicit about what purpose a proof serves, some new proofs for a theorem that are developed after the theorem has been established to be true presumably serve the goal of providing an explanation for why the theorem is true (if the first proof is not explanatory).

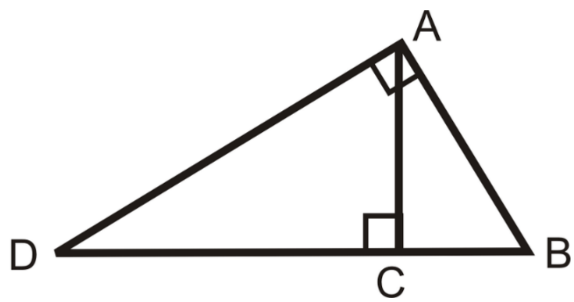
In addition to accounting for these observations, we can note that some philosophers of mathematics have been working on characterizing explanatory proof. For example, Steiner considers two accounts of explanatory proof in (Steiner 1978). The first account, which Steiner rejects, is that a proof is (more) explanatory (relative to other proofs) if it is a (more) general (or

[25] Mejia-Ramos and Inglis (2017) raise some concern about the weight given to mathematicians' explanation talk. They observe that the frequency of explanation talk in mathematical texts is much lower than explanation talk in science texts (which they found using a corpus analysis of mathematical texts on a repository of preprints called ArXiv). If mathematicians' talk of explanation is relatively rare, then perhaps we should not take such talk as strong evidence of there being explanation in mathematics. I do not find their observation to be concerning. I think their observation can be accounted for as follows. While explanation is one of the main goals of scientific practice, it is not one of the main goals of mathematical practice. This accounts for the difference of frequency of explanatory talk between the literatures. Even if mathematicians do not often talk as if proofs are explanatory infrequent talk of this sort still needs to be accounted for.

abstract) proof. The second account Steiner considers (and advocates for) is that a proof is explanatory if it relies on a characterizing property of an entity or structure mentioned in the theorem. More recently Marc Lange (2016) provides an account of explanatory proof where a proof is explanatory if the result proved has a salient feature and the proof of the result exploits a similar salient feature.[26] While I take the two observations mentioned above as carrying more evidential weight, the fact that philosophers have been able to come up with substantial proposals lends further credence to the claim that some proofs are explanations.

I will now give an example of a proof that is supposed to be explanatory, which is taken from (Steiner 1978). The example is a proof of the Pythagorean theorem discussed in Polya's *Induction and Analogy in Mathematics* (and has been attributed to Albert Einstein). The Pythagorean theorem says that for any right-triangle, the area of the square on the hypotenuse is equal to the sum of the areas of the squares on the other two sides of the triangle. If we let c denote the length of the hypotenuse and a and b denote the lengths of the other two sides of the triangle, the Pythagorean theorem is equivalent to the claim that $a^2 + b^2 = c^2$. The question we want to answer is, why is $a^2 + b^2 = c^2$. There are two steps in the proof. The first step is to just point out that $a^2 + b^2 = c^2$ if and only if for any rational number k, $ka^2 + kb^2 = kc^2$. This can be justified geometrically, but we can also easily see that it follows algebraically. The second step in the proof is to show $ka^2 + kb^2 = kc^2$ for all k. Of course, if we show that the result holds in the case of the squares on the sides of the triangle then we will have proven this more general claim. But it is not evident by looking at the squares that $a^2 + b^2 = c^2$. Fortunately, there are figures for which it is evident by looking at them that $ka^2 + kb^2 = kc^2$. Consider the following image.

[26] Lange's account may seem quite similar to Steiner's account. Both accounts are more complicated than what I say here, see (Steiner 1978) and (Lange 2016) for more details. Also, see (Lange 2016) for some examples of proofs that Steiner's account and Lange's account differ on.



Let x be the area of the triangle ABC, let y be the area of the triangle ACD, and let z be the area of ABD. Clearly $z = x + y$. Furthermore, ABC, ACD, and ABD are similar triangles. We can see that ABC and ABD are similar by noting that they share an angle in common and each have a right angle. Similarly, for ACD and ABD. Since we have three similar figures on each side of the triangle ABD there must be some k such that $x = ka^2$, $y = kb^2$, and $x + y = kc^2$, where a is the length AB, b is the length of AD, and c is the length of BD. Thus, we get that $ka^2 + kb^2 = kc^2$, from which it follows that $a^2 + b^2 = c^2$.

Why should we take this proof to be explanatory and what makes it an explanation? As we have seen above, there are a couple different approaches we can take to answering this question. On an understanding-based approach to EUM, since explanation is defined in terms of understanding, we need to evaluate whether this proof conveys explanatory understanding. One way we could do this is to look for phenomenological evidence, i.e. the feeling of understanding. As there is no empirical study that I can point that speaks to this issue, I will leave it up to the reader to assess this. Mathematician Terry Tao writes about the above proof, "it's not particularly earth-shattering, but it is perhaps the most intuitive proof of the theorem that I have seen yet" (Tao 2007). So, there does seem to be something cognitively significant about this proof as compared to other proofs of the Pythagorean theorem. Though, even if there is some phenomenological feeling of understanding associated with learning the proof it is unclear whether we should take this as evidence of understanding, for opposing views on this issue see (Trout 2002) and (Grimm 2009).

Rather than look for a feeling of understanding, we can look for signs that a person who has learned this proof has understanding in the form of abilities they have in virtue of learning the proof. One ability that may be relevant is the ability to easily reproduce the proof. One striking thing about the above proof as compared to other proofs of the Pythagorean theorem is that it uses an exceedingly simple diagram. Once someone understands the proof, they can essentially see the Pythagorean theorem obtains just by looking at a right triangle and visualizing a line from the right-angle perpendicular to the hypotenuse. There may be other relevant abilities, but to give a more substantial argument on an understanding-based approach, we would need an actual account of explanatory understanding.

Another way to argue that the above proof is explanatory is to consider what explanatory proof is in general and then show that this proof fits with the given account. In (Steiner 1978) it is shown that this proof fits with the two accounts of explanatory proof Steiner considers. The above proof is explanatory according to the first account he considers in that it proves $a^2 + b^2 = c^2$ by establishing something more general, i.e. $ka^2 + kb^2 = kc^2$ for all k. The above proof is explanatory according to the second account he considers because it relies on the fact that only a right triangle is decomposable into two triangles similar to each other and itself. We can also motivate the claim that it satisfies Marc Lange's account as follows. The salient feature of the result is that it holds for right triangles but no other triangles, so there is a contrast that needs to be accounted for. The proof then points to a salient feature of right triangles, namely, that they can be decomposed in the way needed for the proof to work.

Both Lange's and Steiner's account seem to be motivated in part by considering the contexts in which explanations arise. Explanations are typically sought and given in cases where there is a contrast between the case in which the result holds and cases in which it does not.[27] The motivation for asking why the Pythagorean theorem obtains is the observation that it holds for all and only right triangles. What is it about right triangles, in contrast to obtuse and acute triangles, such that the area of the square on the hypotenuse is equal to the sum of the areas of the squares on the other two sides of the triangle? Both Steiner's and Lange's account seem to capture this idea (albeit in slightly different ways). Since on both accounts the above proof is an explanatory proof, we have some reason to think it is an explanation. If this proof is indeed an explanation, then some proofs are explanations. Even if this proof is not an explanation, the considerations further above still lend some credence to the claim that some proofs are explanations.

Deduction and EUM

In this section, I will argue that going through the relevant deduction is required for gaining EUM. To do this, I will first briefly say a bit about deduction as a mental process. The distinctive feature of deduction is that we think that something must be the case in virtue of some reasons (as opposed to thinking something must be the case because it is self-evident or because it is unclear how it could be false).[28] There is a sort of experience associated with deduction. For example, suppose I am looking for my keys. I think of the possible places it could be. As I check each place it could be, I infer they must be located at one of the remaining possibilities.

---

[27] See chapter five of (van Fraassen 1980).

[28] Deductive logic codifies our conception of *good* deductive inference, but of course people can and do often make bad deductive inferences, for example, the fallacy of denying the antecedent. So, the certainty associated with deduction need not be a normative certainty. Indeed, as we will see below, the common mistakes that people make in deductive reasoning are data that a psychological theory of deduction needs to be able to account for.

Deduction is whatever process that underlies this conscious experience. Deduction also occurs unconsciously. We can see this by considering that people sometimes explain their own behavior by attributing to themselves a deduction in the past, in circumstances in which they had no conscious experience of deducing. For example, consider a slight variation of the searching for keys example. My roommate observes I am searching the pockets of my jacket repeatedly and asks why am searching them over and over again. Even if I had not consciously deduced that my keys must be in my jacket pocket, it would be normal to explain my actions by pointing out that I already checked all the other places they could be and as a result think they must be in one of my jacket pockets. If this explanation is accurate, i.e. if I really did have those thoughts even if I did not have them consciously, it follows that deduction can occur unconsciously. Indeed, we might think that deduction occurs unconsciously quite often.

I will now argue that deducing is necessary for gaining EUM. Consider the following proof:

> *Theorem*: There are infinitely many prime numbers.
> *Proof* : Let $n > 1$ be an integer. Since n and $n + 1$ are consecutive integer, they must be coprime. To see this consider that $(n + 1) - n = 1$. Let m be a number that divides both n and $n + 1$. Then there exist integers p and q such that $mp = n$ and $mq = n + 1$. We thus have $mq - mp = m(q - p) = 1$. From this it follows that m divides 1, hence m must be equal to 1. Thus the only common divisor of n and $n + 1$ is 1. Then since n and $n + 1$ are coprime the number $N_2= n(n + 1)$ must have at least two different prime factors. Similarly, since the integers $n(n + 1)$ and $n(n + 1) + 1$ are consecutive, and therefore coprime, the number $N_3 = n(n + 1)[n(n + 1) + 1]$ must have at least 3 different prime factors. This can be continued indefinitely. Therefore, the theorem is proved. (Saidak 2006)

For the sake of argument, assume this proof is explanatory. My argument for the claim that deducing is necessary for gaining EUM is similar to the objection to the claim that understanding consists in knowing mentioned above. The issue with the claim that understanding consists in knowing is that, (on some theories of knowledge) I could know the propositions that constitute an explanation and know that together these propositions constitute an explanation through testimony. But, if I do not see how they constitute an explanation, because I do not even see how they constitute a proof, then, intuitively, this is not a case in which I have explanatory understanding. The upshot of these considerations is that in order to gain explanatory understanding you need to see the explanation as an explanation. In the case of explanatory proofs, this also requires seeing the proof as a proof. Due to the nature of a proof, the only way to see that a purported proof really is a proof is via deduction.[29] Hence, in order to have EUM, at least from learning a proof, one must go through the relevant deduction.[30]

The above argument relies on an intuition about a certain type of case. I will provide some further considerations in support of this intuition. First, we can make some considerations about the attribution of understanding in the relevant cases. Someone who learns a proof through testimony, say by reading it in a textbook, could memorize the proof and reproduce it at will. This is probably sufficient in some circumstances for an attribution of understanding. But, I think that in the context of math education there is a higher standard and we can see that there is some effort on the part of instructors to differentiate between those that know a proof and those that understand a proof. For example, some common advice for studying for math exams is to not memorize the proofs but rather to acquire the ability to come up with the proof. [31] This advice

[29] Bourget says the following about this case: "The opaque proof. You are trying to understand a proof. You know (because you have been told by someone you trust) that the conclusion follows from the premises. You have gone through all the steps of the derivation. Still, you don't really see how the conclusion follows from the premises. At last, after going through the steps numerous times, you finally see it. Here the proposition grasped is along the lines of <such and such follows from such and such>" (Bourget 2015). I interpret Bourget as agreeing with my position, i.e. I take the requirement that the person grasp a proposition of the form <such and such follows from such and such> to just mean that someone has gone through the relevant deduction.

[30] I think that in general, gaining explanatory understanding requires making an inference, but since I do not need this stronger claim for what follows I will not argue for it here.

[31] For example: "The only way to prepare for such tests is practicing…. Memorizing proofs from class, or solutions from the homework won't help. However understanding them thoroughly will help you. For instance, one thing you

may just be giving the best strategy for doing well on the test. But, plausibly, part of the impetus for giving this advice is that the math teachers think there is some educational value in being able to prove a theorem without relying on rote memorization. Also, in my experience, math teachers will sometimes change the wording of theorems on exams from how they were presented in class, which makes the theorem more difficult to prove for those who have just memorized a proof, because the language of the proof needs to be adapted to the wording of the theorem. These observations suggest that, in the context of learning and doing mathematics, understanding would not be attributed to someone who merely knew a proof through testimony. Designing exam questions to distinguish between those who merely know a proof through testimony and those that have a grasp on how the proof works, is indicative that there is a difference in the abilities had in each case. We can take the difference of abilities to justify the difference the claim that the person who knows a proof merely through testimony does not have understanding.

Another way to motivate this intuition is to consider the mental representations one has in the case where you know that such and such is a proof through testimony, but you do not yourself see how it constitutes a proof. We can stipulate that you do understand the sentences that represent the propositions that constitute the proof. In this case you presumably have a mental representation of each proposition in the proof. You also have a mental representation of the form 'the propositions expressed by the sentences <P1, P2, … > constitute a proof' (where P1, P2, etc. are the relevant sentences). But someone who has gone through deduction presumably has, in addition to these mental representations, an integrated representation of the

---

could do would be look at a simple result from class/book. Cover up the proof, and see if you can figure it out yourself. If you can't, then maybe see how your notes/book starts the proof. Then cover up the rest of it, and see if you can complete it. The same goes for looking up homework solutions" (<https://www.math.cmu.edu/~gautam/teaching/2011-12/269-vector-analysis/pdfs/exam-notes.pdf>). Also, "Mathematics requires precision, habits of clear thought, and practice. Cramming for an exam will not only fail to produce the desired result on the exam, it will also reinforce a bad habit---that of trying to do mathematics by memorization rather than understanding" (<https://www.math.uh.edu/~dblecher/pf2.html>).

proof and its conclusion. By 'integrated' I mean a representation that represents the propositions that constitute the proof all together. My suggestion is that this mental representation is needed for understanding. The above intuition would then be justified by pointing this difference in mental representation between the two cases.

Why should we think that having an integrated representation of the proof is necessary for having understanding? To have explanatory understanding you need to have a representation that in some way represents the connection between the explanans and the explanandum. There are various views on what propositions are, but for example, consider the view that propositions are sets of possible worlds. My mental representation of the conclusion of an explanatory proof would then be a representation of a set of possible worlds. But, to see why the conclusion must be true given the truth of the premises, what I need is to represent the set of possible worlds in which all the premises are true, which will be a subset of the possible worlds in which the conclusion obtains if the deduction is valid. In the case of someone who knows a proof only through testimony, while they have representations of each proposition, these representations will be disconnected from one another. The explanans and explanandum are in some way connected by the mental representation of the form 'the propositions expressed by the sentences <P1, P2, … > constitute a proof', but they are not connected at the semantic level. This difference justifies the claim that having an integrated representation of the proof is necessary for having understanding.[32]

In this section, I have argued that some proofs are explanations and that going through deduction is necessary for gaining EUM by learning an explanatory proof. These claims will be

[32] I am not claiming that there will necessarily be some conscious experience associated with having the integrated representation, though presumably the feeling of understanding would be associated with having this representation in short term memory.

used in a later section as follows. For any proof from which we do gain explanatory understanding, it must be that one has gone through deduction in learning the proof. So, it must be that the proof is one for which it is possible to deduce. Thus, any psychological theory of deduction will have to account for how it is that deduction occurs in the context of learning the proof. As we will see, in some cases this is challenging for the accounts of deduction to do. I will now turn to the issue of the psychology of deduction.

Psychological Accounts of Deduction

In this section, I will consider two accounts of deduction, the mental logic account and the mental models account. These accounts of deduction fall within the classic computational framework of cognition, i.e. both presuppose deduction is a computational process that is to be accounted for by specifying the mental representations and algorithm underlying the computation. Before getting to the theories, in order to provide some context for understanding them, it is worth considering the scope of what these theories are attempting to account for. The most obvious candidate for behavior that needs to be accounted for by the theories is activity involving deductive arguments. For example, if I provide some argument and ask, ‘must the conclusion be true if the premises are true?’, then answering the question presumably involves deduction. Indeed, the observations the theories have been designed to account for are observations of the performance of individuals in answering questions of this sort. The most pertinent observations about performance in these tasks are observation about the systematic errors people make (errors relative to the normative theory of deductive logic). There are three types of errors that are frequently discussed, errors in categorical syllogistic reasoning, errors in reasoning about conditionals, and errors in the Wason selection task. Since these common errors play a significant role in theorizing about the psychology of deduction, I will briefly review them now.

