THE REPRESENTATIONAL ALIGNMENT HYPOTHESIS OF TRANSFER OF NUMERICAL REPRESENTATIONS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Clarissa Ann Thompson, M.A.

The Ohio State University
2008

Dissertation Committee:

Dr. John Opfer, Adviser

Dr. Vladimir Sloutsky

Dr. Laura Wagner

Approved by

Adviser

Graduate Program in Psychology
Copyright by

Clarissa Ann Thompson

2008
ABSTRACT

Is transfer of learning universally narrow, slow, gradual, and situation specific, or can it also be broad, fast, abrupt, and cross-cut situations? Whether transferring knowledge from one classroom to another, from early in the school year to later in the school year, or from one example to another example, conceptual representations should allow learners to generalize over situations that differ merely in place, time, and superficial details (Murphy, 2002). I hypothesized that narrow transfer of learning—at least in numerical contexts—can occur automatically. In my representational alignment hypothesis, transfer is facilitated not only through the overlap in identical elements across the training and transfer context, but also through participants’ ability to recognize the underlying structural similarity between the two contexts. Numbers are an interesting test case of this theory because surface similarities (e.g., 1 and 7 look more similar than do 1 and 3) as well as relational similarities may be examined (e.g., decimal system; the magnitudes denoted by 1 and 3 are more similar than the magnitudes denoted by 1 and 7). I examined an interesting implication of the representational alignment hypothesis—the implication that representational changes can impose costs as well as benefits for task performance. In my first set of experiments, I investigated the phenomenon of robust transfer and showed that simply highlighting the superficial perceptual details of the contexts in question allowed children to "scale up" a learned linear representation of
numerical magnitude (e.g., when children indicated that 150 is closer to 0 than to 1,000 on a 0-1,000 number line) to increasingly larger numerical magnitude contexts (0-10,000 or 0-100,000). In my second set of experiments, I examined how children estimated salaries denoted in fractional notation on a "money line"—an instance where possessing and employing an automatized linear representation of numeric magnitude led to costs in accuracy. In my final experiment, I investigated a potential experimental artifact common to many previous studies in the transfer literature that I believe helped to explain why researchers so often find that children have trouble with transfer. In sum, the conducted studies demonstrated that transfer is an automatic outcome of generalization, which depends on the similarity of how the training and transfer contexts are represented rather than the similarity of the training and transfer contexts themselves.
Dedicated to my loving father and mother, William & Charlotte
ACKNOWLEDGMENTS

Completing a dissertation is not something that can be done without the help and support of many people. First of all, I would like to thank my adviser, Dr. John Opfer, for guiding me through the dissertation-writing process. I thank John for setting such high expectations for my work, and for piquing my interest in research. I have learned the importance of the question, “So who cares?” and John has taught me how to critically think about my own and others’ claims about cognitive development.

I would also like to thank Dr. Vladimir Sloutsky and Dr. Laura Wagner for serving on my dissertation committee. Vladimir has taught me that there are always two sides to a debate, and I should always be prepared to defend my own theoretical position. I have appreciated hearing Laura’s fresh, female perspective on issues regarding academia, particularly the time she spent chatting with me about my future career options.

Many thanks are owed to my Columbus social support network: Megan Bulloch, Ellen Furlong, Jamie Jackson, Jackie von Spiegel, Anna Yocom, Katie Alexander, Frank Kanayet, and Gewn-hi Park. It was wonderful to go through the ups and downs of graduate school with such a wonderfully talented group of people. I also owe a world of thanks to my Psychology 100 Program supervisor,
Dr. Melissa Beers. I greatly respect Missy’s leadership ability, and she has shown me that there is a way to have it all: family and academic success.

Without the children, parents, teachers, and administrators of the Bentworth school district in Bentleyville, PA, and the various districts in Columbus, OH (Worthington, Upper Arlington, Polaris Christian Academy), I would have been unable to collect the data in this dissertation. I would also like to thank the undergraduate research assistants who helped me collect the child and adult data for my dissertation: Stephanie Heinsons, Jennifer Filicky, Debbie Lurie, and Julia Kennedy.

The last five years of my life would not have been possible without the love and support of my family and fiancé who encouraged me to branch out on my own and pursue my dream of becoming a doctor of psychology. I would like to thank my mother and father for upholding the value of education in our home, and always reminding me of my goals. Finally, I would like to thank my fiancé, Mike Conway, for always lending a truly interested ear to my work and reminding me that it is important to “stop and smell the roses” every once in a while. Mom, Dad, and Mike, I love you all, and I am certain that you will be there for me every step of the way as I continue my career in academia.
VITA

October 15, 1980
Born—Washington, PA

December 2002
B.A., Psychology, California University of Pennsylvania Honors Program

June 2005
M.A., Psychology, The Ohio State University

September 2003—September 2005
Graduate Research Assistant, The Ohio State University

September 2005—December 2007
Graduate Teaching Associate, Introductory Psychology, The Ohio State University

March 2006—June 2007
Introductory Psychology Graduate Student Junior Coordinator, The Ohio State University

June 2007—present
Introductory Psychology Graduate Student Senior Coordinator, The Ohio State University

PUBLICATIONS


**FIELDS OF STUDY**

Major Field: Psychology  
Area of Emphasis: Developmental Psychology  
Minor Field: Cognitive Psychology
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapters:</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 History of the Transfer of Learning Debate</td>
<td>5</td>
</tr>
<tr>
<td>1.2 Representational Alignment and Abstract Similarities</td>
<td>13</td>
</tr>
<tr>
<td>1.3 Criteria for the Representational Alignment Hypothesis</td>
<td>16</td>
</tr>
<tr>
<td>1.4 Numerical Representations as Test Case</td>
<td>20</td>
</tr>
<tr>
<td>1.5 Development of Numerical Representations: The Log-Lin Shift</td>
<td>24</td>
</tr>
<tr>
<td>1.6 Overview of Experiments</td>
<td>30</td>
</tr>
<tr>
<td>1.7 Specific Claims of Representational Alignment Hypothesis</td>
<td>32</td>
</tr>
<tr>
<td>2. Broad Generalization of Linear Representations</td>
<td>36</td>
</tr>
<tr>
<td>2.1 Progressive Alignment</td>
<td>39</td>
</tr>
<tr>
<td>2.2 Analogy as a Mechanism of Developmental Change</td>
<td>40</td>
</tr>
<tr>
<td>2.3 Issues Examined in the Present Studies</td>
<td>43</td>
</tr>
<tr>
<td>2.4 Developmental Trajectory of Log-Lin Switch</td>
<td>46</td>
</tr>
<tr>
<td>2.5 Method</td>
<td>47</td>
</tr>
<tr>
<td>2.5.1 Participants</td>
<td>47</td>
</tr>
<tr>
<td>2.5.2 Design and Procedure</td>
<td>47</td>
</tr>
<tr>
<td>2.6 Results</td>
<td>49</td>
</tr>
</tbody>
</table>

Abstract ............................................................................................................................................. ii

Dedication .......................................................................................................................................... iv

Acknowledgments ................................................................................................................................... v

Vita ....................................................................................................................................................... vii

List of Tables ...................................................................................................................................... xii

List of Figures ....................................................................................................................................... xiii
2.7 Method .................................................................................................................. 54
  2.7.1 Participants........................................................................................................ 54
  2.7.2 Design and Procedure................................................................................. 54
2.8 Results .................................................................................................................... 57
  2.8.1 Training ............................................................................................................ 58
  2.8.2 Generalization .............................................................................................. 59
  2.8.3 Posttest .......................................................................................................... 66
2.9 General Discussion ............................................................................................... 69
3. Non-Strategic Generalization ............................................................................. 72
  3.1 Relation between Estimation Tasks ................................................................. 77
3.2 Method .................................................................................................................... 78
  3.2.1 Participants........................................................................................................ 78
  3.2.2 Number Line Estimation Task .................................................................... 79
  3.2.3 Fraction Line Estimation Task .................................................................... 79
  3.2.4 Design and Procedure ................................................................................. 81
3.3 Results .................................................................................................................... 82
  3.3.1 Age Group Differences in Numerical Estimation ........................................ 82
  3.3.2 Breadth of Age Group Differences ............................................................... 85
  3.3.4 Relation between Estimation Tasks ............................................................... 86
3.4 Method .................................................................................................................... 91
  3.4.1 Participants........................................................................................................ 91
  3.4.2 Tasks ................................................................................................................ 92
  3.4.3 Design and Procedure ................................................................................. 92
3.5 Results .................................................................................................................... 94
  3.5.1 Process of Change in Numerical Estimation ................................................ 94
    3.5.1.1 Source of change ...................................................................................... 94
    3.5.1.2 Rate of change ......................................................................................... 98
    3.5.1.3 Path of change ......................................................................................... 99
  3.5.2 Transfer of Learning to Fractions ................................................................. 101
3.6 Discussion ............................................................................................................. 105
  3.6.1 Effect of Context on Age Differences .......................................................... 106
4. Why Transfer is Rare in the Literature ............................................................... 110
  4.1 The Solomon and Lessac Four-Group Design ................................................. 115
4.2 Method ................................................................. 117
  4.2.1 Participants..................................................... 117
  4.2.2 Number Line Task........................................... 117
  4.2.3 Categorization Task........................................ 117
  4.2.4 Design and Procedure...................................... 118
4.3 Results .................................................................... 120
  4.3.1 Process of Change in Estimation......................... 121
    4.3.1.1 Source of change........................................ 121
    4.3.1.2 Rate of change......................................... 124
    4.3.1.3 Path of change.......................................... 128
    4.3.1.4 Variability of change................................. 130
  4.3.2 Transfer of Learning to Numerical Categorization .... 131
4.4 Discussion ............................................................... 135
  4.4.1 Why Did the Pretest Interfere with Transfer?......... 137
  4.4.2 Implications for Microgenetic Studies.................. 140

5. General Discussion ................................................... 142
  5.1 Summary of Findings ............................................. 143
    5.1.1 Broad Generalization of Linear Representation........ 143
    5.1.2 Non-Strategic Generalization ............................ 143
    5.1.3 Why Transfer is Rare in the Literature ............... 144
  5.2 Support for Representational Alignment Hypothesis ..... 145
  5.3 Broader Implications of Representational Alignment Hypothesis... 149
  5.4 Future Directions................................................ 150

List of References .......................................................... 153
Appendix A: Sex Differences in Magnitude Estimation ............. 164
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Age x scale interaction .......................................................... 51</td>
</tr>
<tr>
<td>2.2</td>
<td>Condition x scale interaction ..................................................... 63</td>
</tr>
<tr>
<td>3.1</td>
<td>Outline of procedures in Experiment 2b ........................................ 93</td>
</tr>
<tr>
<td>3.2</td>
<td>Analysis of age differences at posttest ......................................... 102</td>
</tr>
<tr>
<td>4.1</td>
<td>Experimental groups ........................................................................ 120</td>
</tr>
<tr>
<td>4.2</td>
<td>Rate of change in training phase ..................................................... 126</td>
</tr>
<tr>
<td>A.1</td>
<td>Analyses of sex differences at posttest ........................................... 170</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Log discrepancy .......................................................... 27</td>
</tr>
<tr>
<td>2.1</td>
<td>Illustration of stimuli .................................................. 48</td>
</tr>
<tr>
<td>2.2</td>
<td>Lin and log best-fits of median estimates .............................. 50</td>
</tr>
<tr>
<td>2.3</td>
<td>Percentage of participants best-fit by linear function .............. 52</td>
</tr>
<tr>
<td>2.4</td>
<td>Illustration of experimental conditions ................................ 55</td>
</tr>
<tr>
<td>2.5</td>
<td>Lin and log best-fits of median estimates at generalization .......... 61</td>
</tr>
<tr>
<td>2.6</td>
<td>Accuracy on generalization problems .................................... 63</td>
</tr>
<tr>
<td>2.7</td>
<td>Percentage of participants best-fit by linear at generalization ...... 64</td>
</tr>
<tr>
<td>2.8</td>
<td>Lin and log best-fits of median estimates at posttest ................ 67</td>
</tr>
<tr>
<td>2.9</td>
<td>Percentage of participants best-fit by linear at posttest ............ 68</td>
</tr>
<tr>
<td>3.1</td>
<td>Experimental stimuli ....................................................... 81</td>
</tr>
<tr>
<td>3.2</td>
<td>Long-term changes in numerical magnitude estimation .................. 83</td>
</tr>
<tr>
<td>3.3</td>
<td>Linear best-fit .............................................................. 84</td>
</tr>
<tr>
<td>3.4</td>
<td>Relation between accuracy on NL and FL task ........................... 87</td>
</tr>
<tr>
<td>3.5</td>
<td>Long-term changes in fractional magnitude estimation ................ 89</td>
</tr>
<tr>
<td>3.6</td>
<td>Short-term change in numerical magnitude estimation by condition ..... 96</td>
</tr>
<tr>
<td>3.7</td>
<td>Trial block-to-trial block changes ...................................... 101</td>
</tr>
<tr>
<td>3.8</td>
<td>Short-term changes in fractional magnitude estimation ................ 104</td>
</tr>
<tr>
<td>3.9</td>
<td>Relation between accuracy on NL and FL task on posttest ............. 105</td>
</tr>
<tr>
<td>4.1</td>
<td>Two interactions detectable by four-group design ...................... 117</td>
</tr>
<tr>
<td>4.2</td>
<td>Source of change ........................................................... 123</td>
</tr>
<tr>
<td>4.3</td>
<td>Rate of change ............................................................... 125</td>
</tr>
<tr>
<td>4.4</td>
<td>Path of change ............................................................... 129</td>
</tr>
<tr>
<td>4.5</td>
<td>Breadth of change ........................................................... 133</td>
</tr>
<tr>
<td>A.1</td>
<td>Sex differences in accuracy ............................................... 169</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

A key aspect of cognitive development is the ability to transfer learning from one situation to another. Transfer is the ability to retrieve from memory and effectively use previously learned information to aide in solving related problems at some later point in time. Without the ability to transfer knowledge across related situations or problems, a person would be forced to constantly re-learn information even if very small details differed between the learning and generalization context. For instance, if a teenager learned how to park a car in a mall parking lot and then attempted to park his car in a different lot outside of a grocery store, without the ability to transfer his initial learning, the teenager may not view the two situations as similar and may have difficulty transferring the skill of parking from one related situation to another.

One of the biggest challenges learners face is their attempt to strike a balance between over-extension and under-extension of learning to novel contexts. Participants must ascertain the situations in which extension of newly learned knowledge is appropriate and also correctly identify the scenarios in which the new learning should not be extended (Siegler, 2006). For instance, a child who first learned to play the violin may repeatedly practice his major and minor scales. At some time in the future, when this violinist is expected to perform a brand new concerto in the key of D Major, he should be able to retrieve the knowledge of the scales he had previously practiced and
subsequently apply that knowledge to the concerto to play the appropriate sharp notes (e.g., F sharp and C sharp) despite the fact that he had never before played that exact piece of music. Likewise, early efforts to first teach young students to read may focus on the introduction of “sight words:” one-syllable, common words that generally cannot be “sounded out.” Kindergarten teachers hope that their students will not only recognize these easy words by sight when they are presented on individual note cards in the Kindergarten classroom, but anticipate that these students will recognize the “sight words” in a brand new context, like the bedtime stories their parents read to them at home.

Despite the importance of transferring knowledge—whether from domain to domain, from school to everyday life, or from everyday life to school—spontaneous transfer, transfer that occurs without extensive prompting, is notoriously difficult to elicit. Learners typically generalize new approaches to a much narrower set of problems than is optimal (Barnett & Ceci, 2002; Siegler, 2006; Gick & Holyoak, 1980; Reed, Ernst, & Banajeri, 1974; Spiro, Feltovich, Jacobson, & Coulson, 1991; Bassok, 1990; Bassok & Holyoak 1989; Thorndike, 1922; Thorndike & Woodworth, 1901). This general lack of transfer is not simply because initial learning is incomplete or unstable. Even in microgenetic studies of children’s learning, which allow the stability of new knowledge to be established by obtaining a dense sampling of children’s thinking over time by means of trial-to-trial assessments, children very often under-extend novel solutions (see Siegler, 2006, for a review of 105 microgenetic studies examining the issue). For example, once fourth grade students have learned to solve addition equations in the form: $A + B + C = \_\_\_ + C$, they fail to transfer their new knowledge to
multiplication equations in the form: \( A \times B \times C = \_ \times C \) (Alibali & Goldin-Meadow, 1993).

Numerous explanations have been proposed to identify why transfer is so difficult for learners to achieve. One explanation that has attempted to identify why learners have difficulty transferring knowledge has included the degree of overlap in production rules between the base and transfer domain. As an example, Singley and Anderson (1989) suggested that the amount of transfer of learning one text editor to the next could be predicted by the amount of similar procedural elements the two editors had in common. Second, where attention is directed during learning may also impact transfer (Anderson, Reder, & Simon, 1996). Third, a lack of surface or structural similarity between the base domain and transfer domain may lead to a situation where transfer is non-optimal. For instance, Gick and Holyoak (1980; 1983) found that participants’ ability to transfer problem-solving solutions from a military to medical problem greatly increased when the participants were given a hint that the solution from one problem might help them solve the other problem. Likewise, transfer was facilitated when participants were explicitly told about the relation between a problem where missionaries had to be transported safely by boat in the presence of cannibals and an isomorphic problem pertaining to jealous husbands and wives (Reed, Ernst, & Banajeri, 1974). Fourth, the stability and organization of initial learning may also impact the success of transfer across situations. For example, by thinking of roses, trees, and ferns as members of the superordinate category \textit{plant}, children can generalize novel information (e.g., life status) about any one of these plants to other plants and not to non-plants (Opfer & Siegler, 2004). Finally, insufficient mapping of elements from the base to the target
domain may also inhibit participants’ ability to transfer their learning. Young children heard a story about a genie who rolled up a magic carpet to transport jewels from one bottle to another, and were then asked to solve a similar problem of transporting gumballs from one bowl to another with the help of a rolled up sheet of paper (Holyoak, Junn, & Billman, 1984; Brown, Kane, & Echols, 1986). Indeed, the many possible impediments to transfer have led some investigators to posit the somewhat pessimistic conclusion that learning is universally narrow and “situation specific” (Lave, 1988; Cognition and Technology Group at Vanderbilt, 1997).

In this dissertation, I propose a mechanism of transfer—the representational alignment hypothesis—that can account for many of these obstacles to transfer. This theory posits that transfer is an automatic outcome of generalization, which depends on the similarity of how the training and transfer contexts are represented rather than the similarity of the training and transfer contexts themselves. Specifically, if there is a high degree of similarity between representations (in elements, structure, and degree of activation), transfer across tasks will be facilitated, and performance will be comparable in the training and transfer contexts. An important implication of this view is that transfer can occur not only in situations that contain identical elements across the training and transfer contexts, but it can also occur in situations where an underlying structure unites contexts that possess no identical elements. Although this depiction of transfer is similar to previous accounts (e.g., Gick & Holyoak, 1980; 1983; Gentner, 1983), the wider theory that transfer results from aligning representations of learning and transfer tasks also generates a number of novel (and surprising) predictions that have not been explored previously, especially
about the development of numerical cognition, where children often possess multiple representations that co-exist and compete for use. These predictions concern the conditions that facilitate changes in the representation of numerical magnitude (Experiment 1), conditions under which changes in numerical estimation will transfer inappropriately (Experiment 2), and conditions under which transfer can be blocked (Experiment 3).