A categorical syllogism consists of three categorical propositions. The conclusion shares one common predicate with each of the premises and the premises have a different predicate in common. For example, the following is a categorical syllogism: All dogs are animals. Some dogs chew on shoes. Therefore, some animals chew on shoes. There are 256 distinct forms of categorical syllogisms, only 24 of which are valid, given an Aristotelian interpretation of categorical propositions where the universal propositions are assumed to imply existence, e.g. all dogs are animals does imply there are dogs and animals. On the modern (Boolean) interpretation of categorical propositions, where a universal proposition does not imply existence, there are 15 valid forms. The pertinent observation with respect to human performance in making valid inferences for a categorical syllogism is that, while in many cases most people get the correct answer, there are some cases for which most people get the answer wrong, i.e. they draw logically invalid conclusions. Johnson-Laird (1984) gives the following example. From the premises, some of the artists are beekeepers and all the beekeepers are chemists, most people correctly supply the conclusion that some of the artists are chemists. But, from the premises, none of the archers are boxers and all the boxers are clerks, very few correctly infer (assuming the Aristotelian interpretation) that some of the clerks are not archers.[33]

Broadly speaking, tasks that involve reasoning about conditionals are just tasks where the deduction involves at least one proposition that is a conditional. An example of a conditional

[33] (Chater and Oaksford 1999) is a good resource for more background on research into categorical reasoning. While Chater and Oaksford put forward a probabilistic model to account for performance on tasks involving categorical inference, they compare their model's performance to that of the mental models model and mental logic model I will discuss below. They also compile some of the earlier empirical research on human performance in categorical inference tasks and as such is good source of references to earlier work.

argument is as follows. If the cube is blue, then the ball is green. The cube is blue. Therefore, the ball is green. Again, the pertinent observation is that, while in many conditional reasoning tasks humans perform them correctly, there are certain tasks that many people perform incorrectly. A couple of well-known examples are the fallacies of denying the antecedent and affirming the consequent. Given the premises, if the cube is blue, then the ball is green, and, the cube is not blue, many people incorrectly infer that the ball is not green. Similarly, given the premises, if the cube is blue, then the ball is green, and the ball is green. Many people incorrectly infer that the cube is blue.[34]

The Wason selection task is a task that implicitly involves reasoning about conditionals. The Wason selection task is as follows. An individual is presented with four cards, each of which has a number on one side and a letter on the other. The cards lie flat on a table in front of an individual and on the faces of the four cards are the letters E and K and the numbers 4 and 7, respectively. The goal of the task is to determine if the conditional 'if there is an E on side then there is a 4 on the other' obtains for the four cards. The participants are instructed to specify which cards need to be turned over to see if the rule is true or false. The correct solution to this task is to choose the card with an E and the card with the 7. However, most people give an incorrect answer. Most either answer that only the card with an E on it needs to be turned over or that both the card with an E on it and the card with a 4 on it need to be turned over (Rips 1994, p. 179).

[34] There is large literature on reasoning with conditionals that is made difficult to navigate due the variety of types of conditionals, e.g. material, indicative, causal, etc. Depending on the conditional, it is unclear whether we should take the reasoning involved to be deductive. (Markovits et al. 2012) provides an easy to follow disentanglement of some of these issues and useful citations. Chapter five of (Rips 1994) covers some empirical findings on conditional reasoning and the Wason selection task.

There is some debate on what activities involve deduction, even in these cases just mentioned where completing the task straightforwardly seems to involve deduction. That these sorts of tasks involve deduction has been challenged in two different ways. As pointed out in (Rips 1994, p. 27), Newell (1980) shows that his model of problem solving can be applied to tasks involving answering question about syllogisms. In a sense Newell is not denying that there is deduction in these tasks. But according to Newell, deduction is just a special case of a more general cognitive process, and so we do not need a theory of deduction. In contrast, Oaksford and Chater (2008) argue that many activities psychologists have taken to involve deduction, such as answering questions about arguments like the one given above, do not in any sense involve deduction (at least not usually). Oaksford and Chater do not deny there is a cognitive process of deduction, their claim is rather that it is a very specialized cognitive process used by some individuals in very few circumstances, such as doing mathematics. Oaksford and Chater argue that most people use Bayesian inference in the tasks mentioned above.

It is also an open question whether there are activities that, while do not seem like they involve deduction, in fact do. For example, Rips argues that the mental process of deduction plays a role in a wide variety of cognitive tasks, such as categorization and general problem solving.[35] While the main data he uses to shape his theory is data about systematic error in deduction tasks, his theory is supposed to account for quite a lot. As we will see below, the challenge I raise for his theory are cases in which I claim deduction is involved but for which it is unclear how his theory can provide us with an account of how. While I will consider how Rips would respond to the challenge I raise, I doubt he would respond by denying that deduction is involved.

[35] See chapter eight of (Rips 1994) for more on this.

The rest of this section will consider the two main accounts of the mental process of deduction. I will first go through the mental logic account of deduction and then the mental models account of deduction.

Mental Logic

The mental logic theory of deduction is the theory that deduction is a mental process that constructs a mental representation of a derivation of the inferred proposition. To construct the mental representation of the derivation, the mental process uses syntactic rules on mental representations of the propositions that are the basis of the inference. The syntactic rules are similar to the rules of deductive logic.[36] To illustrate the idea consider the following example from (Rips 1994, pp. 105 - 108). Consider the argument:

(1) If Betty is in Little Rock, then Ellen is in Hammond.

(2) Phoebe is in Tucson and Sandra is in Memphis.

(3) Therefore, if Betty is in Little Rock then, Ellen is in Hammond and Sandra is in Memphis.

The task is to evaluate whether this is a valid argument. The mental process is very similar to how we might consciously think through whether the argument is valid or not. First, when (2) is recognized as a conjunction, the rule of AND-Elimination is applied, which creates two new sentences in working memory, namely ‘Phoebe is in Tucson’ and ‘Sandra is in Memphis’. From this point, it is necessary to work backward from the conclusion by applying Backward IF-

[36] “The central notion in the theory will be that of a mental proof. I assume that when people confront a problem that calls for deduction they attempt to solve it by generating in working memory a set of sentences linking the premises or givens of the problem to the conclusion or solution. Each link in this network embodies an inference rule..., which the individual recognizes as intuitively sound. Taken together, this network of sentences then provides a bridge between the premises and the conclusion that explains why the conclusion follows” (Rips 1994, p. 103). While my overview of the mental logic theory is based on the presentation of it in (Rips 1994), see (Braine and O’Brien 1998) for another version of the mental logic theory.

Introduction, which creates the sentence ‘Betty is in Little Rock’ in working memory as a supposition. The rule IF-Elimination then creates the sentence ‘Ellen is in Hammond’ in working memory. A few more rules are applied to get a complete construction of the derivation in working memory. As the process successfully constructs a derivation, the solution output is that the argument is valid.

Much of the work in developing a plausible mental logic theory of deduction is in specifying the rules and the order and circumstances in which they are applied. To see why this is difficult to do just consider the rule OR-Introduction common to systems of logic. The rule just says that if we have a proposition *p* we can infer *p* OR *q* for any proposition *q*. If our account of the mental process of deduction had an unrestricted OR-Introduction rule then the process would not necessarily ever end, which would obviously not be a plausible account of the mental process. Yet some derivations do require OR-Introduction so we cannot simply do without the rule. This situation illustrates the difficulty and importance of specifying the order and circumstances in which a rule is applied.

Rips has developed a system of mental logic that he gives in a program he calls PSYCOP, the details of which can be found throughout (Rips 1994). The value of PSYCOP is in the way it avoids problems akin to the problem with OR-Introduction and in the way it can account for observations psychologists have made about common errors in deductive reasoning. For example, consider the observation from (Marcus and Rips 1979). It generally takes people longer to correctly answer that the conclusion follows from the premises in an argument of the form modus tollens than it does for an argument of the form modus ponens. This observation can be accounted for by pointing out that it takes more rules for the derivation of an argument of the form modus tollens than it does for an argument of the form modus ponens. Specifically, for modus ponens, only one rule is needed, i.e. IF-Elimination. But, in PSYCOP, a derivation for an argument of the form modus tollens uses two rules, Backward Negation Introduction and If-Elimination, which accounts for the longer time.

Beyond considering that the mental logic theory can account for observations of human performance in tasks involving deduction, we should consider the original motivation for the theory. The primary motivation is phenomenological. Deductive inferences seem to follow rules. Whenever I have a succession of thoughts of the form *p* and then *not p or q* if find it irresistible to conclude *q*. Furthermore, this holds for people who cannot give an explanation using truth tables for why this type of inference is sound. As a result, the mental logic theory has a sort of intuitive appeal. Indeed, historically, deductive logic has been viewed by some as a codification of the laws of thought. Furthermore, the close connection between logic and computation make it easy to see how the theory fits in with a computational view of the mind.[37]

Mental Models

The mental models theory of deductive reasoning is the theory that deduction is the process of constructing mental models (a type of mental representation), making hypotheses, and evaluating those hypotheses using the mental models. The basic idea is that the meaning of the premises (along with background knowledge) is used to construct mental models that represent the possibilities consistent with the premises. A potential conclusion is generated as a hypothesis. Then there is a check to see if the hypothesis is true in all models. If so, the person will infer the hypothesis as a conclusion that follows necessarily from the premises. To illustrate consider the following example from (Johnson-Laird and Yang 2008). Consider the following argument: (1) The triangle is to the right of the circle. (2) The circle is to the right of the diamond. (3)

---

[37] (Bonatti 1998) gives an overview of the historical development of some of the ideas important for the mental logic theory.

Therefore, the triangle is to the right of the diamond. With the theory of mental models we get the following psychological account. In making a deduction we first create a mental model of each distinct possibility contained in the premises. In the case of these two premises there is just one distinct possibility, i.e. the triangle is to the right of the circle, which is to the right of the diamond.[38] Mental models are iconic representations, i.e. they represent something in virtue of a similarity between the vehicle of representation and what is represented. For example, '◇ ◯ △' is an iconic representation of the possibility consistent with the premises. The next step in the process is to check and see if the hypothesis is true in all of the distinct possibilities. In this case, the hypothesis is supplied by the argument we are considering. If the hypothesis does obtain in each mental model, then we infer that it follows deductively from the premises. In this case, it is true of the represented possibility that the triangle is to the right of the diamond, so we infer that the conclusion follows from the two premises.

The main challenge for the mental models theory is to apply the theory to deductions that involve complex propositions that contain logical connectives. A useful example is disjunctions. There are multiple possibilities consistent with a disjunction. So, for a proposition of the form 'P OR Q' there will be mental models of P and not Q, not P and Q, and P and Q. While disjunctions can easily be handled by the theory, negations are a difficult case. For example, consider the proposition that Jake is not a squirrel. It is unclear how we could have an iconic representation of this proposition. One possible solution is to construct a mental model that represents Jake as something other than a squirrel, e.g. a cat. But there are obvious issues with such a solution. For one, based on such a model we could infer that Jake is a cat, even though this does not follow

[38] There are of course many ways the circle can be to the right of the triangle and many ways the triangle can be to the right of the square. The mental representation does not represent any one of these more particular states of affairs but rather represents all of these distinct states of affairs schematically.

from the premise that Jake is not a squirrel. Since people do not generally make mistakes like this, this sort of solution does not seem promising. A better solution, which is the solution Johnson-Laird among others advocate, is as follows. To represent a proposition of the form not P, the proposition P is represented by a mental model and attached to this mental model in some way would be a label that marks it as false.

Before we see how this works in a deduction, one important additional piece of this solution is what Johnson-Laird calls 'the principle of truth', which states that people tend to represent only what is true.[39] So, for example, in the example of representing a disjunction given above. Rather than have a mental model that represents the possibility in which P and not Q obtain, the mental model would just represent P as obtaining. These models could be filled out to include not Q and not P respectively, but this would only occur if were needed.

To see how this all would work in a deduction, consider an argument of the form (1) P or Q, (2) not P, (3) therefore, Q. There are three mental models consistent with (1), a model that represents P, a model that represents Q, and a model that represents P and Q. There is one mental model consistent with (2), a model of not P. Since the model of not P is inconsistent with two of the models already generated, those two models are eliminated and we are left with just one model, which represents not P and Q. Since Q does obtain in all remaining models, the conclusion follows from the premises. Similar accounts can be given for arguments of other forms involving these connectives. Based on these accounts we can formulate an algorithm that handles all the relevant cases in order to develop a computational model of deduction for the

[39] "The principle of truth postulates that individuals normally represent what is true, but not what is false. It does not imply, however, that they never represent falsity. Indeed, the theory proposes that they represent what is false in 'mental footnotes', but that these footnotes are ephemeral. People tend to forget them. But as long as there are remembered, they can be used to construct *fully explicit* models, which represent the true possibilities in a full explicit way" (Johnson-Laird and Yang 2008, pp. 348).

logical connectives, see for example (Bara et. al. 2001). We can then compare the performance of the model to the observations of human performance in order to evaluate the model. For example, a deduction involving an argument that has the form modus ponens requires looking only at one model (which follows form the principle of truth), so it is predicted to be easily recognized as valid. In contrast, recognizing that an argument of the form modus tollens is valid requires the full set of possible models to be constructed and as a result it is predicted that it will typically take longer, which is what we observe in human performance as noted above.[40]

One potential problem with Johnson-Laird's solution to the problem of negation is that it seems to involve admitting that deduction in some cases operates according to rules that utilize only syntactic features of a representation. For example, we can extract the following rule from the deduction above. If model X was generated from a premise that is a disjunction and model not X was generated from a premise for which negation is the main logical connective, eliminate model X. This rule relies just on being able to recognize a formal feature of models. As a result, Johnson-Laird's solution seems to be introducing a mental logic into the theory and this is viewed by some as a problematic feature the solution, see for example (Rips 1994, p. 358) and (Rosa 2017). I will return to this issue below as I will argue that cases of deduction in mathematics exacerbate this problem with the solution.

Since we will be considering some difficulties for the theory below, which may prompt some adjusting of the theory, it is important to have in mind the motivation for the theory. The main motivation for the theory is a generalization from cases of relational reasoning, e.g. spatial reasoning. As we saw in an example above, relational reasoning is easy to account for with the mental models theory. With a mental logic theory, we would need to posit some rules for each

---

[40] Example from (Harman and Kulkarni 2014).

relation or posit some mechanism for translating the premises into ones that would work with existing rules to handle these cases. The mental models theory also has an intuitive appeal as an alternative to the mental logic theory. Where there is an analogy to be made between mental logic and the inference rules of deductive logic, we can make an analogy between mental models and the model theory for a logical system, e.g. truth tables for propositional logic.

Implications for EUM

I will now consider what the psychological theories of deduction imply for gaining explanatory understanding from learning explanatory proofs. Above, I argued that gaining EUM from learning an explanatory proof requires one to go through deduction. So, we can make inferences about gaining EUM by considering how deduction works. I am interested in two sorts of considerations we can make based on these theories of deduction. The first sort are predictions about which proofs it is possible to go through deduction for. The second sort of considerations are comparative predictions about which proofs it will be more or less easy to go through deduction for. In this section, I will say a bit more about each sort of consideration.

All mathematical proofs are in some sense candidates for deduction. The premises of each proof collectively should logically entail the theorem being proved and so the premises of the proof collectively are good reasons for believing the theorem must be true. But it is not controversial that there are proofs in mathematics for which it is impossible to go through deduction. The proof of the four color theorem by Appel and Haken (1977) is an example of such a proof. The proof was constructed in part by a computer that verified the theorem holds for many of the cases covered by the theorem. The conclusion is not deducible from the premises because the proof is just too long. The amount of memory humans have is a constraint on which proofs it is possible to go through deduction for.[41] Of course, how much memory humans have is not something the above theories will differ on. Both theories would correctly predict that it is impossible to deduce the four-color theorem from its proof.

The question of interest for each theory is, what proofs does the theory predict we can go through deduction for and which proofs does the theory predict we cannot go through deduction for. If the theory is accurate, we should find that the theory is extensionally adequate. If the theory is not extensionally adequate, then we have some reason to reject the theory or modify the theory. Of course, we need some pre-theoretic way to assess whether deduction is possible for a given proof. In section one I argued that deduction is required for gaining explanatory understanding and so proofs that we gain explanatory understanding are a class of proofs that we have some theory-independent reason to think deduction is possible for. As we will see in the next section, I will try to apply each theory of deduction to a couple examples of theorems and their proofs. I will show for each theory of deduction that no account can be given of how we could deduce the conclusion from the premises of the proof. We can then infer that each theory of deduction predicts that deduction is not possible for the proof. Of course, my point will be that the theories get it wrong in these cases as they are explanatory proofs.

The other sort of prediction we can derive from these theories are predictions about which proofs it will be easier to go through deduction for. Insofar as gaining EUM from learning proofs is a goal of mathematicians, these sorts of predictions are relevant to mathematical practice. If it is easier to go through deduction for one explanatory proof of a theorem than another, then presumably it is easier to gain explanatory understanding from it and it is a better

[41] Given that we can come to believe the four color theorem must be true on the basis of the proof given by Appel and Haken there presumably is some deduction that we can go through, the point is just that the deduction does not use the premises of the proof.

proof in this way. In the last section of this chapter, I will discuss how these considerations could impact the practice of constructing and presenting proofs. Since we are considering how to produce the best proofs with respect to deduction, we need to consider the experience of someone learning a proof. So, to set up the discussion, here I will distinguish between two parts of the process of learning proofs.[42]

The first step in going through a deduction in the course of learning a proof is translating the external representation of the proof into mental representation of the proof. We need to understand what the premises are that the conclusion is supposed to follow from. So, one thing to consider in gauging how easy it is to go through deduction for a proof is the ease of translation from the external representation of the proof to the relevant mental representations. Since the theories of deduction differ on what mental representations are used for deduction, presumably the theories would make different predictions about how best to present proofs.

The next step in a deduction is to apply to the algorithm underlying the computation to the input. We can consider how the algorithm will perform for various inputs. For example, as we saw above, the theories make predictions about how long a deduction will take for different basic argument forms based on how the algorithm works for them. So, time for deduction is one aspect of a proof we can consider. Another related example of something we can consider is how much memory a proof requires to go through deduction for. Presumably the less time it takes for a deduction and the less memory it takes for a deduction, the easier it is to go through deduction

[42] One cluster of issues relevant to this discussion is the metaphysics of proofs. Perhaps the most relevant question is when are two proof tokens of the same type? If we are going to make comparisons between different proofs we need an account of when we have two different proofs. To avoid having to get into these issues, I will just take on the assumption that if the account of deduction differs for two proof tokens then they are different types of proofs.

for (or at least these are good proxies for ease of deduction).[43] In a section below, I will go into more detail about what we can say in general about what each theory would predict for which proofs would be easier to go through deduction for based on considerations about the algorithm.

Evaluating the Psychological Theories of Deduction

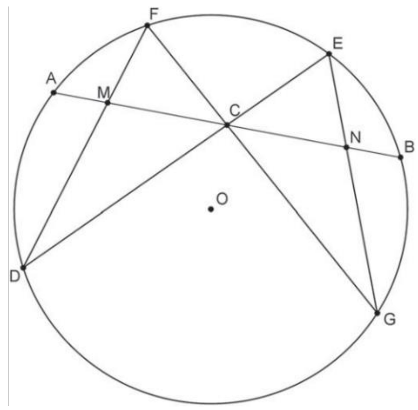
In this section, I will discuss two examples of explanatory proofs that I take to raise some issues for the psychological theories of deduction. As mentioned above, the issues are raised because each example is of an explanatory proof and so it must be possible to deduce the theorem proved from the premises of the proof. Yet, as I will show, it is unclear how to give an account of the deduction based on these theories. So, in light of these examples it will be necessary to alter or add on to the theories. First, I will discuss an example of an explanatory proof for which a diagram plays an integral role. Then, I will discuss an example of an explanatory proof for which identity acts as a logical constant.

Consider the following proof, which is claimed to be explanatory in (Frans and Weber 2014):

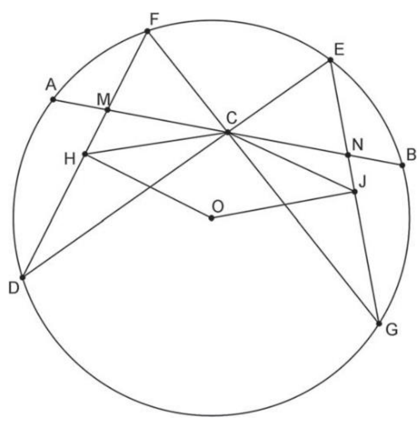
The Butterfly Theorem (of Euclidean Geometry): Let C be the midpoint of a chord of a circle, through which two other chords FG and ED are drawn; FD cuts AB at M and EG cuts AB at N. Then C is the midpoint of MN.

[43] 'Ease of deduction' is essentially just a placeholder. It is certainly worth thinking more about what the desirable features of a deduction are from the perspective of practicing mathematicians. But I will leave this issue for future work as I think the ambiguous idea of how easy a deduction is enough to get my point across that the theories of deduction will have implications for something like ease of deduction across here.

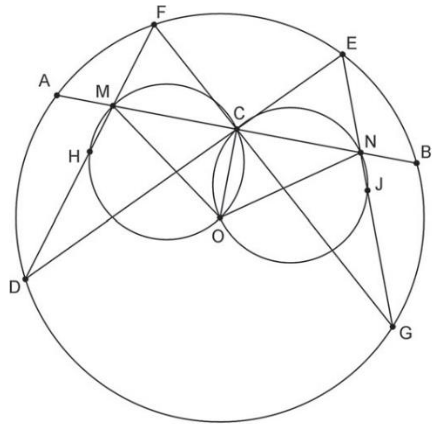
Proof:



(1) ∠FDE and ∠EGF are equal (Inscribed angles)
∠DFG and ∠DEG are equal (Inscribed angles)
(2) △FCD and △ECG are similar (1: Property similar triangles)
(3) FD/FC = EG/EC (2: Property similar triangles)



(4) Construct point H on DF such that OH is perpendicular to DF, and construct point J on EG such that OJ is perpendicular to EG
(5) FD = 2FH (4: Property of a chord)
EG = 2EJ (4: Property of a chord)
(6) FH/FC = EJ/EC (3,5: Substitution)
(7) △FCH and △ECJ are similar (1, 6: Property of similar triangles)
(8) ∠EJC and ∠FHC are equal (Property corresponding angles)



(9) ◻OCMH is cyclic (4: ∠FHO + ∠ACO = 180◦)
◻OCNJ is cyclic (4: ∠EJO + ∠BCO = 180◦)
(10) ∠MHC and ∠MOC are equal (9: Inscribed angles)
∠CON and ∠CJN are equal (9: Inscribed angles)
(11) ∠MOC and ∠CON are equal (8, 10: Substitution)
(12) △OCM and △OCN are equal (10, 11: Property equal triangles)
(13) C is the midpoint of MN (12: Property equal triangles)

To see what issues this proof raises we can simply try to apply the theories to it and see what the difficulties are. I will start with the mental logic theory. The Butterfly Theorem itself is essentially a conditional and the proof proceeds by assuming the antecedent of the conditional and then deriving from this assumption the consequent of the conditional. So, the mental proof begins with an application of the rule IF-Introduction. We can observe that the numbered lines in the external representation of the proof are effectively intermediate conclusions (except for (4)) and so are themselves the result of deductions. For example, (1) follows from the antecedent of the theorem and a fact about inscribed angles, namely that inscribed angles with the same intercepted arc are equal. In this case, the angles ∠FDE and ∠EGF both have the intercepted arc between F and E. Presumably, the best way to account for this inference with the mental logic theory is to posit that the inference is an application of IF-elimination to the premise that ∠FDE and ∠EGF are inscribed angles and the premise that if ∠FDE and ∠EGF are inscribed angles then ∠FDE and ∠EGF are equal. While the latter premise is left implicit in the representation of the proof above, the person who wrote the proof presumably thought that '(inscribed angles)' would prompt the reader to retrieve the needed premise from the long-term memory. So far, the mental logic theory provides a plausible account of how the deduction proceeds for this proof. However, there are some questions raised in applying the mental logic theory to certain parts of the proof.

Consider the justification cited for (13). While (12) and 'property of equal triangles' is cited as the justification for (13), all this really gets us is that CN = CM. In addition to this we need that C is on MN (or that C, M, and N are on AB with C between M and N), which is not explicitly stated anywhere. If we look at the diagram it is clear that C is indeed on MN and so with (12) and this information gathered from the diagram, we do get that C is the midpoint of MN. The point is that we need to consult the diagram and infer from it that C must be on MN, so we need to be able to give an account of this process with the mental logic theory.

To see what the difficulty is, we can first consider why it is appropriate to describe the process of getting the premise that C is on MN as an inference. It may look like the process here is just translating from an external representation that represents C as being on MN, i.e. the diagram, to a mental representation of C as being on MN. But, given that the deduction needs to proceed from the antecedent of the theorem to the consequent of the theorem, we need to derive that C is on MN from the antecedent of the theorem. From the antecedent we have that C, M, and N are all on AB. So, all we need to do is infer that C must be between M and N. It so happens that any way we draw FG and ED, we will end up with C between M and N.[44] The issue for the mental logic account is that drawing a diagram and looking at it is not a rule of inference, at least

[44] Historically, there has been some question on whether proofs that rely on diagrams really deserve to be called proofs. But, recently, it has been shown that the use of diagrams is part of a rigorous practice, see (Manders 2008).

as the theory currently stands. So, there are some aspects of the deduction of the theorem from this proof that the mental logic theory is not equipped to deal with.[45]

One possible solution for the mental logic theory comes from some recent work in formal deductive logic. A few formal systems have been developed that codify the logical inferences used in diagrammatic reasoning.[46] The existence of a formal system that contains rules for these inferences seems to indicate that the mental logic theory could in principle handle these cases. What it would require is an addition of rules of inference to the theory. Of course, the relevant question is not whether the mental logic theory can be modified to accommodate examples of deduction like the one above, but rather whether the mental logic theory in the modified form is a good theory. So, the situation calls for some empirical testing. A related question that would eventually need to be addressed is how these rules are acquired. There of course may be alternative solutions that fare better, but the point is that there is a potential solution that can be explored further.

We can now see what issues this proof raises for the mental models theory by trying to apply it. Deducing a conditional on the mental models theory would involve representing the antecedent of the conditional and checking to see if the consequent is true for that mental model. In this case there is essentially just one possibility consistent with the antecedent of the conditional.[47] Of course, as we can see from the proof above, seeing that C is the midpoint of

[45] A mental logic theorist could reply as follows. While it may be true that the above proof relies on diagrams in way that is not superfluous, it is replaceable (I borrow this terminology from (Giaquinto 2007, ch. 5)). We can replace the role diagram plays in getting that C is on MN with a suitable lemma, something that gets us that C is on MN from the antecedent of the theorem via IF-Elimination. But, while the use of the diagram is replaceable, it is apparent that the diagram is used in thinking through the proof. So, our account of the deduction should reflect that.

[46] See for example (Mumma 2010), (Miller 2007), and (Avigad et. al. 2009). Also, see (Hamami and Mumma 2013) for a discussion of applying the mental models account to diagrammatic reasoning.

[47] Arguably there is more than one, but a single mental model should theoretically be able to represent all the different ways we might draw FG and ED. This issue is discussed more in the appendix.

MN is not a simple matter of looking at the diagram and so analogously it is not just a simple matter of looking at the mental model. The issue is that the midpoint is a particular point with respect to M and N. In any interval around the midpoint, no matter how small, there are an infinite number of points. While it may appear that C is the midpoint of MN we cannot simply dismiss the possibility that it is just a point close to the midpoint. This is why we must reason from the antecedent by bringing to bear other facts such as that fact about inscribed angles. This suggests that we would essentially need to reason deductively about the mental model. But the mental models theory holds that deductive reasoning is a process of constructing models, making hypotheses and checking if the hypotheses hold true of the model. So, the issue is that checking if a hypothesis is true of the model can require deductive reasoning about the model.

While it perhaps sounds odd that deduction can be a nested process, it is arguably not problematic. A mental model theorist could point out that arguments can have complex structures and part of the structure of an argument may consist of sub arguments. Thus, we should expect that in some cases there will be deductions within deductions. But, even if this issue is not a problem for the mental models theory, it does raise questions about the steps theorized to constitute deduction. The statement of the mental models theory is simple, there are just three steps to the process. But this way of describing the theory obscures how complex these three steps can be, which is something the above example demonstrates. Just how complex the process can be is also obscured by the fact that most examples used in explicating the theory and testing the theory have a simple structure. The point is just that applying the theory to complex arguments is not a straightforward process and so a more extensive account of the steps involved is needed so that the theory can more easily be applied to mathematics.

In the above example, identity does play a role in the proof, such as in (1) where it is inferred that ∠FDE = ∠EGF. But in some proofs, identity acts as a logical constant. To see an example of this in an explanatory proof, consider the following example from (Lange 2014):

> Take an ordinary calculator keyboard, though without the zero. We can form a six-digit number by taking the three digits on any row, column, or main diagonal on the keyboard in forward and then in reverse order. For instance, the bottom row taken from left to right, and then right to left, yields 123321. There are sixteen such "calculator numbers" (321123, 741147, 951159 . . .). As you can easily verify (with a calculator!), every calculator number is divisible by 37. But a proof that checks each of the calculator numbers separately does not explain why every calculator number is divisible by 37. Compare this case-by-case proof to the following proof: The three digits from which a calculator number is formed are three integers a, $a + d$, and $a + 2d$ in arithmetic progression. Take any number formed from three such integers in the manner of a calculator number — that is, any number of the form $10^5a + 10^4(a + d) + 10^3(a + 2d) + 10^2(a + 2d) + 10(a + d) + a$. Regrouping, we find this equals to $a(10^5 + 10^4 + 10^3 + 10^2 + 10 + 1) + d(10^4 + (2 \times 10^3)+ (2 \times 10^2)+ 10) = 111111a + 12210d = 1221(91a + 10d) = (3 \times 11 \times 37)(91a + 10d)$. This proof explains why all of the calculator numbers are divisible by 37; as a mathematician says, this proof (unlike the case-by-case proof) reveals the result to be "no coincidence" (Lange 2014, p. 10).

In this proof, one cannot see the conclusion follows unless we make the appropriate substitution based on the identity established in the proof. That is to say, in order to see that for some calculator number x, x is divisible by 37, we need to see that there exists an a and d such that $x = (3 \times 11 \times 37)(91a + 10d)$. Since we know that for any a and d, $(3 \times 11 \times 37)(91a + 10d)$ is divisible by 37, with the previously established identity we can substitute x for $(3 \times 11 \times 37)(91a + 10d)$. The inference based on the identity is integral to grasping the proof, which is a necessary condition for gaining explanatory understanding from the proof. So, the mental models and mental logic theory must give an account of this deductive reasoning.

We will consider how each theory would do this in turn. For the mental logic theory, we have two options. The first option is to handle deductions involving identity by using rules already posited, e.g. If-Elimination. For example, in the case of this proof we can theorize that the premise that if $x = (3 \times 11 \times 37)(91a + 10d)$ and $(3 \times 11 \times 37)(91a + 10d)$ is divisible than 37 then x is divisible by 37 is generated and subsequently If-Elimination is applied. Alternatively, we might think that as part of the mental logic there are inference rules for identity. In this case we would infer x is divisible by 37 directly from the premises that $x = (3 \times 11 \times 37)(91a + 10d)$ and that $(3 \times 11 \times 37)(91a + 10d)$ is divisible 37.[48] Whether there are inference rules for identity is a matter to be settled empirically. So, while there is an important question raised by proofs that treat identity as a logical constant for the mental logic theory, it seems like a question we know how to answer.

It is much more difficult to see how the mental models theory would give an account of this deduction. As we saw above, the mental models account handles each logical constant differently. The question is, how should it handle identity? One possibility is to posit that identity plays a role in constructing mental models. When we have an identity, we can construct a new mental model from another mental model by making the relevant substitution. For example, suppose we have a mental model of the premise that for any a and d we have $(3 \times 11 \times 37)(91a + 10d)$ is divisible by 37. Then, learning the identity $x = (3 \times 11 \times 37)(91a + 10d)$ would cause this mental model to be replaced by a mental model of the premise that x is divisible by 37.

One issue with this potential solution is the same issue raised for the solution given to the problem of negation. It seems that we are introducing a rule of inference into the mental model theory, which raises the question of just what the theory is. If the theory is just that mental models play some role in deduction, then this solution is consistent with the theory. But, if the theory is that deductive inference is a process of constructing models, making hypotheses, and checking to see if the hypotheses are true in all models then it seems that introducing what looks like a deductive inference in the model construction phase is inconsistent with the theory. So, even if this is a viable solution, more work would need to be done to make a case for it. Of course, there may be some other satisfactory way to address the issue raised by identity. The

[48] While Identity-Elimination is not problematic, Identity-Introduction, like Or-Introduction, would need to be restricted in some way as for any x we can introduce the premise x = x.

point is just that it is an issue that needs to be addressed for the mental models theory to be applied to deduction in mathematics.

Implications for Proof Writing

In section one, I argued that going through the relevant deduction is a necessary condition for gaining explanatory understanding. Then in section three, I pointed out that we can make predictions, based on the psychological theories of deduction, about which proofs it will be easier to go through deduction for. This is relevant to mathematical practice since, insofar as gaining explanatory understanding is valuable to mathematicians, the easier it is to gain explanatory understanding the better. In this section, I will talk about what the theories predict and what this means for how mathematicians should construct and present proofs.

Above we saw that both theories predict that a deduction for an argument of the form modus ponens takes less time than a deduction for an argument of the form modus tollens. The mental logic theory predicts this because deduction for modus ponens requires less rules to be applied than deduction for modus tollens (at least in PSYCOP, which is the computational model we were considering). For the mental models theory, it is predicted that deduction for modus ponens takes less time than deduction for modus tollens because deduction for modus ponens requires less mental models to be checked than deduction for modus tollens. Plausibly, shorter deductions are in some sense easier than longer deduction. So, we can infer from the mental logic theory that how many rules need to be applied for a deduction is an important consideration when constructing a proof. Similarly, we can infer from the mental models theory that how many mental models a deduction requires is an important consideration.

Modus ponens and modus tollens are logically equivalent argument forms, so knowing that both theories make this prediction is useful with respect to constructing proofs. For a similar example, consider that an argument of the form modus tollens is logically equivalent to an argument of the form modus tollendo ponens (a.k.a. disjunctive syllogism). We can observe that in PSYCOP, two rules need to be applied to the premises of an argument of the form modus tollens to infer the conclusion, specifically the rule Backward-Negation-Introduction and the rule Forward-IF-Elimination. We can also observe that in PSYCOP, only one rule needs to be applied to the premises of an argument of the form modus tollendo ponens to infer the conclusion, specifically the rule Forward-Disjunctive-Syllogism. So, we get the following advice for proof construction from the mental logic theory. Ceteris paribus, when constructing a proof, modus tollendo ponens is a better argument form to use than the logically equivalent modus tollens.

For the mental models theory, deduction for an argument of the form disjunctive syllogism and deduction for an argument of the form modus tollens both require the same number of mental models. So, according to the mental models theory, there is no advantage to be gained by using disjunctive syllogism rather than modus tollens with respect to ease of deduction.