1.1 History of the Transfer of Learning Debate

One of the first researchers in the history of psychology to systematically study transfer of learning was Thorndike (Thorndike & Woodward, 1901; Thorndike, 1922). Thorndike strongly disagreed with the prevailing educational theory at the turn of the century: the doctrine of formal discipline (Thorndike & Woodward, 1901; Singley & Anderson, 1989). According to the doctrine of formal discipline, if a person attempted to discipline the mind by studying difficult subjects like Latin or Geometry, these attempts should strengthen the mind’s general ability. In this way, the mind might be thought of as analogous to a muscle that could be strengthened through mental exercise. According to the doctrine of formal discipline, the content of the mental exercise was not as important as the exertion of mental effort that strengthened the general faculties of the mind like observation, attention, discrimination, and reasoning (Singley & Anderson, 1989).

In one experiment Thorndike and Woodworth (1901) conducted to combat the theory of formal discipline, participants were asked to judge the area of rectangles that ranged from 10-100 square centimeters. In an attempt to facilitate the participants’ performance during training, participants could reference three comparison squares with known areas. During testing, participants were asked
to estimate the area of novel rectangles within the range of 20-90 square centimeters, and to judge the area of completely new geometric shapes. Since participants made such a large percentage of errors when judging the area of the new geometric shapes, Thorndike and Woodworth concluded that simply teaching a participant to judge the area of rectangles does not improve that participant’s general ability to judge the area of other shapes. This led Thorndike and Woodworth (1901) to claim,

“Improvements in any single mental function rarely brings about equal improvement in any other function, no matter how similar, for the working of every mental function-group is conditioned by the nature of the data in each particular case” (p. 250).

Thorndike’s seminal studies investigating transfer led him to formulate an identical-elements theory: “Any disturbance whatsoever in the concrete particulars reasoned about will interfere somewhat with the reasoning, making it less correct, or slower, or both” (Thorndike, 1922, p. 36). Thorndike believed that transfer between diverse skills was possible only if identical elements united those skills (Detterman, 1993; Singley & Anderson, 1989).

“One mental function or activity improves others insofar as and because they are in part identical with it, because it contains elements common to them. Addition improves multiplication because multiplication is largely addition; knowledge of Latin gives increased ability to learn French because many of the facts learned in the one case are needed in the other (Thorndike, 1906, p. 243).

Thorndike applied an associationist perspective to learning and believed learning occurred when a specific response was paired with a particular stimulus. “In the same organism the same neurone action will always produce the same result, in the same individual the really same situation will always produce the same response” (Thorndike, 1903; p. 7).
However, the general capacity for transfer seemingly must exist, even if transfer is not as perfect as hoped. After all, there are very few instances where two situations are truly identical. It is often a person’s psychological perception that deems two situations as similar (Singley & Anderson, 1989). As Meiklejohn (1908), an educational psychologist, keenly observed,

“What can we say of a theory that the training of the mind is so specific that each particular act gives facility only for the performing again of that same act just as it was before? Think of learning to drive a nail with a yellow hammer, and then realize your helplessness if, in time of need, you should borrow your neighbor’s hammer and find it painted red. Nay, further think of learning to use a hammer at all if at each other stroke the nail has gone further into the wood, and the sun has gone lower in the sky, and the temperature of the body has risen from the exercise, and in fact, everything on earth and under the earth has changed so far as to give each new stroke a new particularity all of its own, and thus has cut it off from all possibility of influence upon or influence from its fellows” (p. 126).

Indeed, consistent with Meiklejohn’s conjecture, researchers have also found cases of symbolic learning that are remarkable for the robustness and breadth of the learning obtained. Given only a small hint, people transfer solutions across situations that have no identical elements (Gick & Holyoak, 1980; Reed, Ernst, & Banejeri, 1974). Moreover, as indicated in work by Biederman & Shiffrar (1987), transfer from very brief explicit lessons to real-world situations is at least sometimes as great as transfer from many years of real-world experience. For instance, when college students with no prior experience sexing day-old chickens were trained to recognize the perceptual feature indicative of the chicks’ sex, their accuracy on this difficult task was comparable to chicken-sexing experts who had sexed millions of chicks over an extended career. Likewise in another famous study, students given an abstract lesson on light refraction immediately transferred the lesson to improve their accuracy in throwing darts
at an underwater dartboard even when the depth of the dartboard was raised closer to the water’s surface (Judd, 1908; Hendrickson & Schroeder, 1941). As these cases of transfer attest, the chief psychological problem is not whether or how widely transfer does or does not occur, but why—given the capacity for transfer—learners so often have trouble with transfer, such as those reported as being widespread in microgenetic studies of learning and transfer (Siegler, 2006).

Thorndike and Meiklejohn represent two extreme poles on the transfer continuum. Thorndike claimed that transfer is naturally narrow, and if there were any hope that learning might transfer from one situation to the next, this transfer would arise due to elements common across the situations. Meiklejohn, on the other hand, was more optimistic and claimed transfer of learning to be naturally broad. To sum up the state of the transfer debate, Singley and Anderson (1989, p. 25) ask, “What then, is the current status of the notion of general transfer? Is it dead, or very much alive?”

It is clear that researchers are split over whether transfer is broad and robust, or more narrow and situation-dependent. As an attempt to more clearly characterize the circumstances under which people successfully access and apply information learned in a previous problem context, some researchers have created what are known as transfer taxonomies (Barnett & Ceci, 2002; Chen & Klahr, 2008). According to Barnett and Ceci (2002), there are two general types of transfer: near and far transfer. Near transfer occurs when the training context and the transfer context are highly similar. Far transfer, on the other hand, occurs when there is less overlap, or similarity, between the base and target domains.

Due to the fact that the terms ‘near’ and ‘far’ transfer are so ill defined (e.g., overlap and similarity may be defined differently from one study
investigating transfer to the next), there is no surprise that there has been little agreement in the literature as to the exact nature and breadth of transfer. Barnett and Ceci’s (2002) taxonomy helps to qualify the near/far transfer continua and shed some light on the reasons why transfer is likely to occur in some situations but not in others. These researchers believe that there are two main dimensions that pin-point whether a particular study observes near or far transfer: 1) the nature of the skill to be transferred (e.g., procedure, representation, heuristic), how the performance across the base and target domain will be measured (e.g., speed, accuracy), and the memory demands required for performance on the transfer task (e.g., recall, recognition), and 2) distance between the training and transfer contexts (e.g., knowledge domain, physical context, temporal context, functional context, social context, modality) (Barnett & Ceci, 2002).

As a further refinement to Barnett and Ceci’s (2002) taxonomy, Chen and Klahr (2008) focused on three dimensions: 1) task similarity (e.g., shared superficial and structural features between two domains), 2) context similarity (e.g., how similar are the contexts in which the solutions and strategies are acquired and used), and 3) time interval (e.g., how short or long is the gap of time between the training and transfer contexts). Chen and Klahr (2008) claim that the degree of transfer that occurs must be inversely proportional to the time delay between the training and transfer contexts and directly proportional to the similarity between the training and transfer contexts. These authors suggest that many previous research experiments claiming to investigate transfer had a limited focus where only superficial similarity between the two situations was manipulated, and context and time dimensions were simply ignored (Chen & Klahr, 2008).
Barnett and Ceci’s (2002) as well as Chen and Klahr’s (2008) taxonomies clearly draw upon Thorndike’s notion of identical elements because the taxonomies attempt to characterize the psychological distance between the training and transfer contexts. What neither of these taxonomies takes into consideration though, is the underlying structural similarity between the training and transfer context and the way the training and transfer task are mentally represented, which my representational alignment hypothesis does address.

In the following paragraphs, I will detail two studies, one characteristic of near transfer and one characteristic of far transfer. One seminal example of near transfer is Gick and Holyoak’s (1980) study on the transfer of analogical approaches to problem solving. Adult participants were first presented with a story about a military officer who decided to use a ‘divide-and-conquer’ technique with his men to simultaneously converge on a fortress after learning that the roads that radiated out from the fortress were mined such that the weight of the entire army would detonate the mines. The participants in the experiment were provided with this base story, and then the goal of the experiment was to see how effective the participants were at abstracting the analogy that related the military problem to the transfer problem—Duncker’s (1945) radiation problem.

In Duncker’s radiation problem, a doctor is attempting to treat a patient’s inoperable stomach tumor through the use of an intense ray. If the doctor uses the ray at its highest intensity, the healthy tissue will be damaged. At low intensities, the ray is harmless to the healthy tissue and it will not impact the malignant tumor. When participants were explicitly reminded to consider the
military solution when attempting to solve the medical problem, their problem-solving success rate drastically increased.

According to Barnett and Ceci’s (2002) taxonomy, this is a good example of near transfer because 1) little temporal delay separated the presentation of the military problem’s solution and participants’ production of solutions for the medical problem, 2) both the base and target domains were administered in a lab setting, and 3) the task in both the base and target domain was administered by paper-and-pencil and was completed by individual participants as opposed to a group of participants. Though it is true that the military and medical problem had very few similar surface features, the underlying structure of the two problems was nearly identical. According to Thorndike’s identical elements view of transfer, transfer should be minimal in this experiment because of the lack in overlap of surface features (e.g., military general vs. surgeon; fortress vs. tumor, etc.). Yet, when participants received a simple clue that highlighted the underlying structural similarity, or analogy, uniting the two problems, transfer was facilitated. Thorndike’s theory cannot successfully account for the transfer that occurred in this experiment, though my representational alignment hypothesis of transfer can account for the transfer.