Consider now an argument that proceeds by considering various cases. For example, if I want to prove that for all integers, n, $P(n)$ I might prove $P(n)$ for all $n > 0$, then show, $P(n)$ for $n = 0$, and finally prove $P(n)$ for all $n < 0$. Alternatively, it may be possible to show the if n is odd then $P(n)$ and show that if n is even then $P(n)$. In the former case, three mental models would need to be checked to deduce $P(n)$ for all n. In the latter case, only two mental models would need to be checked to deduce $P(n)$ for all n. So, ceteris paribus, the latter proof would be better than the former proof with respect to ease of deduction, according to the mental models theory.

Unfortunately, depending on what *P* is, it may not be possible to use the latter type of proof rather than the former type. So, it is unclear how useful this prescription is. But the idea of limiting the cases that one needs to consider could in general be useful.

It is difficult to make generalizations based on the mental models account because, unlike with the mental logic theory, the content of the premises matters for deduction. Furthermore, as we saw in section four, even if there is just one mental model for a deduction, verifying that the conclusion obtains in that mental model could take a lot of work. So, even though, ceteris paribus, less models is better, it will not be clear if the ceteris paribus clause is satisfied until we say more about the rest of the deduction. The mental models theory needs to be applied on a case by case basis. While there is much more that could be said about what each theory implies for constructing proofs, I will leave this for future work and turn to the issues of what the theories imply for presenting proofs.

When it comes to presenting a proof, the theories have different implications for what should be accentuated. It follows from the mental logic theory that the presentation of the proof should accentuate the logical structure of the premises of the proof. It follows from the mental models theory that the presentation of the proof should accentuate what possibilities need to be modelled and perhaps how to model them. Consider the proof of the proposition that for any integer, x, if x is even then x^2 is even. We can construct two different proofs of this proposition that reflect the differences in what should be accentuated according to each theory:

(1) Assume x is even. If x is even, then there exists some integer y such that $x = 2y$. So, there exists some integer y such that $x = 2y$. If $x = 2y$, then $x^2 = 4y^2$. So, $x^2 = 4y^2$. If $x^2 = 4y^2$, then x^2 is even. So, x^2 is even. Hence if x is even, then x^2 is even.

(2) Assume x is even. We can represent x as a line segment with length x and x^2 as the square constructed on this line segment. We can then say that an integer is even if and only if half of its corresponding figure also corresponds to an integer. Since x is even, there is

> some integer y that corresponds to half of the line segment x. Given this, we can see that the square on x can be divided into four squares, each with side y. Thus, half the square x^2 corresponds to the integer $2y^2$. Hence, x^2 is even. Therefore, if x is even, then x^2 is even.

The first presentation of makes the logical structure of the proof explicit. The second presentation of the proof makes explicit one way to model the single relevant possibility. If the mental logic theory is correct, then presumably the first presentation of the proof is better with respect to ease of deduction than the second presentation of the proof. If the mental models theory is correct, then presumably the second presentation of the proof is better. So, we can see that the theories make different predictions, and so which theory is correct is relevant to how we should present proofs.

Conclusion

In this chapter, I began by discussing explanatory understanding in mathematics. I then connected explanatory understanding in mathematics to the psychology of deduction by arguing that one must go through the psychological process of deduction to gain explanatory understanding in mathematics from proofs. The upshot of this argument is that for any proof we can gain explanatory understanding for we can go through deduction for. Consequently, proofs we can gain explanatory understanding for are good test cases for psychological theories of deduction. I then considered a couple psychological theories of deduction, the mental logic theory and the mental models theory. For each I presented examples of proofs we can gain explanatory understanding of that raises issues for each theory. While I presented possible solutions for each issue raised, the point is that more work needs to be done on each theory to make it extensionally adequate. Finally, I considered how the psychological theories of deduction could have implications for the heuristics of proof construction and presentation.

In the first chapter I discussed taking an interdisciplinary approach to the philosophy and cognitive science of mathematics. The work in this chapter illustrates how work in the philosophy of mathematics can be applicable to work in the cognitive science of mathematics. Furthermore, it demonstrates how theories of deduction, especially once informed by work in the philosophy of mathematics, can be applicable to the practice of mathematics.

CHAPTER 4: EXTERNAL REPRESENTATIONS AS COGNITIVE TECHNOLOGLY IN MATHEMATICAL PRACTICE

Mathematical practice involves the development and use of many external representations. Diagrams, numeral systems, and formal theories are examples of representations or systems of representations used in doing mathematics. This chapter, broadly speaking, is concerned with what properties of external representations make them good for doing mathematics. There are several questions about the external representations used in mathematical practice worth investigating. What roles do external representations play in mathematical practice? Is there anything distinctive about mathematical representations or their use? Historically, what criteria did mathematicians use to evaluate external representations? What are the good-making features of external representations? I will touch on these questions and more in this chapter.

To motivate this line of questions, we can make two relevant observations.[49] The first observation is that external representations play an integral role in mathematical practice. In fact,

---

[49] Beyond the observations here are some interesting quotes to consider that Cajori put together in his *A History of Mathematical Notation*: "[Lambert Lincoln] Jackson states: 'Any phase of the growth of mathematical notation is an interesting study, but the chief educational lesson to be derived is that notation always grows too slowly. Older and inferior forms possess remarkable longevity, and the newer and superior forms appear feeble and backward. We have noted the state of transition in the sixteenth century from the Roman to the Hindu system of characters, the introduction of the symbols of operation, +, -, and the slow growth toward the decimal notation. The moral which this points for twentieth-century teachers is that they should not encourage history to repeat itself, but should assist in hastening new improvements.'

The historian Tropfke expresses himself as follows: 'How often has the question been put, what further achievements the patriarchs of Greek mathematics would have recorded, had they been in possession of our notation of numbers and symbols! Nothing stirs the historian as much as the contemplation of the gradual development of devices which the human mind has thought out, that he might approach the truth, enthroned in inaccessible sublimity and in its fullness always hidden from earth. Slowly, only very slowly, have these devices become what they are to man today. Numberless strokes of the file were necessary, many a chink, appearing suddenly, had to be mended, before the mathematician had at hand the sharp tool with which he could make a successful attack upon the problems confronting him. The history of algebraic language and writing presents no uniform picture. An assemblage of conscious and unconscious innovations, it too stands subject to the great world-law regulating living things, the principle of selection. Practical innovations make themselves felt, unsuitable ones sink into oblivion after a time. The force of habit is the greatest opponent of progress. How obstinate was the struggle, before the decimal division met with acceptation, before the proportional device was displaced by the equation, before the Indian numerals, the literal coefficients of Vieta, could initiate a world mathematics.'

there are multiple roles that external representations play that are integral to mathematical practice. I'll briefly mention two. The most obvious role is as a means of communication. Mathematics is accomplished as a coordinated effort over time and so requires mathematicians to communicate their ideas to each other. The only means we have of communication (at least at this point) is via external representations. Another role external representations play is as a means of developing new mathematics. External representations can do this because the creation of external representations can enable new ways of thinking. One example of this is the development of symbolic writing, which enabled mathematicians to ask and answer new questions that they did not have the resources to ask or answer before, specifically questions about the structure of algebras (Van Dyck and Heeffer (2012) at least hypothesize this is the case (p. 3)).

While the fact that external representations are integral to mathematical practice is enough to motivate questions about how to evaluate external representations in mathematics, it does not tie this line of questioning into the theme of this dissertation, which involves utilizing the cognitive science of mathematics to gain insight into such philosophical questions and vice versa. The second observation does this. The observation is just that, typically, the external representations used by mathematicians were designed prior to the development of cognitive science and so were designed without any insight into how external representations interact with

---

Another phase is touched by Treutlein: 'Nowhere more than in mathematics is intellectual content so intimately associated with the form in which it is presented, so that an improvement in the latter may well result in an improvement of the former. Particularly in arithmetic, a generalization and deepening of concept became possible only after the form of presentation had been altered. The history of our science supplies many examples in proof of this. If the Greeks had been in possession of our numeral notation, would their mathematics not present a different appearance? Would the binomial theorem have been possible without the generalized notation of powers? Indeed could the mathematics of the last three hundred years have assumed its degree of generality without Vieta's pervasive change of notation, without his introduction of general numbers? These instances, to which others from the history of modern mathematics could be added, show clearly the most intimate relation between substance and form.'" (Cajori 1928, Vol.1, pp. 227 - 229).

the cognitive system based on recent work in cognitive science. So, even if the mathematicians who designed these external representations relied on some explicit or implicit theory of good representation design (which would presumably be based on the experience of doing mathematics and perhaps folk psychology), it may be worth revisiting questions about what makes a representation good in order to take advantage of any recent developments relevant to this issue.[50]

To reiterate, I am interested in what features of external representations make them good for doing mathematics, and how we can then use such criteria to evaluate the external representations currently used in mathematical practice and guide the design of new representations. In the context of this dissertation, more specifically I want to apply cognitive science to the question of what makes external representations good. To that end, I will argue that what I call 'affinity' and 'adoptability' are good-making features of systems of external representations. Roughly, the adoptability of a system of external representations is the extent to which it can be internalized, mentally speaking, and thereby enables the cognitive system to perform relevant computations mentally. The affinity of a system of external representations is the extent to which the external representations works efficiently with the cognitive system in situations in which there is dynamic interaction between the cognitive system and the world while making use of the system of external representations.

The ideas of adoptability and affinity will be made clearer later. But to briefly illustrate, we can see that the Hindu-Arabic numeral system has some degree of adoptability by considering that I can divide 234,567 by 10 mentally, not by thinking about the numbers themselves but just by thinking about how the numeral system works to represent numbers. So, I

[50] A third observation that I will not discuss is how the development of computer graphics and new printing methods has significantly changed the range of external representations that can be developed.

can just reason syntactically to get the correct answer, i.e. 23,456.7. The Hindu-Arabic numeral system seems to have some non-trivial degree of affinity as well. This can be seen when we consider how doing long division can be efficiently carried out on paper by carrying out a dynamic procedure involving mental arithmetic and writing down numerals on the paper following a set procedure.

The usefulness of these concepts is that they enable us to ask questions of the form, given our understanding of mathematical cognition, how adoptable is some given system of external representations? With a full account of the roles a particular representation plays, and a given theory of mathematical cognition we can assess the degree to which a representation has adoptability and/or affinity, perhaps not in a precise way but at least the relative degree of adoptability or affinity. Whether, in practice, affinity and adoptability are important considerations is largely an empirical question. So, I am by no means trying to settle the issue of their importance here. I will consider this chapter successful if I can show these properties are of interest and worth further investigation.

This chapter will proceed as follows. First, I will consider in a bit more detail some of the vital roles external representations play in mathematical practice. Then I will give a bit of background on the development of external representations and their evaluation in the past. Next, I will delve into the cognitive psychology relevant to the task at hand. After that I will consider some recent work that demonstrates how to assess the comparative cognitive effects that systems of external representations can have. Lastly, I will discuss affinity and adoptability after which it should be clear why they are good-making features of systems of external representations for mathematical practice.

Some Roles of External Representation in Mathematical Practice

The purpose of this section is just to highlight some of the roles external representations can play in mathematical practice so that we have some examples beyond the obvious one of communication. It would be valuable to have an exhaustive account of the roles external representations play and the relative importance of these roles in mathematical practice. But, giving such an account is beyond what I want to accomplish here. I will begin by discussing the role external representations play as a means of computation. I will then talk about the role that external representations play as a basis for making conjectures. Next, I will discuss the role that external representations play as a means for reification. Lastly, I will talk about external representations as a means reducing bias in reasoning. This should provide a good variety of examples for what follows.

Some systems of external representations are such that we can create algorithms that carry out computations relevant for doing mathematics. The important feature of these algorithms is that they operate only on the syntactic features of the representation. So, no part of the algorithm requires any thoughts about what is represented to execute. A simple example is using a look-up table to do multiplication. Suppose I want to multiply 356 by 84. I can perform this computation without ever having myself to think about the numbers if I have a multiplication table and addition table for the numbers one through nine, i.e. tables which for each pair of numbers specifies the sum or product. With these tables I can do long multiplication, which for any given step only involves adding or multiplying two numbers between one and nine, which is a calculation that I can just look up on the table. Once I write down the numerals corresponding to my multiplicand and multiplier, each step of the process just involves reading or writing symbols. Once complete I just interpret the resulting symbol and I have the desired answer. Of course, I could also just use a digital calculator which uses an electronic circuit to carry out an algorithm for multiplication on binary numerals.

Multiplication is obviously a rather elementary computation. But this same role gets played in more advanced mathematics too. A more advanced example of a computation with external representations is the use of formal languages and theories in automated theorem proving (or similarly automated proof verification). Formal proofs play an important epistemic role in mathematical practice and automating the creation of them is important because at least sometimes it would not be feasible for us humans to do these computations ourselves. Indeed, digital computers do computations we cannot in practice perform. So, computation is vital for doing some of the mathematics mathematicians want to do.

Another way of conceiving why computation with external representations is useful is that it can give us a way of substituting a question about mathematics with a question about the vehicles of representation. To illustrate just consider that in some sense modern mathematics, with its axiomatic structure, proceeds by demonstrating that some statement can be computed from some other statements. For example, some mathematician wants to know if every even integer greater than 2 can be expressed as the sum of two primes (Goldbach's conjecture). We can replace this question with the question of whether 'Goldbach's conjecture' is provable in first order logic from some given axioms. We have replaced a question that is ostensibly about some abstract objects with a question about the vehicles of representations. As we can perform computations on the vehicles of representations, this gives us another way to try to solve problems in mathematics (though it is presumably quite rare that the substituted question is significantly easier to solve for cases in which the mathematical question is really tough to solve).

Another role that external representations play is as a means of reification. 'Reification', at least as I will use the term here, denotes the act of using something concrete as a proxy for something abstract. The sense in which external representation serve as a means of reification is that they literally give us something concrete that we are able to look at. This in turn enables us to use visualization and spatial reasoning skills to do mathematics. So external representations in this way enables us to use a variety of cognitive resources to do mathematics.

We have already seen an example of how visual reasoning can play a role in mathematics in chapter three. To briefly review, the example was a proof of the Butterfly theorem for which learning the proof via deduction required an inference based on looking at a diagram. So clearly the visual aspect of external representations can sometimes play a role in learning a proof of a theorem. Speaking more broadly, it is hard to imagine doing geometry and mathematics in general without the aid of visualization (and so without iconic external representations).[51]

The next role that external representation play I want to talk about is as a basis for making conjectures. Systems of external representations can serve as a basis for making conjectures due to the systematic aspect of the external representations. The relevant sense of 'systematic' here is that there are ways of building up more complex representations out of simpler ones in such a way that the more complex representation is still meaningful. A simple example of this sort of systematicity is sums of numbers. We can put any numerals on either side of '+' and the resulting string will represent a number.[52] This aspect of external representations

[51] See Giaquinto's *Visual Thinking in Mathematics* for more examples of how what I call the 'reification' role that external representations play in mathematics impacts mathematical practice through mathematical cognition.

[52] Another example is the language of first order logic. We build up formulas out of simpler formulas and the logical connectives. We can build up more and more complex formulas and once an interpretation is given, they will correspond to meaning propositions.

can lead to promising conjectures because patterns in the vehicles of representations can reflect important patterns in what is represented.

For example, we can consider the derivative of a quadratic function. Suppose we look at the derivatives for several quadratic functions. More specifically, we will look at the polynomial equations that specify the functions and their derivatives. By attending only to syntactic features of the equations that represent the functions and their derivatives, we may notice that for quadratic function, which is of the form $f(x) = ax^2 + bx + c$, where *a*, *b*, *c* are real numbers, the derivative is always of the form $f'(x) = 2ax + b$.[53] Consequently, we might make a conjecture that this will always be the case (indeed I recall my first calculus teacher continued to show us examples until we recognized the pattern). Given this example, we can see that external representations can enable us to see patterns in the syntactic features of the representations that reflect patterns in the mathematical objects, properties etc. that are being represented. While observing a pattern in some finite number of cases cannot justify a universal claim in mathematics, it can be the basis of a conjecture and so can guide mathematicians.

The last role of external representations I want to consider is as a means of reducing bias in reasoning. In *Formal Languages in Logic* (2012) Novaes argues that formal languages can have this effect of reducing bias. But the idea can presumably be extended to any system of external representations where computations can be performed using only the syntactic features of the representations. Novaes argument is quite detailed and long. So, I will just give the gist of her idea here. One sort of bias that people have when it comes to reasoning is a bias toward coming to conclusions that the person takes to cohere strongly with their background beliefs and

[53] It is worth noting that it is the visual nature of external representations that enable us to find patterns by looking at them. So, there is a sense in which the role that external representations play as a basis for making conjectures is enabled by their role as a means of reification.

expectations and rejecting conclusion that do not. For example, if we have spent significant time and effort trying to prove that Euclid's parallel postulate follows from Euclid's other postulates, we may mistakenly convince ourselves that what really is a bad argument for this conclusion is a good one. So, this bias can have bad effects and it would be good to reduce the bias.

Formal languages can help to avoid this sort of bias, because we can operate on the representations as mere objects. As Novaes puts it, formal languages can be de-semantified. This is a key feature of the sense in which these languages are 'formal' in mathematical practice (see Novaes 2012 pp. 12 - 14).[54] But just as formal languages are de-semantified, so too can other systems of external representations in mathematics as illustrated by the example of multiplication with a look-up table given above. Sometimes we may be surprised by the result of some computation. But after checking the computation is done correctly, we can confidently accept the result even though it may not cohere with expectations.