Research by Chen and Klahr (1999) provide a good example of an experiment that reports on the effects of far transfer. In this experiment, children were taught the control of variables strategy (CVS). The CVS taught children to design valid physics experiments across a variety of contexts (e.g., slopes, springs, and sinking designs) where the experimental variables were un-confounded with one another. Chen and Klahr’s empirical question investigated whether training a child in one physics domain (e.g., inclined planes) would
allow him to create un-confounded experiments in another physics domain (e.g., springs). The reason the experiment is considered to be a representative example of a far transfer study (Barnett & Ceci, 2002) is because the children in the experiment were tested after a seven-month period of delay on a paper-and-pencil posttest that was administered in the children’s classroom, whereas the training session involved children physically manipulating the variables in the experiment (e.g., slope of inclined plane, surface material of the inclined plane, etc.) and occurred in a laboratory setting (Chen & Klahr, 1999). According to my representational alignment hypothesis, children successfully transferred their knowledge across physics problems because they were able to abstract the general schema of the control of variables strategy and apply it across a wide array of problems by drawing an analogy across the underlying structural similarity in each of these problem contexts.

In summary, the breadth of transfer has been debated since the turn of the century when the psychologist, Thorndike, first began to systematically study how broadly people generalize their knowledge. Thorndike was in opposition to the educational theory, the doctrine of formal discipline, that claimed disciplining the mind in one topic area should strengthen it for optimal performance in others. Thorndike thought that transfer was more elusive, and that transfer was closely tied to the amount of overlapping elements between the base and target domains. Yet, there have been a great number of experiments investigating transfer since Thorndike’s time that have uncovered broad transfer despite a lack of overlapping elements. Recently researchers have formulated “transfer taxonomies” in an attempt to tease out the circumstances under which transfer is likely and unlikely to occur. In this dissertation I will focus on the
instances in which there are no identical elements in the training and transfer domain, yet transfer appears to broadly occur. I call this theory the representational alignment hypothesis, and I provide the details of this theory in the next section.

1.2 Representational Alignment and Abstract Similarities

In this section, I will detail a general theory of transfer that I developed—the representational alignment hypothesis. This transfer hypothesis draws upon elements of Thorndike’s theory of identical elements, but it makes some important departures from the theory as well. For instance, my hypothesis addresses how transfer can successfully occur even when there are no identical elements present between the training and transfer domains.

The basic premise of my representational alignment hypothesis is that transfer is simply generalization that occurs automatically. Generalization is the act of perceiving similarities or relations between different stimuli. The closer the stimuli are in psychological space, the more similarly they will be perceived and the more likely generalization will occur across the stimuli (Shepard, 1987). I hypothesize that generalization can occur automatically at two levels: 1) generalization over surface similarities (e.g., overlapping features), and 2) generalization over abstract similarities (e.g., underlying abstract structural relations; Medin, Goldstone, & Gentner, 1993).

One example of generalization over surface similarities is Pavlov’s experiments that examined dogs’ digestion. In these classic learning experiments, Pavlov’s dogs were actually classically conditioned to salivate whenever they heard a bell ring. The bell was predictive of the meat powder the dogs were about to receive. Stimulus generalization occurred when the dogs also reflexively
salivated in response to hearing similar sounding bells, such as those of different pitches. The dogs generalized across the different sounds, ignoring the differences in pitch and focusing on the similarities of the bells, leading to a conditioned association between the sound of the bell and the arrival of the food powder.

On the other hand, Gentner (1977) provides one example of generalization over structural or relational similarities. In this experiment, children were shown a picture (e.g., tree or mountain) and asked, “If this tree had a _____ (body part), where would it be _____? For instance, the child might be asked to point to where a tree’s stomach would be located on a schematic drawing of a tree in different spatial orientations. Children were very successful at mapping their knowledge of the relation between human body parts (e.g., the head is above the shoulders) and the location of these parts on objects like trees or mountains (e.g., if a tree had knees they should be located above the tree’s feet).

Because it is my assumption that transfer and generalization are interchangeable terms descriptive of essentially the same phenomena, it follows that findings from research on generalization informs the understanding of research on transfer. As highlighted in the two examples of generalization I provided, generalization can occur under circumstances where there are identical elements in the base and target domains, and also under circumstances where there are no identical elements, but an underlying structural similarity that relates the base and target domains. In accordance with Thorndike’s theory of identical elements and Shepard’s (1987) view of similarity, my hypothesis does not dispute that as the number of identical elements between the base and target domains increases, transfer becomes faster and more frequent. Shepard’s
mathematical model provides further evidence that the probability of
generalization approximates an exponential decay function based upon the
distance in psychological space between the base and target domains, or the
amount of feature overlap. Though one major departure from Thorndike’s
identical elements theory is that my representational alignment hypothesis posits
that generalization can also occur over relational similarities (e.g., Gentner, 1983),
so therefore, transfer does not necessitate the presence of identical elements.

Thinking about transfer as a type of generalization is important because it
allows for a reevaluation of the near vs. far transfer taxonomy (Barnett & Ceci,
2002; Chen & Klahr, 2008). For instance, near transfer is traditionally thought of
as generalization of knowledge to situations that have very many identical
elements in common with the original learning scenario, whereas far transfer has
been characterized as generalization of knowledge to situations that have no
identical elements with the original learning situation. Thus, two very different
situations potentially fall on the “far” side of the transfer continua when the
variable of relational similarity is ignored: 1) a situation with identical relations,
but no identical elements, and 2) a situation with no identical relations and no
identical elements. I would like to argue that the situation where the base and
target domain have no relational similarity or identical elements is a situation
where transfer seems very unlikely, whereas the situation where there are at
least identical relations underlying the structure of the base and target domains
bodes much better for successful transfer.

Thus, the representational alignment hypothesis posits that there are
conditions under which broad and narrow transfer alike should be anticipated.
The hypothesis suggests that participants automatically deploy their abstract
representations when solving tasks in the base and target domains. That means, participants do not think consciously about the deployment of their abstract representations. I hypothesize that it is an impossible feat for participants to be consciously aware of the representation that they are using in a particular task. In instances where 1) participants possess the same abstract representation across the original learning context and the transfer context, likely due to a substantial overlap in elements common to the two tasks, and 2) this abstract representation is appropriate for both the training and transfer context, I hypothesize that transfer will be broad from situation A to B. When a participant’s abstract representation of the original learning context differs from their abstract representation of the transfer context, likely due to a lack of featural overlap between the training and transfer context, I hypothesize narrow transfer of learning will occur. That is, in circumstances that lack identical elements, participants may not abstract an underlying structural relationship between the training and transfer context, and therefore less than optimal transfer may occur. I hypothesize that participants align the commonalities between the base and target domain (either the similar elements or the similar relations), judge whether the tasks belong to a set of situations entailing the same consequences, and subsequently engage in the same behaviors across both of those tasks (Shepard, 1987).

1.3 Criteria for the Representational Alignment Hypothesis

What evidence might lend support to the representational alignment hypothesis? I discuss the criteria for this transfer hypothesis in the next sections.

First, the hypothesis holds that learners are not consciously aware of the abstract representations generating transfer. That is, learners do not strategically
apply their representations to a given task. Consequently, learners cannot optimize their performance for accuracy. In contrast, if generalization and transfer were strategic, breadth of generalization would be optimized for accuracy.

Second, these automatically deployed representations may be modified with implicit training that aligns the underlying structures of the training and transfer domains. Transfer of learning over problems that contain non-identical elements may only occur with implicit feedback if participants draw upon the underlying relational similarity of the problems.

Third, training that highlights the irrelevant surface similarities/features or irrelevant relations between the training and transfer context will distract from learning about the relevant surface similarities/features or relevant underlying structural relationships. An example from the literature on cross-mapped objects (Ratterman & Genter, 1998) is a good example of what I am trying to illustrate. Identical object matches inhibited young children from using relational structure in a spatial mapping task where choosing a relational match would have led to accurate performance. In the task, there were two sets of objects that were arranged in order according to increasing size: the child’s set (e.g., house, cup, car) and the experimenter’s set (e.g., flowers, house, cup). A sticker was hidden under one of the objects (e.g., house) in the experimenter’s set, and the child was told to find the other sticker by “looking in the same place” in their own triad. Without the use of relational language (e.g., “Daddy,” “Mommy,” and “Baby”), preschoolers had grave difficulty inhibiting the perceptual object match (e.g., choosing the house in their triad) even though the children were given feedback as to the correct location of the sticker on each trial (e.g., cup in child’s triad).
Fourth, if one assumes that transfer occurs automatically with representational alignment, either surface similarity of the experimental context or the internal state of the learner may affect transfer of learning. Evidence for the impact of both variables can be found in the literature. For instance, accuracy has been tied to context-dependent factors in the experimental setting (e.g., location where training and transfer tasks are administered; Godden & Baddeley, 1975). Likewise, state-dependent factors (e.g., mood congruency effects; Bower, 1981) may also impact transfer of learning. The more overlap in context- and state-dependent factors across the training and transfer domains, the more successful transfer one is likely to observe. The less overlap in these elements, the less likely successful transfer will be observed.

By simply manipulating the perceptual similarity of the training and transfer tasks (e.g., the elements involved in these tasks) the representation that is automatically deployed by a participant could be impacted (e.g., the more elements common across the problems, the more likely transfer will be facilitated because the participant will represent the problems similarly). I consider the mental representation of the task at hand to be a part of a participant’s internal state that may impact transfer of learning. Specifically, the level of activation of the participant’s recently used representation may impact their task performance. If participants complete a task, they were able to do so because they mentally represented the task at hand in some way. If these participants were later asked to complete the same task again, the way they previously represented the task would be “fresh” in memory (depending on the amount of time that has elapsed), and this representation should be more highly activated than any other representation that the participants might choose to use (cf. social
psychology’s availability heuristic or mental set literature; Duncker, 1945; Luchins, 1942; Luchins and Luchins, 1950). Transfer of learning can be negatively impacted when an inappropriate mental representation becomes entrenched during unsupervised practice on a task. Then, this inappropriate representation will be highly activated, and therefore more likely to be used each time participants come across this task in the future.