I am reminded of the following example. Treat the earth as a sphere and imagine we have wrapped a rope around the circumference of the earth. Now imagine that similar to how telephone wires are sometimes suspended on poles above the ground, we want to suspend a rope 10 feet off the ground all the way around the earth. We want to know how much longer than the rope on the ground that this rope will need to be. If you are inclined to expect that quite a lot more rope is needed, maybe many miles more, you might be very surprised to learn that only about 60 more feet is actually needed. We want to determine the difference between circumferences, $C_{new} - C_{old}$. We have $C_{old} = \pi \cdot D_{old}$ (where D is the diameter) and from the description of the problem we have $D_{new} = D_{old} + 20$. So, $C_{new} - C_{old} = \pi \cdot 20 = 62.8$. Once we have

[54] For example, Novaes points to this quote of Carnap from his *The Logical Syntax of Language*. "A theory, a rule, a definition, or the like is to be called *formal* when no reference is made in it either to the meaning of the symbols (for example, the words) or to the sense of the expressions (e.g., the sentences), but simply and solely to the kinds and order of the symbols from which the expressions are constructed."

everything properly represented, it is trivial to make the calculation. Even if the result is something we would have been disinclined to believe before, once the computation is performed by operating on these symbols, it is easier to accept the conclusion.

There is a sense in which de-semantified external representations are not representations. A vehicle of representation only serves as a representation when it has some content. But this is just a feature of how symbolic representations work, we give them the content and so we can treat them as representations or not. This ability to switch back and forth is what enables them to serve as a means for reducing bias. Due to the fact that formal languages are apt for being treated as lacking semantic content, when we reason with the aid of a formal language, we can focus just on ensuring the conclusions are deductively valid. Similarly, with other system of representations for which there are associated computations.

A Few Observations on the History of Representation in Mathematics

In this section, I will make a few observations about how external representations develop. The observations are only intended to give a very rough sense of the topic. I will give an extended example to provide some motivation for the observations. I will then note a few features of external representations that mathematicians take to be good-making. Lastly, I will briefly consider an example that illustrates how views on whether a representation has a particular good-making feature can change over time.

My observations on the development of external representations in mathematics are mainly based on Cajori's *A History of Mathematical Notation* (1928).[55] While it obviously does not contain any insight into the most recent developments of mathematical notation, it covers

[55] See pp. 336 - 340 of Vol.2 of Cajori's *A History of Mathematical Notation* for his own observations.

quite a bit of history. The main observation relevant to the point of this chapter is that there is no standard method for developing external representations for doing mathematics. Typically, external representations are developed by individual mathematicians. The rationale behind the design is typically not articulated along with its introduction, though often we can make reasonable guesses as to what the rationale is. Subsequently, this use will spread or not depending on how widely the use comes to be known and whether other individual mathematicians decide to adopt the use of the external representation. How widely an external representation becomes known seems to largely depend on how important (or influential) the mathematics that was done and communicated via the external representations is. Mathematics education presumably also has a large role in spreading the use of particular external representations and also in entrenching its use.

As a result of this typical pattern of development, it is common for there to be multiple external representations with roughly the same content. It also can result in an arguably better external representation being less widely used than an inferior one. Also, as we will see in the example below, sometimes due to the way use of external representations spread, an external representation can to some extent be the result of historical accident more so than intentional design.

To illustrate these points, we can use the history of the notation of the square root. Currently, there are two common notations for the square root of a number. The first is '$\sqrt{\quad}$' and the second is '$\square^{1/2}$'. The symbol '$\sqrt{\quad}$' is actually a combination of the symbol '√' and a vinculum, which is just a horizontal line drawn over a group of terms to indicate that they be treated together (functionally similar to how parentheses are used in mathematics). The combination of these two symbols can be traced back to Descartes (Vol, 1, p. 375). There are two main theories for how the symbol ‘√’ developed. One theory is that the symbol is essentially a deformed ‘r’, i.e. someone made the intentional choice to represent the square root function with the symbol ‘r’ (the first letter of the word ‘radix’ which is Latin for root), but over time the convention changed (unintentionally). The competing theory is that the symbol is essentially a deformed dot (the dot turned into a dot with a kind of tail attached and that eventually turned into ‘√’). Both a dot and ‘r’ had been used to represent the square root function prior to the symbol ‘√’ being used, which goes to why there are the competing theories. On either theory though, it is actually through some miscommunication that we get ‘√’ (which is not to say that fact makes ‘√’ bad notation) (Cajori 1928, pp. 324-338).

The development of the fraction ½ as an exponent to represent the square root has a much simpler origin. This notation is traced back to a single individual, namely Isaac Newton (Cajori 1928 pp. 355). This representation has the advantage of making clear the relationship between the square root function and exponential functions. Rather than represent the square root function independently, it is represented as a particular instance of a more general type of function. Furthermore, we can use algebra on the exponents to determine what the composition of exponential functions will be.

We can see the above observations hold for this case. In particular, we have individual mathematicians using various methods to create external representations. One mathematician uses an ‘r’ to represent the square root, presumably because ‘r’ alludes to what is being represented. One mathematician uses a dot, presumably because it is perhaps the easiest symbol to write. However it occurs, we eventually get the symbol ‘√’, which is arguably an improvement over ‘r’ or the dot simply because we do not have to rely on the context of use for interpretation and so it is a more effective means of communication. But, ‘√’ is arguably inferior to the fractional exponent representation as it is part of a system of representations. Even though these representations are not equally good, they both persist. Of course, this one example does not constitute an argument that these observations hold generally, but this example should at least help to illustrate these observations.

I will now consider a few examples of the features of external representations that mathematicians clearly hold are good-making features. The first is wide-spread use, the second is convenience, and the third is epistemic adequacy. Widespread use is just what it sounds like. An external representation has widespread use when a significant proportion of mathematicians use the representation when needed. It is an extrinsic feature of a representation and so we need to do more than consider the representation itself to determine if it has this feature. It can also come in degrees; a particular representation may have more widespread use than another. The importance of widespread use stems from the role that external representations play in facilitating communication. Since this is perhaps the most important role external representations play, and effective communication is difficult if different meanings are attached to the same symbols (in the same context), it is vital to have a commonly used representation so that how to interpret the representation used is common knowledge.

Indeed, while studying mathematics, I have on multiple occasions heard a professor remark that while they personally feel some notation is not good (presumably with respect to other good-making features), they use it in the lesson because it is widely used. To further illustrate the importance of this feature, consider the following quote. Cajori writes,

> No topic which we have discussed approaches closer to the problem of a uniform and universal language in mathematics than does the topic of symbolic logic. The problem of efficient and uniform notations is perhaps the most serious one facing the mathematical public. No group of workers has been more active in the endeavor to find a solution of that problem than those who have busied themselves with symbolic logic - Leibniz, Lambert, De Morgan, Boole, C. S. Peirce, Schroder, Peano, E. H. Moore, Whitehead, Russell. (Cajori 1928, Vol. 2 p. 314).

While Cajori goes on to remark that the logicians were not very successful in standardizing the representations used, the fact that it was goal demonstrates that mathematicians value the property of widespread use.

The convenience of an external representation is the relative ease with which an instance of the representation is produced. Roughly, we can say that an external representation is convenient if it takes little effort and skill to create a recognizable instance of it relative to the alternatives.[56] The convenience of representation is somewhat context dependent as how easy the representation is to instantiate will depend on what materials one has on hand. For example, the development of computer graphics and widespread availability of digital computers has generally made visually complex representation much more convenient than they were previously.

To see that convenience has been valued by mathematicians historically, we can consider how the same convenient symbols are used to represent all sorts of things. Perhaps the most convenient representation is the dot and the dot is used to represent all sorts of things in different contexts throughout history.[57] We can also note the prevalent use of symbols that act as shorthand for less convenient representations. For example, '+' serves the same exact role as 'plus'. It is presumably inevitable that we will have a verbal representation for this operation, so

---

[56] There are several closely related features of external representations we could also consider. For example, we can define the compactness of a representations to be the amount of information contained in a representation relative to its size. Highly compact representations would be valuable for similar reasons to convenient representations.

[57] From (Cajori 1928, Vol. 1 p. 285): "Probably no mathematical symbol has been in such great demand in mathematics as the dot. It could be used, conveniently, in a dozen or more different meanings. But the avoidance of confusion necessitates the restriction of its use. Where then shall it be used, and where must other symbols be chosen? Oughtred used the dot to designate ratio. That made it impossible for him to follow John Napier in using the dot as the separatrix in decimal fractions. Oughtred could not employ two dots (:) for ratio, because the two dots were already pre-empted by him for the designation of aggregation, :A+B: signifying (A+B). Oughtred reserved the dot for the writing of ratio, and used four dots to separate the two equal ratios. The four dots were an unfortunate selection. The sign of equality (=) would have been far superior. But Oughtred adhered to his notation."

we could do without '+'. There may be multiple reasons why '+' is better than 'plus' in some circumstances, but one is certainly that '+' is more convenient than 'plus'.

Convenience is a valuable property mainly because convenient representations require little effort to produce.[58] As the use of external representations are integral to practicing math in multiple ways, convenient representations makes mathematics easier to do. But convenience is interesting in that it can easily be trumped by other concerns. For example, while a dot may be a convenient representation, if it is overused, i.e. used to represent many different things depending on context, then there is an increased cost with respect to interpretation. So, while convenience is a good-making feature, it highlights how various good-making features need to be weighed against each other.

The epistemic adequacy of a representation has to do with the justificatory role external representations can play. We can say roughly that a representation is epistemically adequate if it is a good basis for making a judgment or inference. Epistemic adequacy is important because mathematicians want to gain knowledge and understanding, and these epistemic states are generally agreed to have some justificatory requirements that need to be met. I will illustrate the idea of epistemic adequacy with a couple examples.

Views on the use of Euclidean diagrams in proofs have changed over time. I discuss this more in the appendix so here I will be brief. Euclidean diagrams play a justificatory role in Euclidean proofs, i.e. the proofs involve inferences based on the diagrams. Concerns arose about this use of these diagrams in rigorous mathematical proofs, because there can be particular details in such a diagram that do not hold generally for the conditions in the antecedent of the theorem being proved. So, one could theoretically infer something from the diagram that does

[58] Convenience can also be important for mass producing some external representation of some mathematical work, especially in the past when less printing technology was available.

not follow from the antecedent of the theorem being proved. However, it has more recently been argued that if we restrict the inferences that can be made based on the diagram to those based only on co-exact features, then we will avoid this problem. So, Euclidean diagrams used according to a particular practice arguably are epistemically adequate. The point to take away here is that these shifting views impacted mathematical practice significantly. The shift away from the use of diagrams led to a new axiomatization of geometry. Clearly epistemic adequacy is an important feature of external representations.

Another example of a sort of external representation where epistemic adequacy is a concern is formal theories. Formal theories are used to provide rigorous justification for theorems. For the justification to be epistemically adequate, the formal theory has to be an accurate representation of the informal theory. Ideally, the formal theory will have elements that correspond to the basic concepts, statements, and arguments of the informal theory it is supposed to represent. If the formal theory is an accurate representation then when we derive a formula in the formal theory, we can be confident that we have justification for taking the corresponding statement to be true.

I have considered three features of external representations that mathematicians hold are good-making features. But of course, these are by no means the only features valued by mathematicians. For example, Kenneth Iverson considers several others in his "Notation As a Tool of Thought" (1979).[59] I will of course be arguing below that affinity and adoptability are also good-making features of external representations in mathematics.

[59] This paper was a significant source of inspiration for how I frame my thesis and is definitely worth a read. Another source of inspiration is Andy Clark's "Material Symbols" in *Philosophical Psychology* 19, no. 3 (2006).

A Couple Relevant Concepts from Cognitive Science

In this section, I will briefly consider two related ideas discussed in the cognitive science literature that provide the motivation for this chapter and will be useful for thinking about the claims made in this chapter. The first idea I will consider is that of an epistemic action. I will then go on to consider the idea of cognitive technology. The point here is just to tie the claims and arguments of this chapter into existing work and thought in cognitive science.

In (Kirsh and Maglio 1994), an epistemic action is defined as "a physical action whose primary function is to improve cognition by: 1. reducing the memory involved in mental computation, that is, space complexity; 2. reducing the number of steps involved in mental computation, that is, time complexity; 3. reducing the probability of error of mental computation, that is, unreliability" (p. 514).[60] Kirsh and Maglio illustrate the idea of an epistemic action with an example involving the game Tetris. Tetris is a game in which the player must manipulate a given shape in a limited amount of time by moving it horizontally across the screen or by rotating it, where the goal is to get the shape in the best position before the player runs out of time (it is not important here what constitutes the best position). When a new shape is introduced, skilled Tetris players often begin manipulating the shape well before they could possibly formulate a plan for how best to position the shape. The action of manipulating the shape can actually make the process of determining how to position the shape more effective and efficient than just trying to think of how best to position the shape and then executing the

[60] To put this idea in context, Kirsh and Maglio are responding to the view that rational action is action that alters the world in a way that brings it closer to some desired state of the world. Epistemic action does not meet this requirement yet is rational action and why it is needed to explain the function of some actions. Thus, they introduce this idea in order to argue that it is required to make sense of some rational actions. On the view that Kirsh and Maglio reject, we would expect that a player would formulate some plan of action and then manipulate the shape in the planned way to get it to the desired position. But, this conflicts with observations made of skilled Tetris players.

manipulations necessary to get it into that position. The upshot is that some actions are epistemic in the sense that their purpose is best explained as generating some information to be input into a cognitive process. One of Kirsch and Maglio's innovations is to show, through experimentation, that these epistemic actions can occur in situations like playing Tetris.

We can relate the idea of an epistemic action to the above discussion by looking at the ways that mathematicians use external representations in mathematics. If I create some mathematical representations as a means of communicating some ideas, then this is not a case of performing an epistemic action, creating the representation serves my practical aim of communication. In contrast, creating mathematical representations in the course of performing some computation are epistemic actions because the actions serve to generate some new information for the person performing the action. Similarly, creating the diagram in the proof of the butterfly theorem can be an epistemic action because in some contexts its purpose is to generate some new information for the person performing the action. Furthermore, De Cruz and De Smedt argue that the use of some mathematical symbols should be considered epistemic actions because the symbols enable us to think about some mathematical ideas that we couldn't without the use symbols (De Cruz and De Smedt 2010, p. 15).

The takeaway here is that the fact that producing some external representation(s) are epistemic actions points to the possibility that details about how mathematical cognition work are applicable to what makes a representation good. Of course, considerations of how cognition works in general could be applicable even in cases where the action is not epistemic, e.g. communication. But in the case of epistemic actions that involve producing external representations, in order to evaluate the efficacy/efficiency of the action we need to countenance the impact the representations have on whatever cognitive process follows. So, the fact that producing some mathematical representations are epistemic actions implies that these representations are properly viewed as a tool to aid in mathematical cognition. This brings us to the idea of cognitive technology.

I am not aware of any standard definition of cognitive technology, so I will introduce a few for consideration. Roy Pea defines cognitive technology as “any medium that helps transcend the limitations of the mind (e.g., attention to goals, short-term memory span) in thinking, learning, and problem-solving activities” (Pea, Cognitive Technologies For Mathematics Education, p. 91). Marcelo Dascal defines cognitive technology as “every systematic means – material or mental – created by humans that is significantly and routinely used for the performance of cognitive aims” (Dascal 2002). Andy Clark defines cognitive technology as “any external and/or artificial cognitive aid” (Clark 2001, p. 140). Clark gives the following example involving the behavior of some expert bartenders. Evidently some expert bartenders, when taking a series of drink orders, will set up a series of distinctly shaped glasses. These distinctly shaped glasses act as persistent memory cues that help the bartender to remember the drink orders (Clark 2001, p. 141). In this example, the distinctly shaped glasses act as cognitive technology by aiding in the cognitive process of remembering. But it is unclear whether the distinctly shaped glasses count as cognitive technology under Dascal’s definition because it is unclear whether to satisfy Dascal’s definition the object needs to have been created to serve humans’ cognitive aims or just that it need have been created by humans and has since been used by humans to reach their cognitive aims. It is also a bit unclear whether the example satisfies Pea’s definition as it is unclear whether the being better at remembering drink orders involves transcending the limits of the mind. Another important difference between the definitions is that Clark’s and Pea’s definitions leave it ambiguous whether something that could be used but is not currently used as a cognitive aid counts as cognitive technology whereas Dascal's definition explicitly requires that cognitive technology is actually used to reach some cognitive aim. Although a minor point, it is also unclear why Dascal restricts cognitive technology to things created by humans as opposed to intelligent creatures in general. Neither Clark's nor Pea's definitions have this restriction.

The idea of cognitive technology relates to the current discussion in that in certain contexts we can view external representation used in mathematics as cognitive technology. This follows from the claim that producing some mathematical representations are epistemic actions. Consider that in cases where external representations are produced because they enable us to think thoughts that we wouldn't have been able to otherwise, the external representations satisfy each definition of cognitive technology. Even the less exotic example of using a numeral system for performing some computation show that numeral systems can be considered an example of cognitive technology. The takeaway here is essentially the same as above, the fact that in some contexts there external representations operate as cognitive technology points to the applicability of details of how mathematical cognition works and how the representations work with the cognitive system to the matter of how to evaluate them.