Fifth, if a participant automatically deploys an inappropriate representation that does not optimize for accuracy during unsupervised practice on a task, like on a pretest where participants receive no feedback from the researcher, it is likely the case that the next time the participant encounters this exact task, like at posttest, the participant will again automatically deploy his practiced, yet inappropriate, representation. This will hold true even in a circumstance where during training, a participant is explicitly taught to represent a related task in a way that optimizes accuracy.

One interesting implication of the hypothesis that transfer of conceptual representations occurs automatically is that this hypothesis ties together the discrepant literature on transfer. For instance, researchers like Thorndike propose that transfer is rather narrow, slow, situation-specific, and occurs gradually over extended amounts of time, whereas theorists like Meiklejohn believe that transfer of learning is fast, abrupt, abstract, and cross-cuts various situations and contexts. My representational alignment hypothesis posits that because transfer is an automatic effect of abstract representation, representational changes can impose costs as well as benefits leading to unavoidable setbacks in the course of learning that may persist for many years and across many contexts. Therefore, my hypothesis pertaining to the transfer of learning leaves room for
costs, like transfer of an automatic representation from a context where the representation is appropriate for the task at hand to a situation where the automatic deployment of the same representation is actually detrimental to accuracy because it is an inappropriate representation for that transfer context.

1.4 Numerical Representations as Test Case

Numerical representations used in estimation contexts will serve as a test case for my representational alignment hypothesis. The types of numerical estimation contexts that I focus on will be tasks of whole number estimation, fractional magnitude estimation, and category judgments. As will be elaborated on in the Development of Numerical Representations: Log-Lin Switch section, the primary estimation task that I will focus on is the number line estimation task where participants will be shown a line flanked by either whole numbers or fractions, and will be asked to estimate the position of a whole number or fraction on the line. In the category judgment task, participants will categorize numbers with the help of the category labels: “really small,” “small,” “medium,” “big,” and “really big.” Estimation of numerical magnitudes is an important test case of my representational alignment hypothesis because it allows me to investigate the way participants mentally represent approximate magnitudes. For instance, if a participant labeled the numbers 150 and 625 as “big” numbers, then this is a glimpse at the underlying numerical representation that characterizes the way the participant thinks about numbers.

Why did I decide to focus on the transfer of numerical representations? Numerical representations are an interesting test case for the representational alignment hypothesis of transfer because numbers are abstract symbols. This is important because my hypothesis posits that it is not simply the concrete surface
similarities across the training and transfer task that might lead participants to assess the two contexts as similar enough to perform the same on the two tasks, but the underlying structural similarity of the two tasks may also lead to successful transfer. I hypothesize that participants are non-consciously analyzing the overlapping features and/or structural similarities in the task by comparing the way they mentally represent each of the tasks. If the two tasks are mentally represented in a similar fashion, this will lead to similar performance across the two tasks.

What exactly do I mean by stating that numbers are abstract entities? The Arabic symbols that denote numerals are arbitrary, but these Arabic symbols represent the quantity or magnitude of the numerals. For instance, seeing the Arabic numeral “9,” reading the word, “nine,” or viewing 9 daisies in a field should all activate the same underlying numerical representation. Because numbers can be represented abstractly, the following two statements are equivalent: 1) 5 is more than 3 and 2) five > three. It is important to note, however, that Arabic numerals may also be thought of as concrete objects that may or may not have surface features in common with one another. For instance, the surface similarity of 1 and 7 is much higher than is the surface similarity between 1 and 3 (e.g., straight edges vs. rounded edges), yet the quantities denoted by 1 and 3 are more similar than the quantities denoted by 1 and 7. So, surface similarity and abstract similarity can actually be placed in opposition to one another. Since numbers are concrete objects, participants can compare across their surface similarities; likewise since numbers are also abstract, participants can compare across their underlying structural similarities.
Just as the Arabic numerals are abstract, so are the numeric mental representations that are associated with these arbitrary symbols. For instance, “oneness” should be represented the same regardless of whether a person is thinking about one ice cream cone or one cocker spaniel. Likewise, abstract numeric representations possess relational properties, or the relations among the various numeric representations. For instance, the abstract numeric relationship between two sets of objects—like two ice cream cones and six ice cream cones (e.g., the set of two cones contains less items than the set of six cones; in the set of six cones, there are exactly four more ice cream cones than in the set of two cones, etc.)—is constant regardless of the objects to which a person is referring.

A second, more practical, reason why I studied the transfer of numerical representations is because previous research, which is elaborated on in the Development of Numerical Representations: The Log-Lin Switch section, indicated evidence of representational change in the domain of numeric representations. For instance, research suggests that young children initially think about numbers in a logarithmic manner (e.g., children overestimate numbers at the low end of the range, and compress their estimates for numbers at the high end of the range, much like the spacing on a logarithmic ruler), whereas adults think about numbers in a more linear manner (e.g., there are equal, linear spaces between consecutive numbers across the entire numeric range).

With increasing age and experience, children’s thinking shifts and they begin to represent numbers in a more adult-like manner. That is, representational change signals that there is learning to be had in the domain of number. It may be the case that some previous studies that set out to study transfer of learning uncovered poor transfer because no learning actually occurred during the
training. One cannot transfer unless sufficient learning in the base, or original learning domain, occurred. Evidence of representational change is important to my representational alignment hypothesis of automatic transfer because it indicates that learning of a new numeric representation occurred at some point over development or was due to specific experiences to which the participant was exposed during the experimental session.

A final reason why numerical representations serve as an optimal test case for my representational alignment hypothesis of transfer is based on previous research that I also elaborate on in the Development of Numerical Representations: The Log-Lin Switch section. This research indicated that there are high correlations in performance and broad transfer across various estimation tasks. These results suggest that many estimation tasks are actually tapping similar numeric representations. Thus numerical representations are prime candidates for a theory of automatic transfer since it seems that many numerical estimation tasks are composed of similar surface (e.g., Arabic numerals flanking number/fraction lines) and underlying structural features (e.g., decimal system is the same regardless of 0-100 or 0-1,000 numerical magnitude estimation context). Though correlations between numeric estimation tasks have been uncovered, no mechanisms underlying these correlations have been investigated. Another goal of the representational alignment hypothesis is to posit a mechanism, automatic transfer of abstract representations, such as numerical representations, when the surface features and/or underlying structural similarity overlaps in the training and transfer contexts.
1.5 Development of Numerical Representations: The Log-Lin Switch

In this section, I will detail the literature on children’s modification of their default numeric representations with increasing age and experience that lends support to the claim that learning of new numerical representations occur.

Across a wide range of tasks, children normally improve their expectations about the magnitudes denoted by symbolic numerals. For example, on a number line estimation task, children are presented with a series of lines flanked by a number (e.g., 0 and 1000), a third number above the line (e.g., 230), and no other markings. When asked to estimate the position of this third number, children’s estimates of the positions of the numbers would ideally increase linearly with the actual value of the third number, thereby reflecting representation of the ratio characteristics of the formal decimal system. In fact, however, children’s estimates do not increase linearly—at least not initially. On 0-1,000 number lines, sixth graders’ estimates increase linearly, but second graders’ estimates increase logarithmically (Siegler & Opfer, 2003). On 0-100 number lines, second graders’ estimates increase linearly (Geary, Hoard, Nugent, & Byrd-Craven, in press; Geary, et al., 2007; Siegler & Booth, 2004; Siegler & Opfer, 2003), whereas kindergartners’ estimates increase logarithmically (Siegler & Booth, 2004). On 0-10 problems, kindergartners’ estimates increase linearly, whereas preschoolers’ estimates increase logarithmically (Opfer, Thompson, & Furlong, 2007). Moreover, on an inverse position-to-number task, where children are asked to assign a number to a position on the number line, children’s estimates increase as an inverse of the logarithmic function (Siegler & Opfer, 2003). Finally, evidence for logarithmic-to-linear shifts is not unique to number
lines. Parallel changes have been found in the estimation of real objects, money, answers to arithmetic problems, measurements of novel units, and the categorization and comparison of symbolic numbers (Booth & Siegler, 2006; Laski & Siegler, 2007; Opfer & DeVries, in press).

Theoretically, these initial expectations that numerical magnitudes increase logarithmically are interesting because they are consistent with Fechner’s Law, in which the apparent magnitude of a quantity increases logarithmically with actual value. This logarithmic representation of numerical magnitude is apparently widespread across species and age groups and is consistent with the quantitative performance of time-pressured adults, young preschoolers, human infants, and such non-human animals as pigeons, rats, and monkeys (Brannon, 2005; Banks & Hill, 1974; Feigenson, Dehaene, & Spelke, 2004; Moyer & Landauer, 1967; Gallistel & Gelman, 1992; Roberts, 2005; Xu & Spelke, 2000). In these groups, too, the difference between 1 and 10 seems larger (or is more quickly detected) than the difference between 101 and 110, much as these numbers would be spaced on a logarithmic ruler. Moreover, single-cell recordings in the parietal cortex of monkeys show a similar pattern of neural activation, with the numerical tuning functions of neurons that fire to large sets (e.g., 8 or 9 dots) showing greater signal overlap than neurons firing to small sets (e.g., 1 or 2 dots) (Nieder, Freedman, & Miller, 2002; Nieder & Miller, 2003).

Thus, it appears that the natural mental number line is logarithmically scaled, unlike the decimal system that children must eventually learn in school (Dehaene, Dehaene-Lambertz, & Cohen, 1998).