Some Relevant Differences Between Cognitive Processes

The goal of this section is to move to talking in more general and systematic terms about the sorts of cognitive processes that are of interest. When I briefly introduced the ideas of affinity and adoptability above, I mentioned arithmetic computation as a relevant example. Here I want to distill out what it is about a couple ways cognitive processes can perform arithmetic computations that makes them of interest. Using these examples, I will make a distinction between basic and derived cognitive processes. Then I'll introduce a distinction between embedded and non-embedded cognitive processes. Lastly, I will distinguish between pure and mixed cognitive processes. Once again, I will rely on examples involving arithmetic to provide a simple illustration.

Consider a case of mental arithmetic, e.g. someone adding 539 and 54 in their head. We might imagine the following inner monologue: "First I'll take nine and four and add them to get thirteen. So, I will have a three in the ones place. I have to carry the one, so I'll add that to three and five and end up with nine. So, I will have three in the ones place and nine in the tens place. So, we end up with 593." If the cognitive process is structurally similar to our conscious experience of it, then this is clearly an acquired cognitive process. Adding 3 and 5 to get 9 is a subprocess that may be accomplished via association. But the process as a whole relies on using a base ten numeral system since it involves a ones place, a tens place, etc. An algorithm for addition associated with the numeral system is simply internalized. This would be an example of what I call a 'derived' cognitive process. Even though no external representation is used in the course of the cognitive process, the cognitive process makes use of knowledge of the system of external representation in the course of carrying out the process. We can easily contrast these sorts of acquired cognitive processes with innate cognitive process.[61] For example, when I considered the research on numerical cognition in chapter two, we saw that there is a consensus that the ability to estimate quantities is innate. Of course, we can only acquire cognitive processes by bootstrapping up from the innate cognitive processes and so I call the innate cognitive processes 'basic' and the acquired cognitive processes 'derived'.

[61] I do not take these two categories to partition the cognitive processes, there are presumably acquired cognitive processes that don't make use of any knowledge of an external system of representations. I am just using innate cognitive processes here to provide a clear contrast.

To get at the idea of a cognitive process being embedded we can reuse the example of working out long multiplication on paper. In such cases there is a dynamic interaction between the cognitive system and the environment. There is a series of sub-processes, some of which are carried out in the head and some of which involve physical action. If for example I am multiplying 345 and 27, I begin by multiplying 7 and 5 in my head. I then write down the product 35. I then multiply 7 and 4 in my head and subsequently write down 28 below 35 lining up the 8 below the 3, and so on. We can relate these physical actions to the idea of an epistemic action. While any one of these actions of writing down a product may not itself be an epistemic action, they are clearly part of the epistemic action of finding out what addition problem is equivalent to the given multiplication problem.[62] The crucial (and perhaps a bit contentious) claim is that the sub-processes are indeed parts of a more complex cognitive process, doing long multiplication on paper.[63] We can contrast these sorts of cognitive processes with those that do not involve physical interaction with the world, such as purely mental arithmetic. In both cases, we are doing the same sort of computation, but one way of doing the computation inherently involves active physical interaction with the world. I will call cognitive processes where performing the computation requires dynamic interaction with the world *embedded* cognitive processes and the rest will be *non-embedded* cognitive processes.

---

[62] See (Robbins and Aydede 2009) on how the idea of epistemic actions serve as motivation for the idea of embedded cognition. There are some debates in cognitive science about the extent to which cognition in general is embedded and what this means for how cognitive science should be conducted. Fortunately, we need not get into those here as I am only making use of a fairly trivial observation that some cognitive processes involve dynamic interaction with the world.

[63] Some may object that it is incorrect to consider this more complex process a cognitive process. This issue is certainly part of the debate over embedded cognition in cognitive science. Fortunately, it is not a debate relevant to my overall argument. Certainly, there are many processes in the course of doing mathematics that involve dynamic interactions between the cognitive system and the external environment, importantly that involve manipulating external representations. That is all I need for what follows.

If we again consider the example of long multiplication on paper, we can note that the process involves multiple types of representations. It involves the numerals written on paper and whatever mental representations are involved in the mental arithmetic, at a minimum our impressions of the numerals (the details will depend on the theory of numerical cognition). I will call cognitive processes that involve multiple types of representations ‘mixed’ and otherwise they are ‘pure’. It is notable that a cognitive process does not have to be embedded to be mixed. For example, in the LASS theory of cognition considered in chapter two, there are two distinct types of mental representation and both can be involved in a cognitive process. We can contrast this with something like the language of thought hypothesis, under which it seems to follow that only mixed cognitive processes would be embedded. Looking at the relationship between mixed and embedded cognitive processes from the other direction, embedded cognitive processes need not be mixed. The Tetris example considered above is embedded but not mixed, as the epistemic actions do not involve creating or manipulating representations.

What do these distinctions have to do with the affinity and adoptability? Well there are two sorts of cognitive processes that will be of interest going forward, namely mixed embedded cognitive processes and derived cognitive processes. Derived cognitive processes will be the focus when we discuss adoptability and mixed embedded cognitive processes will be the focus when we discuss affinity.

A Few Ways Cognition Can Be More or Less Efficient

In this section, I will talk about three ways that comparable cognitive processes can be more or less efficient. Two different cognitive processes can perform different computations to accomplish the same task. Two different cognitive processes that compute the same thing can be more or less efficient than each other, because they use different algorithms. Two cognitive processes can use the same algorithm on different inputs to perform the same computation and as a result be more or less efficient. I'll give examples to illustrate each of these cases. Efficiency is of course relative, and here we will be evaluating efficiency relative to the various cognitive processes that can accomplish a given mathematical task.

There are fairly simple examples that illustrate how different computations can accomplish the same task. To take an example from philosophical practice, suppose my aim is to advocate for a philosophical position in a debate. I might accomplish this task by making a positive argument for the position. But I could also accomplish this task by crafting an objection to the competing positions or by undercutting an objection to my favored position. Depending on the situation, one of these strategies may be easier than the alternative. Mathematics provides some interesting examples because of the close relationship between distinct areas of mathematics, e.g. algebra and geometry or proof theory and model theory. Due to these close relationships, mathematicians can sometimes answer a difficult question in one branch of mathematics by answering a related question in the different branch of mathematics (using the methods of that branch). One simple example is determining how some geometric figure would appear under some given transformation, say a rotation. We can often accomplish this by simply imagining what it will look like. Alternatively, we can use techniques from the branch of computational geometry (the math behind how your digital computer can do all sorts of stuff with its display) to compute what the geometrical figure will look like. In these sorts of cases, simply imagining what the figure will look like will be much more efficient than carrying out some algebraic computation.

It is also fairly trivial to recognize that different algorithms can be used to compute the same thing and that these algorithms can be more or less efficient. In fact, computer scientists study the efficiency of algorithms and so we can borrow an example from computer science to illustrate. Sorting algorithms are algorithms that put a list of elements in order. Suppose we have the list, [3, 8, 5, 11, 2, 16] and we want to order it from smallest to largest. There are several sorting algorithms we can use, two of which are bubble sort and merge sort. Bubble sort is done by going through the list from left to right, we compare each number with the neighbor to its right and if the number on the right is bigger, we swap them. We continue looping through the list from left to right until it is completely sorted. In the first loop we swap 8 and 5 as well as 11 and 2. In the second loop we swap 8 and 2, and so on. Merge sort is done by breaking the list up into its individual parts and then merging these sublists back together, sorting as we merge. Once we have merged all the sublists back into one, we will have a sorted list. We begin by dividing out given list into its components, [3], [8], [5], [11], [2], and [16]. We begin merging them back together into sorted lists, [3, 8], [5, 11], and [2, 16]. Then [3, 5, 8, 11] and [2,16]. With one last merge we get [2, 3, 5, 8, 11, 16]. While merge sort does require more memory resources than bubble sort, it is generally faster.

In the previous chapter we already saw an example of a case where different inputs into an algorithm lead to differences in efficiency. It has been observed that arguments in the form of modus ponens take less time to go through deduction for than arguments of the form modus tollens. So in this case the relevant computation is deducing the theorem and while we saw there were some competing theories of the psychology of deduction, both theories we considered gave a rough sketch of the one algorithm that is supposed to underlie the psychological process of deduction. To give a simpler example, consider the task of counting the dots in some given picture. In this case all we are doing is mentally attending to each dot once and keeping a count until we have attended to all the given dots. To ensure we can agree that the algorithm is consistent between the two scenarios let's suppose we start at the top most and left most dot and proceed by finding the next top most left most dot and count that one next. If the picture I give you has all the dots sort of randomly placed this will be a lot more difficult and less efficient than if I give you a picture a dots that are in orderly lines.

Assessing the Impact of External Representations on Cognition

We can now turn our attention to the issue of how differences in the systems of external representations being used to complete some task can lead to differences in efficiency of cognition used in completing that task. The discussion will just be an overview of Schlimm and Neth's paper "Modeling Ancient and Modern Arithmetic Practices: Addition and Multiplication with Arabic and Roman Numerals." There are several things this paper accomplishes that is relevant here. It illustrates how there can be a difference in the relative efficiency of performing some computation depending on which external system of representation is used. Notably the point Schlimm and Neth make is that in the case they examine there is not a substantial difference in relative efficiency. But in making this case they demonstrate one way such investigations can be carried out. They also make some interesting observations on difficulties with these sorts of investigations that are worth reiterating here.

Schlimm and Neth present the results of their investigation as debunking the idea that the sort of arithmetic computations, e.g. addition and multiplication, we regularly perform are much more difficult if not impossible to do with Roman numerals. In the context of this chapter, we could consider the similar hypothesis that mental computations using the Hindu-Arabic numeral system are substantially more efficient than those using the Roman numeral system. Schlimm and Neth aim to demonstrate that arithmetic computations using the Roman numeral system are not only possible, but qualitatively are not substantially different from arithmetic computations using the Hindu-Arabic numeral system. They make their case by creating models of agents performing these cognitive processes. Using these models, they simulate the performance of many arithmetic computations (specifically they simulate all additions of three numbers in the range one to one hundred and all multiplications of two numbers in the same range). By measuring various aspects of the simulations, such as the total number of attentional shifts required by the agent, the total number of rules that need to be remembered, the total number of motor actions required, etc. for each sort of computation and numeral system, they can compare the results.[64]

Let's consider in some detail their model of an agent that can perform arithmetic using the Roman numeral system. The task is to add three numbers and the form of the input is three Roman numerals stacked vertically and right-justified, e.g.:

X X V
I I I I
L X X I

The agent computes the sum by creating a working table where each line of table contains one kind of symbol involved in the problem. So, for the above problem the working table has four rows, one for each of 'L', 'X', 'V', and 'I'. The agent writes the symbols in the given problem in the appropriate row:

L
X X X X
V
I I I I I

[64] Schlimm and Neth suggest they will extend their modelling techniques to allow for adding in time measurements in future work, p. 2102.

The agent then applies simplification rules, that are stored in long term memory, to each row in the working table starting at the bottom. In this case ‘IIIII’ simplifies to ‘V’. The agent then writes another ‘V’ on the appropriate row. ‘VV’ then simplifies to ‘X’ and so on until the agent ends up with the final table consisting of just one ‘C’, which is the solution to the problem (and in general one can concatenate the rows from top to bottom in the table to get the answer). This is just a sketch of the algorithm and as Schlimm and Neth point out, how one fills in the details will have an impact on the metrics they measure.

Schlimm and Neth make many interesting observations on how the simulations of the two addition models and simulations of the two multiplication models compare. I will highlight just a couple. For the Roman numeral addition simulations, there is significantly more perceptual, attentional, and motor activity required than for the Hindu-Arabic numeral addition simulations, e.g. 244,166 perceptual actions in the case of the Roman model compared to 8,116 perceptual actions in the case of the Hindu-Arabic model. However, the Hindu-Arabic model requires the agent to store 100 single-digit addition facts in memory, whereas the Roman model does not require the agent to store any addition facts in memory (it does require the ability to apply the simplification rules, which in Schlimm and Neth’s model uses long term memory). So, while Schlimm and Neth are arguing that on the whole there is not a substantial difference in efficiency between the cognitive process that perform arithmetic computations using these two systems of representation, the example illustrates how it could be that there is a substantial difference in efficiency between cognitive processes using two different systems of representations and how we could measure the difference.

Beyond the observations about the simulations, Schlimm and Neth make some notable methodological observations about the sort of factors that need to be considered when setting up a study like this. First, it is important to use a variety of computations because any particular computation may be easier in one system than others. Second, the results of the simulations can depend a lot on the form of the inputs. Third, using different artifacts (other than pencil and paper in their simulation), such as an abacus could have a substantial impact on the results. Fourth, it is important to keep in mind that there can be multiple algorithms for a given system of representations for completing a task. The takeaway here is just that one must be careful how the experiment is set up to ensure that it is not biased against one system of representation. Ideally, we could just simulate and compare all ways of performing a computation for all computations with each system of representations. But since that is not feasible, we must rely on carefully crafted empirical investigations to determine for a given task, the relative efficiency of the applicable systems of representation and associated ways of using them to accomplish the task.

In section four, I discussed what derived cognitive processes and what mixed embedded cognitive processes are. Then in section three, I discussed a few different ways cognitive processes can be more or less efficient. And in this section, I discussed some work by Schlimm and Neth that show how we can evaluate the relative efficiency of using systems of representations to complete mathematical tasks. In the next section we will bring this all together in our discussion of affinity and adoptability.

On Affinity and Adoptability

In this section, I will discuss affinity and adoptability. Recall that the goal of this chapter is to argue that affinity and adoptability are good making features of systems of external representations. The argument consists in showing that systems of representations with these features are generally better for doing mathematics than their alternatives. I'll take on an assumption that a more efficient mathematical practice is better than a less efficient alternative. Suppose it can be shown that systems of representations with a higher degree of adoptability or affinity than the alternatives are generally more efficient for completing the relevant mathematical task in virtue of having that property. Then, given the assumption they are better for doing mathematics than the alternative options, it follows that adoptability and affinity are good making features of systems of representations. So, to make my case that affinity and adoptability are indeed good-making features of systems of representation, it is sufficient to show that they are more efficient for doing mathematics than their alternatives. In particular, I am concerned with cases where the mathematical task involves some cognitive process and the goal will be to show that affinity and adoptability lead to more efficient cognition. As I mentioned above, whether in practice affinity and adoptability are important considerations is largely an empirical question. So, I am by no means trying to settle the issue of their importance here. I aim to show these properties are of interest and worth further investigation.

In the case affinity, since efficiency is built into the idea as I define it here, it will be rather trivial to make the case that systems of representation that have a high degree of affinity are more efficient than the alternatives. So, my focus below will be on discussing what we can expect systems of representation that have affinity to be like. As we will see, there is not some underlying feature such systems of representation must share, which is why we cannot reduce affinity to something more fundamental. In the case of adoptability, it is clearer what systems of representation with adoptability are like so there will be less discussion about that, but the connection between adoptability and efficiency is less clear, so that will be the focus of the discussion. I will begin by discussing adoptability.

We saw that one of the main uses of external representation is as a means of mathematical computation. We saw that different systems of representations come with their own algorithms for performing computations and that these algorithms can be more or less efficient relative to one another. We also considered a couple examples of derived cognitive processes that involved a person carrying out an algorithm using a system of external representations mentally, e.g. mental arithmetic with Hindu-Arabic numerals. Given this context, I will say that the degree to which a system of external representations is *adoptable* is the degree to which its associated algorithms can be successfully carried out entirely mentally.[65] So, adoptability has to do with deriving cognitive processes from a system of external representation. Presumably, for any given system of representation we can contrive at least one simple computation that can be carried out mentally. Therefore, it is better to talk in terms of degrees of adoptability. The issue talk of degrees raises is how we can calculate the degree of adoptability for a given system of representation. For my purposes it is advantageous to leave this relative to the context. Here the context is current mathematical practice. So for a system of representation, the question becomes what typical computations can be done as cognitive processes derived from that system of representation that mathematicians perform in the course of mathematical practice. We can then relate the degree of adoptability to the quantity and usefulness of the computations that can be done as derived cognitive processes.

Conveniently, we can use Schlimm and Neth's investigation to introduce considerations relevant to evaluating the relative adoptability of the Hindu-Arabic numeral system and the Roman numeral system. Schlimm and Neth's model involves motor actions and so is not a model

[65] This of course makes adoptability relative to the cognitive system. Insofar as we are able to come up with *the* theory of human mathematical cognition (or *the* model of human mathematical cognition), we will be able to judge *the* degree of adoptability for a given system of external representation.

of purely mental arithmetic computation. But we can suppose that one is mentally simulating the motor actions and storing the intermediate results in short term memory and consider what implications this would have. From their investigation of simulating all additions of three numerals between 1 and 100, the simulations for Hindu-Arabic addition involved writing 3,637 symbols on paper compared to writing 18,700 symbols on paper for the Roman addition. The difference presumably stems from the fact that addition with the Roman numeral systems, as they model it, requires the construction of a working table. The constraints on short term memory raise the question of whether it is feasible to perform all computations they consider, but if we suppose it is feasible, it is safe to assume that it in general arithmetic computation carried out mentally would be easier with the Hindu-Arabic numerals, just in virtue of requiring much less use of short-term memory.

Since the binary numeral system, like the Hindu-Arabic numeral system, is a place-value system, it is easier to compare those two. The algorithm for addition with binary numerals is simpler in the sense that only two addition facts need to be stored in long term memory. But, the length of the numerals can be so much greater than those involved doing addition with Hindu-Arabic numerals that it is hard to imagine it is feasible to carry out these computations mentally given the constraints on short term memory. My point here is not to claim that the Hindu-Arabic numeral system is more adoptable than the Roman numeral system or binary numeral system, a lot more would need to be done to see whether that is the case.[66] But, these considerations do show how it could be that there are differences in the degree of adoptability between systems of representation that serve the same function.

[66] The main thing to investigate when considering purely mental computation is what the best algorithms for each system would be given how the cognitive system works and then model and compare those. Also, as Schlimm and Neth point out, how the problem is presented can have a significant impact on things.

Let's suppose for the sake of argument that we can determine the degree to which a system of representation is adoptable. Why is it better for a system of representation to be more adoptable? There are a couple different ways adoptability can relate to efficiency. We saw above that different systems of representations come with different algorithms and that these algorithms can be more or less efficient in terms of how many steps involved when carried out mentally. So, there is the possibility for an adoptable system of representation that the associated cognitive process will be more efficient than the alternative in that it requires less steps and presumably is faster. Of course, being a mere possible gain in efficiency, arguably this is insufficient for making adoptability a good-making feature. But there is another relevant sense of efficiency to consider, efficiency in terms of requiring less resources. Adoptable systems of representation enable us to perform computations relevant to doing mathematics without the need for external resources, e.g. pen and paper. So adoptable systems of representation always increase efficiency of mathematical practice in this sense and thus adoptability is a good-making feature of systems of representation. Now that I have made my case for why adoptability is a good-making feature of systems of representations in mathematics, we can turn our attention to affinity.

Above we considered the idea of a mixed and embedded cognitive process and saw how in such processes, external systems of mathematical representation can be used in conjunction with mental representations to perform computations relevant to doing mathematics. We have also seen a detailed illustration of such cognitive processes and how their relative efficiency can be compared in Schlimm and Neth's models. Given all this, I will say the degree to which a system has affinity is the degree to which computations carried out by mixed embedded cognitive processes are efficient relative to comparable systems of representation. It follows that any system of representation used in a mixed embedded cognitive process has some degree of affinity and we can describe Schlimm and Neth as arguing that the Roman Numeral system and the Hindu-Arabic numeral system have a comparable degree of affinity.

Unlike adoptability, there is no distance between the idea of affinity as I have defined it and the question of whether systems of representation that have a higher degree of affinity are more efficient to use than their alternatives. It just follows directly for the definition that this is the case. The downside of defining affinity in this way is that it risks making my claim that having a high degree of affinity is a good-making feature of systems of representation vacuous. It is a true claim, but there needs to be a takeaway beyond the observation that some systems of representations are used in a particular way and that the ones that are more efficient for these uses are better. The value I see in defining affinity is that it gives us a convenient way to frame various questions. Are there any systems of representation that have a high degree of affinity relative to the alternatives? If so, what they are like and what commonalities do they have? While doing the empirical investigation into which systems of representations have affinity is beyond the scope of what I am prepared to accomplish here, there is more to be said about affinity and about what we can expect systems of representation that have affinity will be like.

We can start by considering how Schlimm and Neth's model segments the mixed and embedded cognitive process that makes use of a system of representations to perform addition and multiplication. On the input side there is the reading of the representations and whatever attentional shifts doing so requires and on the output side there are the motor actions of writing new representations and deleting representations. What these input and output processes are like, as well what mental process that occurs between the inputs and outputs is like, will generally depend on the system of representation. Next, we can consider that the number of actions required in each segment will depend to some extent on the system of representation used. For example, addition with Roman numerals required producing more symbols but required less use of long-term memory. Lastly, we can consider how in the case of multiplication with Hindu-Arabic numerals we can distinguish between different sub-processes involved, namely single digit addition and single digit multiplication. Of course, other mixed and embedded cognitive processes could have several distinct sub-processes.

From these considerations it is clear that due to the numerous aspects of these processes that depend on the system of representation used, there are a variety of dimensions along which any particular system of representation could be more efficient to use than another. Thus, systems of representation with high degrees of affinity will likely need to have some balance in the sense that it is unlikely to be enough to just be very efficient for one aspect of the process. Furthermore, we can see why it is difficult to pick out one feature or set of features of systems of representation in virtue of which they have affinity (in which case I would have just called that features or those features the feature of affinity). The thing these systems of representation have in common is that they balance many competing concerns and so there need not be one feature, or some features, that they all have in common.

To illustrate this idea of balance we can consider some different aspects of numeral systems and see what trade-offs occur for basic arithmetic computations carried out as mixed and embedded cognitive processes. If we consider various place-value systems that differ in what base value is used, we can look at the tradeoffs associated with increasing and decreasing the base value by extrapolating from Schlimm and Neth's results. We can compare a base sixty place-value numeral system with a base two place-value system to evaluate the differences. First, we can consider the difference in the number of facts that must be stored and accessed in memory. A base two numeral system will have just a few addition facts and a few multiplication facts, whereas a base sixty numeral system will have many more. But, when it comes to producing the symbols required, the length of the expressions will be much shorter for the base sixty numeral system than for the base two numeral system. So, there appears to be a tradeoff here between ease of producing the required expressions and the number of facts required for the mental portion of the computation. Where we find the ideal balance will depend on comparing the cost of producing expressions to the cost of storing and using addition and multiplication facts.

Following (Hayes 2001), two related aspects of these numeral systems we can consider is how long the numerals can become and how many distinct symbols are needed. For example, in a base ten numeral system, we need ten distinct symbols to form the numerals and the numeral for the number one million is seven places long. For the binary numeral system, we need two distinct symbols to form the numerals and the numeral for the number one million is twenty places long. This pattern continues in that the fewer symbols needed to form the numerals, the greater the number of places needed to represent numbers. As mentioned above, one downside to having long numerals is that they take more effort to produce. It can also be harder to recognize what number is represented by the numeral, though in the arithmetic computations considered this is not an issue as the algorithm only requires recognizing single digits at a time. The downside to needing many distinct symbols to form the numerals is that it will increase the difficulty of distinguishing between the symbols (we could try to make the symbols complex enough that distinguishing between symbols would not be an issue, but this would introduce its own difficulties). Hayes recounts a historical anecdote in which it was theorized that the product of the length of the numeral used to represent some fixed number, say one million, multiplied by the number of symbols needed to form the numerals was related to the relative cost of creating computers that utilized the given numeral system. Given that idea, the prudent course of action would be to minimize that product and design computers that utilized the associated numeral system. The same idea could be employed in the context of assessing the degree of affinity of these numeral systems. Of course, the relative impact of the length of the numerals and the number of symbols in the numerals on efficiency would need to be assessed, but the point is that we could do such a calculation to make such an assessment.

While we have extracted quite a bit from the example that Schlimm and Neth provide, it would be great to move beyond it to other examples. The difficulty in doing so is that it is difficult to say much without having models of the cognitive processes to consider. One case where we did see a partial sketch of some algorithms that could be part of a model is in the context of deduction. So, we can consider what there is to say about the relative affinity of various languages we could use to write proofs in. In the example we will consider, which is admittedly somewhat contrived, one language will be the textual language currently used I mathematical practice and the other will be that same language but will fewer logical connectives. Specifically, we will restrict the logical connectives to just ‘and’ and ‘not’, which is sufficient for expressing everything the current language can express, logically speaking.

In Rips’ mental logic theory, the psychological process of deduction involves constructing a mental representation of a derivation of the inferred proposition by applying syntactic rules on mental representations of the propositions that are the basis of the inference. Consider the following argument in the standard textual language of mathematics,

(1) If ten is even, then two is a factor of ten.

(2) Ten is even.

(3) Therefore, two is a factor of ten.

We can compare this to a logically equivalent argument in the alternative language with ‘and’ and ‘not’ as the only logical connectives.

(1) It is not the case that both ten is even and two is not a factor of ten.

(2) Ten is even.

(3) Therefore, two is a factor of ten.

Based on Rips model PSYCOP, to go through deduction for the first argument, just one rule needs to be applied, namely IF-Elimination. For the second argument, many more rules need to be applied. Even without running simulations of PSYCOP using a wide variety of arguments in each language, it seems safe to infer that in general, more rules need to be applied in deduction when the restricted language is used. To tie this into affinity, we can consider the dynamic process of working out a proof on paper and compare this process with one language to this process with the other. Since all the other aspects of the process are essentially the same, it follows from the fact that the process of working out proofs on paper using the restricted language will be less efficient that the restricted language has a lower degree of affinity than the standard language used in mathematics.

While this particular example is contrived, mathematicians do make choices about how to develop the language of mathematics, such as how to define new terms. How these new terms are defined can impact the structure of mathematical arguments. So, this is an example where research on affinity has the potential to be applicable to mathematical practice.

Conclusion

Mathematicians use systems of representation for various purposes in mathematics. Since one of the chief purposes is communication, it is important that the systems of representation used are widely adopted. This in turn makes it valuable to know what features of systems of representations make them good for doing mathematics, because then we can compare the alternatives and (at least in theory) coordinate our usage of the best ones. In this chapter I have argued that adoptability and affinity are features of some systems of representations used in doing mathematics that are worth considering when it comes to evaluating how a system of representation compares with its alternatives. The argument relies on supposing that more efficient mathematical practice is better than an alternative less efficient mathematical practice. Along with this supposition, I then argued that whether a system of representation has affinity or adoptability is an important consideration when evaluating the efficiency of practices involving the use of these representations compared to the alternatives.

In chapter one, I discussed an interdisciplinary approach to the philosophy and cognitive science of mathematics. Then in chapter two and chapter three I demonstrated a couple ways in which the philosophy and psychology of mathematics are applicable to each other. In this chapter I have showed how this sort of interdisciplinary work can be applicable to mathematical practice. The main idea is that to determine that affinity and adoptability are good-making features and which external representations have affinity or adoptability (or to design ones that do), we need to understand how mathematical cognition works. So, theories of mathematical cognition are in this way relevant to and can have an impact on mathematical practice. Of course, throughout this dissertation I have pointed to or raised many important questions that need to be answered in order for the philosophy and cognitive psychology of mathematics to more concretely inform our views on mathematical practice. But, as the work I have referenced throughout this dissertation show, important progress has and continues to be made in these interrelated fields.

REFERENCES

Ansari, Daniel, and Ian M. Lyons. "Cognitive neuroscience and mathematics learning: how far have we come? Where do we need to go?." *ZDM* 48, no. 3 (2016): 379-383.

Appel, Kenneth, Wolfgang Haken, and J. Koch. "Every planar map is four colorable." *Illinois Journal of Mathematics* 21 (1977): 439-657.

Avigad, Jeremy, Edward Dean, and John Mumma. "A formal system for Euclid's Elements." *The Review of Symbolic Logic* 2, no. 4 (2009): 700-768.

Bara, Bruno G., Monica Bucciarelli, and Vincenzo Lombardo. "Model theory of deduction: A unified computational approach." *Cognitive Science* 25, no. 6 (2001): 839-901.

Barsalou, Lawrence W. "Perceptions of perceptual symbols." *Behavioral and brain sciences* 22, no. 4 (1999): 637-660.

Barsalou, Lawrence W., Ava Santos, W. Kyle Simmons, and Christine D. Wilson. "Language and simulation in conceptual processing." *Symbols, embodiment, and meaning* (2008): 245-283.

Baumberger, Christoph, Claus Beisbart, and Georg Brun. "What is understanding? An overview of recent debates in epistemology and philosophy of science." *Explaining understanding: New perspectives from epistemology and philosophy of science* (2017): 1-34.

Baumberger, Christoph. "Understanding and its Relation to Knowledge." *Epistemology: Contexts, Values, Disagreement: Papers of the 34th International Wittgenstein Symposium* (2011): 16-18.

Benacerraf, Paul. "Mathematical truth." *The Journal of Philosophy* 70, no. 19 (1973): 661-679.

Bonatti, Luca. "Why it took so long to bake the mental-logic cake: Historical analysis of the recipe and its ingredients." *Mental Logic*. Mahwah, New Jersey: Lawrence Erlbaum Associates (1998): 7-23.

Bourget, David. "The role of consciousness in grasping and understanding." *Philosophy and Phenomenological Research* (2015).

Braine, Martin DS, and David P. O'Brien, eds. *Mental logic*. Psychology Press, 1998.

Cajori, Florian. "A History of Mathematical Notations. 2 vols." *Open Court* (1928).

Campbell, Jamie ID, ed. *Handbook of mathematical cognition*. Psychology Press, 2005.

Casazza, Peter G., and Janet Crandell Tremain. "The Kadison–Singer problem in mathematics and engineering." *Proceedings of the National Academy of Sciences of the United States of America* 103, no. 7 (2006): 2032-2039.

Chater, Nick, Joshua B. Tenenbaum, and Alan Yuille. “Probabilistic models of cognition: Conceptual foundations.” *Trends in cognitive sciences* 10, no. 7 (2006): 287-291.

Chater, Nick, and Mike Oaksford, eds. *The probabilistic mind: Prospects for Bayesian cognitive science*. Oxford University Press, 2008.

Chater, Nick, and Mike Oaksford. “The probability heuristics model of syllogistic reasoning.” *Cognitive Psychology* 38, no. 2 (1999): 191-258.

Clark, Andy. *Mindware: An introduction to the philosophy of cognitive science*. New York: Oxford University Press, 2001.

Clark, Andy. “Material symbols.” *Philosophical psychology* 19, no. 3 (2006): 291-307.

Dascal, Marcelo. “Language as a cognitive technology.” *International Journal of Cognition and Technology* 1, no. 1 (2002): 35-61.

De Cruz, Helen. “Numerical cognition and mathematical realism.” *Philosophers' Imprint* 16, no. 16 (2016).

De Cruz, Helen, and Johan De Smedt. “Mathematical symbols as epistemic actions.” *Synthese* 190, no. 1 (2013): 3-19.

de Regt, Henk W. *Understanding Scientific Understanding*. Oxford University Press, 2017.

Dehaene, Stanislas. *The number sense: How the mind creates mathematics*. Oxford University Press, 1997.

Dehaene, Stanislas, and Laurent Cohen. “Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic.” *Cortex* 33, no. 2 (1997): 219-250.

Devlin, Keith. “A mathematician reflects on the useful and reliable illusion of reality in mathematics.” Proceedings of the workshop Towards a New Epistemology of Mathematics, held at the GAP.6 Conference in Berlin, September 14-16, 2006. *Erkenntnis*, Vol. 68, No. 3, May 2008, pp.359-379

Dretske, Fred. *Knowledge and the Flow of Information*. The MIT Press, 1981.

Elgin, Catherine. “Understanding and the facts.” *Philosophical Studies* 132, no. 1 (2007): 33-42.

Feigenson, Lisa, and Susan Carey. “Tracking individuals via object-files: evidence from infants’ manual search.” *Developmental Science* 6, no. 5 (2003): 568-584.

Fodor, Jerry A. *Psychosemantics: The problem of meaning in the philosophy of mind*. The MIT press, 1987.

Frans, Joachim, and Erik Weber. “Mechanistic explanation and explanatory proofs in mathematics.” *Philosophia Mathematica* 22, no. 2 (2014): 231-248.

Gallistel, C. R., and Gelman, R. “Mathematical Cognition.” In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning*. New York: Cambridge University Press, (2005): 559-588.

Ganesalingam, Mohan. “The language of mathematics.” In *The Language of Mathematics*, pp. 17-38. Springer, Berlin, Heidelberg, 2013.

Giaquinto, Marcus. “Mathematical proofs: The beautiful and the explanatory.” *Journal of Humanistic Mathematics* 6, no. 1 (2016): 52-72.

Giaquinto, Marcus. *Visual thinking in mathematics*. Oxford University Press, 2007.

Gowers, W. T. “Mathematics, memory and mental arithmetic.” In M. Leng, A. Paseau & M. Potter (Eds.), *Mathematical knowledge* (2007): pp. 33-58. Oxford, United Kingdom: Oxford University Press

Greco, John. “Episteme: Knowledge and understanding.” *Virtues and their vices* (2014): 285-302.

Grimm, Stephen R. Christoph Baumberger, and Sabine Ammon, eds. *Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science*. Taylor & Francis, 2017.

Grimm, Stephen R. “Understanding.” *The Routledge Companion to Epistemology* (2011).

Grimm, Stephen. “Reliability and the Sense of Understanding.” *Scientific understanding: Philosophical perspectives* (2009): 83-99.

Grimm, Stephen R. “Is understanding a species of knowledge?.” *The British Journal for the Philosophy of Science* 57, no. 3 (2006): 515-535.

Hafner, Johannes, and Paolo Mancosu. “The varieties of mathematical explanation.” In *Visualization, explanation and reasoning styles in mathematics*, Springer, (2005): 215-250.

Hájek, Alan. “Philosophical heuristics and philosophical creativity.” *The Philosophy of Creativity: New Essays* (2014).

Hájek, Alan. “Philosophical Heuristics and Philosophical Methodology.” *The Oxford Handbook of Philosophical Methodology* (2016).

Hamami, Yacin, and John Mumma. “Prolegomena to a cognitive investigation of Euclidean diagrammatic reasoning.” J*ournal of Logic, Language and Information* 22, no. 4 (2013): 421-448.

Harman, Gilbert and Sanjeev Kulkarni. “Review: Models in the Mind.” *The American Journal of Psychology* 127, no. 1 (2014): 129-131.

Hayes, Brian. “Third base.” *American scientist* 89, no. 6 (2001): 490-494.

Iverson, Kenneth E. “Notation as a tool of thought.” *ACM SIGAPL APL Quote Quad* 35, no. 1-2 (2007): 2-31.

Jenkins, Carrie S. *Grounding concepts: An empirical basis for arithmetical knowledge*. Oxford University Press on Demand, 2008.