How does the logarithmic-to-linear shift in numeric representations take place with increasing age or experience? An important mechanism implicated in
previous work on number line estimation (Opfer & Siegler, 2007) is children’s use of analogy to structure their generalization of “log discrepant” information. According to this view, children normally encounter information that does not match their logarithmic representation of numerical magnitudes (e.g., hearing 150 referring to a relatively small part of 1,000 items). If children already apply linear representations in some numerical contexts (e.g., for small numeric ranges), such experiences of log discrepancy may lead them to draw analogies between the two contexts and to extend the linear representation to numerical ranges where they previously used logarithmic representations. For example, if a second grader is shown that her estimate of the position of 150 on a 0-1,000 number line is too high, and also is shown the correct position of 150 within that range, she may draw the analogy “150 is to the 0-1,000 range as 15 is to the 0-100 range.” This analogy may lead her to rely on a linear representation for the 0-1,000 range on subsequent estimation problems.

What types of experiences would be most likely to stimulate such an analogy? Opfer & Siegler’s (2007) log discrepancy hypothesis predicts that if children are using a logarithmic representation, then the magnitude of change in their estimates in response to feedback should be positively related to the discrepancy between the logarithmic and linear functions for the problems on which the children receive feedback. The discrepancy between a logarithmic and a linear representation of the values on a 0-1,000 number line (with both functions constrained to pass through 0 and 1,000) is illustrated in Figure 1.1. As the figure shows, the difference in estimates varies as a function of the number presented. The maximum difference occurs at 150, where the logarithmic representation predicts an estimate of 725 and the linear representation predicts
an estimate of 150, resulting in a discrepancy of 575 (57.5% of the line). For purposes of comparison, the absolute numerical discrepancy between the estimates predicted by the linear and logarithmic representations of both 5 and 725 is 228 (22.8% of the line).

Figure 1.1 Log discrepancy: The discrepancy between a logarithmic and linear representation of numeric values on a 0-1000 number line is greatest at 150; the discrepancies for 5 and 725 are equal to each other and about half as great as that at 150 (from Opfer & Siegler, 2007).

Consistent with the log discrepancy hypothesis, Opfer and Siegler (2007) demonstrated that larger discrepancies between children’s estimates and the linear function (e.g., feedback on the magnitude of 150) were more likely to
provoke representational changes than smaller discrepancies (e.g., feedback on the magnitudes of 5 and 725). Indeed, after a single feedback trial, the best fitting function switched from logarithmic to linear for 85% of children in the 150-feedback condition and for more than half of all children who received feedback. Moreover, once children’s estimates first conformed to the linear function, the linear model continued to provide the best fit on more than 80% of subsequent trial blocks, regardless of the problems that led to the apparent switch of representations.

In school-aged children, the change from logarithmic to linear representations of numerical quantity is neither unique to microgenetic studies (e.g., Opfer & Siegler, 2007) nor to the number line estimation task. The timing of development in number line estimation coincides with parallel logarithmic-to-linear changes in numerosity estimation (i.e., generating a set of approximately N objects), in measurement estimation (i.e., drawing a line that is approximately N-units long), and in number categorization (i.e., categorizing N as “very small, like 0” to “very large, like 1,000”), with consistent individual differences emerging across all four estimation tasks (Booth & Siegler, 2006; Laski & Siegler, 2007). Linearity of number line estimates (measured by the $R^2$ value of the best fitting linear regression function) also correlates strongly with other tests of school children’s understanding of numerical magnitudes, including with speed of magnitude comparison (e.g., deciding whether 4 is greater than 6; Laski & Siegler, 2007), with learning of solutions to unfamiliar addition problems (Booth & Siegler, in press), and with overall math achievement on standardized tests ($r$’s typically between .50 and .60; Siegler & Booth, 2004; Booth & Siegler, 2006).
These strong correlations between number line estimation and mathematical proficiency likely stem from the fact that they both rely on children’s representations of magnitude, which are strongly associated with symbolic numbers and thus easily activated by them. Thus, when numerals appear either on a number line task, an arithmetic problem, or on a computer screen, it is difficult for children to inhibit the magnitudes associated with the numeric symbol, regardless of whether these magnitudes intrude on accuracy of task performance (Besmer & Coltheart, 1979; Berch, Foley, Hill, & Ryan, 1999; Opfer & DeVries, in press; Opfer, Thompson, & Furlong, 2007). For instance, when participants were asked to choose the physically larger of two numerals presented side-by-side, participants were slower to respond when presented with trials where the physically larger number was actually the smaller magnitude (e.g., 3 vs. 9 where the 3 was physically larger). The authors suggested that the participants had difficulty inhibiting the numerical magnitudes associated with the symbols in order to choose the physically larger numeral as instructed (Besner & Coltheart, 1979). These automatic magnitude activations, of course, are also beneficial, in that they aid children in generating approximately correct answers to arithmetic problems and in more swiftly rejecting errors that differ largely in magnitude from the correct answers (Ashcraft, 1992; Siegler, 1988). This positive relation between accurate representations of numerical magnitude and math achievement is also an important reason that the Standards published by the National Council of Teachers of Mathematics have consistently recommended that improving estimation skills be made a high educational priority (e.g., NCTM, 2000).
1.6 Overview of Experiments

Across three experiments, I will provide evidence for the automatic, unmonitored deployment of numerical representations that lead to transfer of learning in the domain of number. The purpose of my first experiment is to determine just how broadly the linear representation will generalize without direct feedback. First, I will examine the type of representation (e.g., linear or logarithmic) second grade students, sixth grade students, and adults use when estimating large magnitudes on 0-1,000, 0-10,000, and 0-100,000 number lines to track the trajectory of the logarithmic-linear shift across the lifespan. Since previous research (Siegler & Opfer, 2003) indicated that second grade students provide a logarithmic series of estimates in the 0-1,000 range, this age range will be the target of a subsequent training intervention in my first experiment. The intervention with second grade students will aim to motivate representational change (e.g., logarithmic-linear shift) within an experimental setting by highlighting the underlying relational structure of the training (0-100 number lines) and transfer tasks (0-1,000, 0-10,000, and 0-100,000 number lines).

The hallmark of my representational alignment hypothesis of transfer is that it predicts benefits in automatic deployment of representations as well as costs. In my second experiment, I will investigate an instance where possessing and employing a linear representation of numeric magnitude may actually lead to costs as well as benefits in the context of a fractional magnitude task where children were asked to estimate salaries on a “money line” ranging from $1/1 minute to $1/1,440 minutes.

Despite the fact that transfer of conceptual representations tends to be rather broad and automatic, in some situations transfer of learning is rather
difficult to illicit. What might be the trouble with transfer? In my third experiment, I will investigate a potential experimental artifact that has plagued many transfer studies already reported in the literature: pretesting of the transfer task (see Siegler, 2006 for a review of narrow transfer in the context of 105 microgenetic studies). Specifically, I will provide children with unsupervised practice on a pretest where they deploy an inappropriate representation in their attempt to solve the task, and then I will test them again on this same task at posttest. I will investigate how pretesting a transfer task may actually block later transfer of a more appropriate representation learned during a training session.

This series experiments will contribute to the transfer literature in general and also to the development of representations for use in numerical estimation literature in particular. The transfer literature has offered many reasons why some researchers find robust transfer, whereas others uncover more modest transfer. My representational alignment hypothesis will shed some light on the reasons that researchers are finding such discrepant results in their investigations of transfer. Namely, conceptual representations, like the linear representation, are deployed automatically, and this is advantageous when the base and target domains require the same representations (e.g., number line estimation and number categorization), but is detrimental when the representation used in the base domain is an inappropriate representation for the target domain (e.g., fractional magnitude estimation). Therefore, robust transfer, when the representation is appropriate for the transfer task, as well as narrow transfer, when the representation is inappropriate for the transfer context, can be accounted for and are predicted by my representational alignment hypothesis.
1.7 Specific Claims of Representational Alignment Hypothesis

Earlier, I described the general criteria for the representational alignment hypothesis. In this section, I will detail how my proposed numerical estimation tasks will directly test the representational alignment hypothesis.

The first claim that the numerical estimation tasks will test pertains to the quickness of responses between the training and transfer contexts. It is hypothesized that participants will quickly make whole number and fractional magnitude estimates. Similarly, participants will not overly contemplate when asked to make category judgments. I hypothesize that participants will have grave difficulty inhibiting their practiced numerical representation (e.g., linear for adults and older children and logarithmic for younger children). For instance, if an older child is presented with a number line estimation task, she will employ her linear representation in the service of estimating the location of numbers on this line. If she received no further intervention before being presented with another numerical estimation transfer task directly thereafter, she should automatically employ her linear representation to this numeric transfer task as well. If the linear representation is not an appropriate representation for the task at hand (e.g., fractional magnitude estimation), the participant will not consciously consider ways to make her performance more optimal. Instead, accuracy on the task will suffer.

There are very few instances in development where it is hypothesized that younger children can outperform older children or adults on cognitive tasks. What is novel about the representational alignment hypothesis is that it predicts accuracy advantages for younger children in some estimation situations. For instance, if a young participant automatically employs his logarithmic
representation in an estimation task where this is a more optimal representation to use than an older participant’s automatic linear deployment, one would expect to see the younger participant outperform the older participant in terms of accuracy. As an extension of this criterion, it would also be possible to assume that one sex (e.g., males) might be more likely than the other (e.g., females) to employ a linear representation, and thus, when a linear representation is suboptimal for a particular estimation task, one might expect rampant sex differences in the task at hand.

The second claim that the numerical estimation tasks will test pertains to the ease with which numerical representations can be modified with training. I hypothesize that both explicit and implicit information can lead to representational change across the numerical estimation tasks that I conducted. In Chapters 3 and 4, I will provide participants with direct, explicit feedback that capitalizes on the maximally discrepant point between a logarithmic and linear function in the hopes of inducing representational change (e.g. logarithmic-linear switch). In Chapter 2, I will provide more implicit information to modify children’s numeric representations by aligning low numerical contexts (0-100) and high numerical contexts (0-1,000) in the hopes that the underlying structural similarity (e.g., decimal system) will be highlighted.