Johansen, Mikkel Willum. “Embodied strategies in mathematical cognition.” *PhMSAMP Philosophy of mathematics: Sociological aspects and mathematical practices* (2010): 179-196.

Johnson-Laird, Philip N., and Yingrui Yang. “Mental logic, mental models, and simulations of human deductive reasoning.” *The Cambridge handbook of computational psychology* (2008): 339-358.

Johnson-Laird, Phillip N., and Bruno G. Bara. “Syllogistic inference.” *Cognition* 16, no. 1 (1984): 1-61.

Kaye, Kelley L., and T. GR Bower. “Learning and intermodal transfer of information in newborns.” *Psychological Science* 5, no. 5 (1994): 286-288.

Kellman, Philip J., and Martha E. Arterberry. *The cradle of knowledge*. The MIT Press, 1998.

Kersten, Daniel, and Alan Yuille. “Bayesian models of object perception.” *Current opinion in neurobiology* 13, no. 2 (2003): 150-158.

Kelp, Christoph. “Understanding phenomena.” *Synthese* 192, no. 12 (2015): 3799-3816.

Khalifa, Kareem. “Inaugurating understanding or repackaging explanation?.” *Philosophy of Science* 79, no. 1 (2012): 15-37.

Kirsh, David, and Paul Maglio. “On distinguishing epistemic from pragmatic action.” *Cognitive science* 18, no. 4 (1994): 513-549.

Kitcher, Philip. “Explanatory unification.” *Philosophy of science* 48, no. 4 (1981): 507-531.

Kitcher, Philip. *The nature of mathematical knowledge*. New York: Oxford University Press, 1984.

Kvanvig, Jonathan L. *The value of knowledge and the pursuit of understanding*. Cambridge University Press, 2003.

Lakoff, George, and Rafael Núñez. *Where mathematics comes from*. Basic Books, 2000.

Lange, Marc. *Because without cause: Non-causal explanations in science and mathematics*. Oxford University Press, 2016.

Lange, Marc. "Aspects of mathematical explanation: Symmetry, unity, and salience." *Philosophical Review* 123, no. 4 (2014): 485-531.

Larkin, Jill H., and Herbert A. Simon. "Why a diagram is (sometimes) worth ten thousand words." *Cognitive science* 11, no. 1 (1987): 65-100.

Lipton, Peter. "Understanding without explanation." *Scientific understanding: Philosophical perspectives* (2009): 43-63.

Manders, Kenneth. "The euclidean diagram." In *The Philosophy of Mathematical Practice*, Oxford University Press, (2008).

Marchitello, Max, and Megan Wilhelm. "The Cognitive Science behind the Common Core." *Center for American Progress* (2014).

Marcus, Sandra L., and Lance J. Rips. "Conditional reasoning." *Journal of Verbal Learning and Verbal Behavior* 18, no. 2 (1979): 199-223.

Margolis, Eric, and Stephen Laurence, eds. *Concepts: core readings*. The MIT Press, 1999.

Markovits, Henry, Hugues Lortie Forgues, and Marie-Laurence Brunet. "More evidence for a dual-process model of conditional reasoning." *Memory & cognition* 40, no. 5 (2012): 736-747.

Marr, David. *Vision.* New York: Freeman, 1982.

Mejia-Ramos, J.P. and Inglis, M., 2017. "'Explanatory' talk in mathematics research papers." Presented at the Twentieth Annual Conference on Research on Undergraduate Mathematics Education, San Diego, February 23-25th.

Miller, Nathaniel. *Euclid and his twentieth century rivals: Diagrams in the logic of Euclidean geometry*. Stanford: CSLI Publications, 2007.

Mumma, John. "Proofs, pictures, and Euclid." *Synthese* 175, no. 2 (2010): 255-287.

Murphy, Gregory. *The big book of concepts*. The MIT press, 2004.

Newell, Allen. "Reasoning, problem solving and decision processes: The problem space as a fundamental category." *Attention and Performance VIII* (1980).

Novaes, Catarina Dutilh. *Formal languages in logic: A philosophical and cognitive analysis*. Cambridge University Press, 2012.

Núñez, Rafael E. "Mathematics, the ultimate challenge to embodiment: Truth and the grounding of axiomatic systems." In *Handbook of Cognitive Science*, pp. 333-353. Elsevier, 2008.

Núñez, Rafael, and George Lakoff. "The Cognitive Foundations of Mathematics." *Handbook of mathematical cognition* (2005): 109-124.

Paivio, Allan. *Mental representations: A dual coding approach*. Oxford University Press, 1990.

Pantsar, Markus. "An empirically feasible approach to the epistemology of arithmetic." *Synthese* 191, no. 17 (2014): 4201-4229.

Pea, Roy D. "Cognitive technologies for mathematics education." *Cognitive science and mathematics education* (1987): 89-122.

Polya, George. *How to solve it: A new aspect of mathematical method*. Princeton university press, 2014.

Prinz, Jesse. *Furnishing the mind*. The MIT Press, 2002.

Pylyshyn, Zenon W. *Computation and cognition: Toward a foundation for cognitive science*. The MIT Press, 1986.

Ramsey, William M. *Representation reconsidered*. Cambridge University Press, 2007.

Rips, Lance J., Amber Bloomfield, and Jennifer Asmuth. "From numerical concepts to concepts of number." *Behavioral and brain sciences* 31, no. 6 (2008): 623-642.

Rips, Lance J. *The psychology of proof: Deductive reasoning in human thinking*. MIT Press, 1994.

Robbins, Philip, and Murat Aydede, eds. *The Cambridge handbook of situated cognition*. Cambridge: Cambridge University Press, 2009.

Rosa, Luis. "Inferential basing and mental models." *Philosophical Psychology* 30, no. 1-2 (2017): 102-118.

Saidak, Filip. "A new proof of Euclid's theorem," *Amer. Math. Monthly*, 113:10 (2006): 937-938.

Salmon, Wesley C. *Scientific explanation and the causal structure of the world*. Princeton University Press, 1984.

Schlimm, Dirk. “Conceptual metaphors and mathematical practice: On cognitive studies of historical developments in mathematics.” *Topics in cognitive science* 5, no. 2 (2013): 283-298.

Schlimm, Dirk, and Hansjörg Neth. “Modeling ancient and modern arithmetic practices: Addition and multiplication with Arabic and Roman numerals.” In *CogSci*, pp. 2097-2102. 2008.

Schwitzgebel, Eric. “The unreliability of naive introspection.” *Philosophical Review* 117, no. 2 (2008): 245-273.

Searle, John R. “Consciousness, explanatory inversion, and cognitive science.” *Behavioral and Brain Sciences* 13, no. 4 (1990): 585-596.

Simon, Herbert A., and Allen Newell. “Human problem solving: The state of the theory in 1970.” (1972).

Starkey, Prentice, Elizabeth S. Spelke, and Rochel Gelman. “Numerical abstraction by human infants.” *Cognition* 36, no. 2 (1990): 97-127.

Steiner, Mark. “Mathematical explanation.” *Philosophical Studies* 34, no. 2 (1978): 135-151.

Strevens, Michael. “No understanding without explanation.” *Studies in history and philosophy of science* Part A 44, no. 3 (2013): 510-515.

Susac, Ana, and Sven Braeutigam. “A case for neuroscience in mathematics education.” *Frontiers in human neuroscience* 8 (2014).

Schwitzgebel, Eric. “The unreliability of naive introspection.” *Philosophical Review* 117, no. 2 (2008): 245-273.

Tao, Terry. “Pythagoras' Theorem.” *What's New*, WordPress, 14 Sept. 2007, terrytao.wordpress.com/2007/09/14/pythagoras-theorem/.

Trout, J. D. “Scientific explanation and the sense of understanding.” *Philosophy of Science* 69, no. 2 (2002): 212-233.

Van Dyck, Maarten, and Albrecht Heeffer. “Script and symbolic writing in mathematics and natural philosophy.” *Foundations of Science* 19, no. 1 (2014): 1-10.

Van Fraassen, Bas C. *The scientific image*. Oxford University Press, 1980.

Wilkenfeld, Daniel A. “Functional explaining: A new approach to the philosophy of explanation.” *Synthese* 191, no. 14 (2014): 3367-3391.

Wilkenfeld, Daniel A. “Understanding as representation manipulability.” *Synthese* 190, no. 6 (2013): 997-1016.

Xu, Fei. “Numerosity discrimination in infants: Evidence for two systems of representations.” *Cognition* 89, no. 1 (2003): B15-B25.

Xu, Fei, and Elizabeth S. Spelke. “Large number discrimination in 6-month-old infants.” *Cognition* 74, no. 1 (2000): B1-B11.

Young, Richard M., and Tim O'Shea. “Errors in children's subtraction.” *Cognitive Science* 5, no. 2 (1981): 153-177.

APPENDIX: MENTAL MODELS AS REPRESENTATIONS

In this appendix, I will discuss some issues related to the nature of a mental models qua representations. There are two issues I want to address. One issue, already raised in chapter three, is that mental models of negations are in part symbolic representations. But furthermore, this symbolic aspect of mental models seems to play a logical role in deduction. I will consider an objection to this move and the solution offered in its place given in (Rosa 2017). As we will see, Rosa's objection and solution raise a deeper issue with the mental models theory. I will show that some of the work on diagrammatic inference in the philosophy of mathematics could possibly be used to address this deeper issue. The other issue I want to address is the issue of how it is we can have mental models of various mathematical propositions. It is hard to imagine what an iconic representation would be like for some mathematical propositions and so it is hard to see how the mental models theory works for deductions involving such propositions. I will then revisit the notion of S-representation discussed in chapter two and consider how it may help address this issue.

We saw above that the mental models theory has difficulty accounting for deductions in which negation plays a role. To reiterate what the issue is, first consider that in some deductive arguments we have premises of the form *not p*. According to the mental models theory, the first part of the process of deduction is create a mental model of each possibility consistent with each premise. In the case of a premise of the form *not p*, there is one possibility consistent with the premise, namely *not p*. But mental models are iconic representations and we cannot represent *not p* iconically, because there is no state of affairs to represent. The solution considered in chapter three was to represent *p* with a mental model and then the mental model would be marked in such a way that indicates what is represented by the mental model is false.

Rosa's objection to this solution is as follows. The problem is that this proposed solution amounts to admitting that sometimes syntactic processes are required for deductive reasoning and this undermines two of the advantages of the mental model theory with respect to the mental logic theory. One of the advantages of mental model theory is its explanatory scope. It is supposed to be a theory that can account for all deductive reasoning and so is a genuine alternative to the mental logic theory. Yet, the proposed solution requires we posit a mental logic, so the scope of explanation for mental models alone is reduced. The other advantage of the mental models theory that is undermined by the solution is that the theory does not require us, as the mental logic theory does, to posit that people need to learn syntactic rules in order reason deductively. But the proposed solution would require us to posit that a syntactic rule is acquired for deductions involving negation.

Rosa's solution to the problem posed by negation is just to argue that for any proposition of the form *not p* there will be a mental model that can represent it. For example, we can represent the possibility that it is not raining with a mental model that represents the state of affairs as sunny. To make sense of why this mental model would suffice, Rosa introduces a distinction between the explicit content of an iconic representation and its implicit content. In the iconic representation that it is sunny, the explicit content is just that it is sunny, the implicit content is everything else that holds true of the represented state of affairs. Importantly for the case here, it is true that it is not raining.

Above, I raised a potential issue with this sort of solution. The issue was that it seems one could infer from the model which represent the state of affairs as being sunny, that it is sunny. But, we should not (and I presume do not) infer that it is sunny from the premise that it is not raining. So, on this way of handling negation, the mental models theory would seem to predict that people would make inferences that people do not actually make. In response to this objection, Rosa points out that this is an issue the mental models theory faces even in cases in which there is no negation. He gives the following example. Consider a mental model of the premise that the square is to the left of the circle. The mental model will represent the square as being in some particular location with respect to the circle. So, we could infer from the mental model that the square is above, below, or neither above nor below the circle, depending on which mental model we use to represent the proposition. But none of these inferences would be warranted and presumably we would not make these inferences. So, for the mental models theory to work at all we need to posit that people are able to ignore the features of models that are not representative of the intended proposition (Rosa 2017, pp. 110 - 111)

While I think Rosa's response to the objection makes sense, it seems that in solving one issue a deeper concern is raised for the mental models theory. The mental model theory seems to imply that we would make bad deductive inferences regularly and we have not been given any story about why this is not an actual implication of the theory. Furthermore, it is by no means obvious that a story could be told to address the issue and if an adequate story cannot be told then the mental models theory would be implausible. Fortunately, there is some recent work on diagrammatic inference that gives us some evidence that some sort of story could be told, even though it does not show us what it would be.

To see why the work on diagrammatic inference is relevant we can consider some concerns that have been raised about proofs in geometry that involve diagrams. While Euclid's proofs in *Elements* had been viewed by many as a prime example of good deductive inference and a testament to what can be accomplished by it, there are others who have argued that Euclid's arguments in his *Elements* are not rigorous enough to count as proofs. The reason has to do with fact that the arguments rely on diagrams. The concern can be raised that diagrams often contain more information than what they represent. For example, in the diagram in the proof of the Butterfly theorem above, we could have happened to construct the diagram such that FM = 2MD. But it would be wrong to infer that FM = 2MD follows from the antecedent of the theorem, because logically it does not. So, making inferences based on diagrams is not part of a logically rigorous deductive practice and so cannot be part of a truly rigorous proof. But recently, there has been work done to show that Euclidean diagrammatic inference is indeed part of a logically rigorous deductive practice.

In (Manders 2008), Manders makes a distinction between exact (metric) and co-exact (topological) features of a diagram and he then shows that diagrammatic inferences in Euclid's proofs are only made based on co-exact features of the diagram. He argues that if inferences are only made on co-exact features of the diagram then there will be no logically invalid inferences based on the diagram. So, Euclid's proofs are rigorous and so are legitimately considered proofs. While I do not take the exact/co-exact distinction in itself to solve the issue raised for the mental models theory, that a solution was found for a very similar problem is promising. What we need to find is some similar distinction or set of distinctions that we can make about the features of mental models. If we can find such distinctions, then we will have a solution.

I will now address the concern that, since mental models are iconic representations, it is in general unclear that we could have mental models of mathematical propositions. I will first motivate the concern. An iconic representation is a representation whose vehicle is structurally similar to what is represented. An image is an example of an iconic representation. The spatial relations between the parts of the image, and, the spatial relations between the things represented by the parts of the image, may be different. For example, in a digital photograph of two trees, the distance between the parts of the image that represent each tree may be just a couple inches apart while the trees themselves are a couple feet apart. But, structurally speaking, the image and what it represents will be similar in that all the spatial relations in the image will be reduced by the same proportion. Another example of an iconic representation is a number line. A number line iconically represents the ordering relations of numbers as follows. That zero is less than one is represented by zero being to the left of one on the line.

While there are some domains of mathematical propositions for which we do have some idea of how to represent the propositions iconically, for example Euclidean geometry, there are many difficult cases.[67] Consider for example the proposition that three is odd. Whether we can represent this proposition iconically may depend on our ontology of numbers. But certainly there are some accounts of the ontology of numbers for which it appears we cannot have iconic representations of individual numbers. For example, one plausible account of the ontology numbers is that they are essentially positions in the structure that is the natural numbers. On this account, three itself would have no intrinsic structure and so could not be represented iconically independently of the other natural numbers. Given this it seems that we cannot have an iconic representation of the proposition that three is odd.

To address this concern, it is worth considering using the notion of S-representation in place of mental models. To see how replacing mental models with S-representations could work consider the following example. In chapter two, I talked about how mental numeral representations can serve as S-representations of numbers. In this case, three would be

[67] Rosa does give an example that shows the idea of implicit content could be useful in addressing this concern. Rosa writes, "for example, any mental model that contains a representation of two objects and two other objects will also be a model that represents four objects, thus verifying the sentence 2 + 2 = 4." So, the idea is that at least some mathematical statements can be represented implicitly by iconic representations. Even granting that Rosa's account in this case is plausible, it still seems implausible in general that when mathematicians reason deductively they use iconic representation with the appropriate implicit content.

represented symbolically, but the symbol would represent three in virtue of being part of a representation of the natural numbers. So, ultimately there would be an iconic representation involved, i.e. the representation of the structure that is the natural numbers, and this representation would play a role in explaining how cognition of numbers works. But, the representation of any particular number would be symbolic and so we can avoid the concern raised about how we could represent three iconically.

We can view the mental models theory theories as a particular version of a more general claim that deductive reasoning relies on the processing of semantic information. It is in virtue of the fact that the representations involved represent things in the world that deduction is possible. This is contrasted with the mental logic theory, in which the content does not play a role in the deduction, just the form of the propositions represented. We can also observe that iconic representations are a sort of S-representation. So, if we changed the theory such that the role of mental models was played by S-representations, then we would of course have a substantially different theory, but one of a very similar spirit.

The advantage of S-representations is just that there are less constraints on what constitutes an S-representation as compared to an iconic representation. So, there is more flexibility when it comes to what could serve as an S-representation of mathematical objects, properties, relations, and operations. For example, as was discussed in chapter two, the Language and Situated Simulation theory posits a sort of S-representation called simulators that constitute a conceptual system, which can be used to represent propositions. While some further work needs to be done in order to apply the Language and Situated Simulation theory to mathematics, theoretically it would give us an account of S-representations that could be used to represent any mathematical proposition. Furthermore, with S-representations we would retain the advantage of being able to explain why the process of deduction is effective.