The third claim that the numerical estimation tasks will test pertains to children focusing on relevant and irrelevant features/underlying structures in the training and transfer contexts. In Chapter 2, I progressively aligned low and high numeric contexts in the hopes that participants draw the analogy of the decimal system across these contexts and subsequently “linearize” their estimates. The experimental conditions in this series of experiments (e.g., no
alignment, full alignment, focused alignment) will be presented with numeric stimuli that contain both relevant and irrelevant similarities that the participants may focus upon, and I predict that the less irrelevant similarities that a participant’s attention can be drawn to, the more likely the person will be to make accurate estimates in increasingly larger orders of magnitude.

The fourth claim that the representational alignment hypothesis of transfer posits is that because there are a plethora of 1) identical surface features (e.g., number lines oriented left-right, Arabic numerals, paper-and-pencil exercises), 2) features of the experimental context (e.g., same researcher across training and transfer tasks, both tasks conducted in the same room with little temporal delay separating the two tasks), and 3) underlying structural features (e.g., decimal system), there should be successful transfer of participants’ automatic numeric representations across the experiments that I conducted.

And finally, the fifth claim that the representational alignment hypothesis suggests is that children will have a difficult time inhibiting a practiced numerical representation that does not optimize for accuracy. In Chapter 4, participants will be given a category judgment pretest without receiving feedback from the researcher, will be trained to produce a linear series of estimates in a related number line estimation task, and then will be asked to make category judgments in the same numerical categorization task that was administered at pretest. I hypothesize that these participants will continue to represent the category judgment task at posttest the way it was represented at pretest, and thus accuracy will suffer. In the experimental condition that does not receive a category judgment pretest, this inappropriate representation will not be
entrenched, and thus the more appropriate linear representation will be transferred from the training context to the category judgment posttest.
CHAPTER 2

BROAD GENERALIZATION OF LINEAR REPRESENTATIONS

Experiment 1a: Numerical Representations Underlying Estimates in Large Magnitude Contexts

Across the life span, does the logarithmic or the linear function provide a better fit for participants’ number line estimates in large numeric contexts (e.g., 0-1,000, 0-10,000, and 0-100,000)? Previous research has indicated that children abandon an inappropriate logarithmic representation in exchange for adopting a more appropriate linear representation, often after receiving only one trial of maximally discrepant feedback that highlights the difference between the linear and logarithmic functions. Opfer and Siegler (2007) hypothesized that one candidate mechanism that sets in motion this representational change is analogy. Analogy is a potentially powerful mechanism of representational change (Gentner et al., 1997).

In a recent empirical example of representational change, children’s knowledge of the “living thing” concept was investigated (Opfer & Siegler, 2004). Children who initially failed to categorize plants as living things were explicitly told that plants move towards goals to acquire food or sunlight. These children likely drew the analogy: animals move towards goals and they are alive, so if plants move towards similar goals, they are likely alive as well.
Subsequently, these children began categorizing a diverse array of plants as living things.

Even in the history of scientific discoveries, there is evidence that analogy can spur conceptual change. Take for example Gentner et al.’s (1997) case study of Johannes Kepler’s use of analogy (e.g., just as light appears to spread out/become weaker as the distance that the light must travel increases, so too does the influence of the sun on the planets causing planets further from the sun to move slower than those planets that follow pathways closer to the sun). Kepler’s analogies helped him to conceptualize the elliptical pathways of the planets around the sun, which was quite a departure from the widely held belief at the time that the planets actually revolved around the earth, the supposed center of the universe. One of the questions I would like to investigate in my present series of experiments is whether analogy can serve as the mechanism of representational change in numerical estimation.

Why might analogy serve as a mechanism of representational change in the domain of numbers? Numbers are concrete as well as abstract entities; analogies can be drawn over concrete surface similarities or abstract underlying relations. For instance, analogies can be drawn between contexts that share overlapping elements (e.g., 1 and 7 look more similar than 1 and 3), or they can be drawn between contexts that share similar relations or underlying structural similarities (e.g., the magnitudes denoted by 1 and 7 are less similar than the magnitudes denoted by 1 and 3). Further, with increases in age and particular experiences, the accuracy of children’s numerical estimates increase (see further discussion in section 1.5).
Previous research on representational changes that occur in numerical estimation (e.g., Siegler & Opfer, 2003) have indicated that the same children know and use multiple numeric representations depending on the particular context at hand. For instance, second grade students who produced an adult-like, linear series of estimates in a 0-100 context often produced a logarithmic series of estimates in a 0-1,000 context (e.g., 150 was marked much closer to 1,000 than to 0). Interestingly, by providing second grade students with corrective feedback about the placement of their estimates (see Opfer & Siegler, 2007), these students abruptly adopted a linear representation of numbers across the entire 0-1,000 range, often after receiving feedback on only one number that maximized the discrepancy between a linear and a logarithmic function (e.g., 150). These studies of representational change in numerical representation are similar to the representational change seen in Opfer & Siegler’s (2004) living things experiment. In both instances, participants were given maximally discrepant information that allowed them to abandon their current representation in lieu of adopting a new, more adult-like representation.

Why might second grade students make broad generalizations within a particular order of magnitude (e.g., 0-1,000), whereas they have difficulty transferring their linear representation in the 0-100 context to the 0-1,000 context? What type of experimental intervention could be undertaken to promote transfer of the linear representation across increasingly larger orders of numeric magnitude? I would like to focus on the potential mechanism of analogy.
2.1 Progressive Alignment

To test whether analogy could serve as the mechanism of developmental change across numeric representations, I examined whether progressive alignment—a means of fostering analogies in young children (Kotovsky & Gentner, 1996)—led numeric representations used at low numerical ranges (0-100) to be spontaneously generalized to progressively higher numerical ranges (0-1,000, 0-10,000, and 0-100,000) simply by highlighting the perceptual similarities across these numeric contexts.

What is structural/progressive alignment, and how does it act as a mechanism of representational change? According to Kotovsky and Gentner (1996), progressive alignment allows children to make similarity comparisons over concrete, perceptual similarities (e.g., when shown stimuli arrays like oOo and xXx, children might describe them as having two little ones on either side of a big one). After children practice making similarity comparisons, children’s ability to notice higher-order relational commonalities is facilitated (e.g., arrays like oOo and xXx can be described with the relational terms: "baby," "daddy," "baby").

Further, progressive alignment assumes that the process of making similarity comparisons is actually one where representational structures are aligned and mapped onto one another. In this way, carrying out a concrete similarity comparison highlights the relational structure common between the stimuli and makes this relational structure more salient. When the relational structure becomes more salient, the structure can be abstracted and transferred to other related problems. Thus, progressive alignment allows children to recognize the abstract relational similarity uniting their mental representations. Therefore,
progressive alignment can actually decontextualize relations from initially situated representations such that these representations can be matched across diverse domains (Kotovsky & Genter, 1996). For instance, when Gick and Holyoak (1983) asked adults to list the commonalities between their famous military problem and a second problem where a firefighter attempted to put out a fire with a series of buckets or fire hoses, the participants were more likely to abstract a general schema that facilitated performance on a transfer task (e.g., Duncker’s [1945] radiation problem). Kotovsky and Gentner claim that the process of recognizing commonalities across the concrete, surface level allows participants to find "richer and deeper" commonalities at the structural level.

In the next sections, I will detail 1) analogy as a mechanism of developmental change across numeric representations, and 2) the empirical questions that were examined in the present studies.

2.2 Analogy as a Mechanism of Developmental Change

A person’s use of analogies can prompt representational changes. Analogies can be elicited by drawing a person’s attention to surface similarities between two contexts or by highlighting the similar underlying structural relations between the two contexts at hand. This general perspective on representational change, drawn from computational models of cognition (Doumas, Hummel, & Sandhofer, 2008; Gentner, 1983; Hummel & Holyoak, 2003), artificial grammar learning in infants (Marcus, Vijayan, Bandi Rao, & Vishton, 1999), and historical changes in scientific concepts (Gentner et al., 1997; Holyoak & Thagard, 1995), immediately suggested analogy as a candidate mechanism for development of numerical representations.
In Opfer and Siegler (2007), analogy was suggested as the candidate mechanism of representational change when children adopted a linear representation of number after receiving maximally discrepant feedback. Opfer & Siegler believed that children were drawing an analogy between the familiar 0-100 context in which they already possessed a linear representation and the less familiar 0-1,000 context in which they possessed a logarithmic representation. The analogy that Opfer and Sigler assumed children were drawing was: 15 in the context of a 0-100 number line should be placed in the exact same location as 150 in the context of a 0-1,000 number line. It is important to note that children’s use of analogy was not directly tested in Opfer and Sigler’s experiment, so it could have been the case that children were only learning the general ratio characteristics of the order of magnitude in which they received explicit feedback (e.g., 0-1,000 only).

One of the goals that motivated the present series of experiments was to determine if the representational change seen in Opfer & Siegler (2007) was more widespread than just the order of magnitude in which participants received feedback and thus impacted the linearization of children’s estimates across even larger orders of magnitudes. I hypothesized that progressive alignment of units in the 0-100 numeric range and larger orders of magnitude may elicit the analogy highlighting the numerical decimal system. This alignment might make the similarities across the increasingly larger orders of magnitude more apparent for children as they attempted to make their estimates.

For instance, after indicating that differences across units can be ignored when estimating magnitudes on a number line (e.g., regardless of whether one is estimating 15 cherries, 15 cocker spaniels, or 15 balloons, the hatch mark for 15
should be placed in the same location on a 0-100 cherry line, 0-100 cocker spaniel line, or 0-100 balloon line), children may be more likely to recognize that differences across orders of magnitudes may also be ignored (e.g., because of the way the numeric decimal system works, one can ignore the one extra zero in the 0-1,000 context, two extra zeros in the 0-10,000 context, or the three extra zeros in the 0-100,000 number line context and effectively think about these contexts as analogous to the 0-100 context). By getting children to think about large numeric contexts in the same way they think about small numeric contexts, children may be prompted to “linearize” their number line estimates.

To put it another way, children need to learn the applicability of the linear representation for increasingly larger orders of magnitude. I hypothesized that linearization of increasingly larger orders of numerical magnitude is not an all-or-none, stage-like process. Rather, the linearization of numerical magnitudes is likely to occur more gradually, with children producing more linear estimates in one order of magnitude at a time. I hypothesized that children make linear estimates in contexts in which they are more familiar and logarithmic estimates in less familiar contexts. This hypothesis is consistent with Siegler’s (1996; 2006) Overlapping Waves Model, which hypothesizes that the strategies that children know and use are constantly in flux, with some strategies gaining prominence, whereas others lose prominence as a child becomes more aware of the general applicability and success of new and old strategies for the problems at hand.

Why are adults and older children so successful in their attempts to produce estimates in large numeric ranges? I hypothesized that adults and older children extend the analogy children were believed to have used in Opfer and Siegler (2007): 9 in the context of a 0-10 number line, 90 in the context of a 0-100
number line, 900 in the context of a 0-1,000 number line, 9,000 in the context of a 0-10,000 number line, and 90,000 in the context of a 0-100,000 number line are numbers that all play the same “role” across these diverse contexts. That is, when making estimates on number lines in these analogous contexts, a person should make a mark in the exact same location on each number line for the numbers 9, 90, 900, 9,000, and 90,000 for the most accurate performance.

I hypothesized that the analogy a participant uses has differing “strengths” depending how far the participant is trying to “scale up” his or her linear representation. A person is hypothesized to be most successful at scaling up a linear representation from a context in which he already possesses a linear representation (e.g., 0-100) to the next larger order of magnitude (e.g., 0-1,000). The ability to scale up the linear representation beyond the “neighboring” order of magnitude deteriorates the further away the transfer context is in orders of magnitude (e.g., 0-10,000 or 0-100,000). This is likely because the child becomes less and less familiar with the numbers as they increase in orders of magnitude. This hypothesis suggests age differences in the estimates made across increasingly larger orders of magnitude and the underlying numerical representations that are pressed in the service of making these estimates.

2.3 Issues Examined in Present Studies

The central purpose of the present studies is to 1) trace participants’ use of the linear representation in large numeric contexts across the lifespan and 2) determine if using an analogy that aligns low and high numeric ranges will prompt children to “linearize” their numeric estimates without receiving explicit instruction to do so.
The purpose of Experiment 1a is to determine age differences in the underlying representations that children and adults used when making estimates on number lines of increasing numerical magnitude and track the developmental trajectory of participants’ linearization of estimates on number lines of increasing magnitudes. To investigate these age differences in numerical representations, I examined second and sixth grade students’ number line estimation performance in large numerical contexts (e.g., 0-1,000, 0-10,000, and 0-100,000) prior to the students receiving any corrective feedback from the experimenter. These children’s results were then compared to the performance of college-aged adults who were asked to perform the very same task.

From the results gained from Experiment 1a, I hoped to identify the age group where an intervention would be most effective to prompt children to spontaneously transfer their linear representations to larger numeric contexts. In Experiment 1b, I attempted to 1) determine how widely children generalized linear representations of numerical magnitudes after they were given feedback on the correctness of their estimates and 2) determine whether the mechanism of analogy was sufficient to produce generalization to larger numerical contexts without children receiving explicit corrective feedback in these contexts. To meet my goals for Experiment 1b, I brought number lines for high scales (0-1,000, 0-10,000, and 0-100,000) into progressive alignment with scales that children already represented linearly (0-100).

The alignment procedure I used is considered “progressive” in two senses. First, all children’s comparisons were supported by increasing the perceptual similarity of low (training problems) and high scales (generalization problems) by matching the color of units and orders of magnitude. Second, in the
alignment conditions, children’s comparisons of low to high scales were supported by allowing children to directly compare 0-100 problems with 0-1,000, 0-10,000, and 0-100,000 problems; in contrast, children in the no alignment condition were not given this opportunity. Finally, to test for representational change, I examined numerical estimates on a 0-1,000 posttest, where children were not given feedback, perceptual support, or the opportunity to compare problems to lower scales.

Participants were randomly assigned to one of three experimental conditions in Experiment 1b where the amount of progressive alignment participants received (e.g., none, full, focused) was manipulated to determine how much alignment was necessary to prompt the formation of an analogy between low and high numeric contexts. There was a no alignment group and two alignment groups: full alignment and focused alignment. As will be described in further detail in the Method section of Experiment 1b, participants in the full alignment group had the ability to compare the current problem they were trying to solve with all other previously answered training and generalization problems. Participants in the focused alignment condition were only able to compare across related training and generalization problems and were unable to reference previously presented generalization problems.

I hypothesized that the focused alignment participants would outperform the full alignment participants because participants in the focused alignment group would be able to compare across aligned units and magnitudes, and would be more successful at linearizing their estimates. Participants in the full alignment condition, on the other hand, could compare units on one training problem to units on another training problem, magnitudes on one generalization
problem to magnitudes on another generalization problem, compare the appropriate units to the appropriate magnitudes, or compare units to a non-related generalization problem. If participants’ attention was not directed to the appropriate comparison between units and magnitudes, this could lead to lower accuracy when participants attempted to scale up their linear representation of number. Comparing and aligning units to related magnitudes should highlight the fact that children can ignore the irrelevant zeros in the magnitude problems (drawing the analogy between the familiar numeric context in which they were already producing linear series of estimates and this less familiar numeric context), and therefore kids will be prompted to transfer their linear representation from small numerical contexts to larger numerical contexts.

2.4 Developmental Trajectory of Log-Lin Switch

The current experiment investigated the developmental trajectory of second graders’, sixth graders’, and adults’ numeric estimates in the context of large numeric ranges (e.g., 0-1,000, 0-10,000 and 0-100,000). Previous number line estimation results (Siegler & Opfer, 2003) have indicated that numerical representations appear to broadly generalize to some contexts and narrowly generalize to other contexts. Adults and sixth graders produced a linear series of estimates when presented with estimation problems in a 0-1,000 context, whereas second graders produced a logarithmic series of estimates in this very context and a linear series of estimates in the more familiar 0-100 context, despite the fact that some of the same numerosities were presented in both contexts. Though I hypothesized that second grade students would continue to produce a logarithmic series of estimates in all larger numeric contexts in which they were unfamiliar, and I hypothesized that sixth grade students and adults would
continue to produce a linear series of estimates in larger numeric contexts with which they were familiar, this claim had not been investigated previously.

### 2.5 Method

#### 2.5.1 Participants

Participants were 24 second graders \((\text{mean age} = 7.9, \ SD = .33; 14 \text{ girls and} \ 10 \text{ boys})\), 24 sixth graders \((\text{mean age} = 12.06, \ SD = .36; 12 \text{ girls and} \ 12 \text{ boys})\), and 24 college-aged adults \((\text{mean age} = 19.96, \ SD = 1.8; 16 \text{ women and} \ 8 \text{ men})\). The children were recruited from public elementary schools in upper-middle class suburbs of a large city in the Midwestern United States. The adults were recruited from an introductory psychology course at a large university in the same Midwestern city. The children participated in return for a small prize (e.g., a sticker), whereas the adults received credit towards their introductory psychology course. One female graduate student and one female undergraduate served as experimenters.

#### 2.5.2 Design and Procedure

All children and adults were given a number line estimation task that consisted of a line flanked by two hatch marks where the left hatch mark was labeled “0,” and the right hatch mark was labeled with either “1000,” “10000,” or “100000” (see Figure 2.1). Order of magnitude presented in the number line estimation task was a between-subjects variable. On this number line task, the participants were asked to estimate the position of a third number appearing above the midpoint of the number line by making a hatch mark through the line. The to-be-estimated numbers were chosen to maximize the discriminability of the logarithmic and linear functions and to minimize the influence of specific knowledge (e.g., 500 is halfway between 0 and 1000).
Participants were presented with one number line estimation problem per page, which ensured that the participants were unable to reference their previous
estimates. Participants from each age group were randomly assigned to one of three conditions where they completed 10 number line estimation problems without feedback from the experimenter: 1) 0-1,000 context (numerosities: 20, 50, 80, 110, 150, 250, 490, 610, 730, and 940 in randomized order), 2) the 0-10,000 context (numerosities: 200, 500, 800, 1100, 1500, 2500, 4900, 6100, 7300, and 9400 in randomized order), or 3) the 0-100,000 context (numerosities: 2000, 5000, 8000, 11000, 15000, 25000, 49000, 61000, 73000, and 94000 in randomized order).

2.6 Results

To determine whether the hypothesized logarithmic-to-linear shift seen in other number line estimation tasks (e.g., Siegler & Opfer, 2003; Opfer & Siegler, 2007) was apparent in the present study, I compared the fit of the best fitting linear and logarithmic functions to the median numerical estimates across the three age groups (see Figure 2.2). Across all orders of magnitude, adult participants produced more linear than logarithmic series of estimates (0-1,000, $\text{lin } R^2 = 1 > \log R^2 = .83$; 0-10,000, $\text{lin } R^2 = 1 > \log R^2 = .82$; 0-100,000, $\text{lin } R^2 = 1 > \log R^2 = .83$), as did sixth graders (0-1,000, $\text{lin } R^2 = 1 > \log R^2 = .84$; 0-10,000, $\text{lin } R^2 = .98 > \log R^2 = .87$; 0-100,000, $\text{lin } R^2 = .89 > \log R^2 = .81$). Second grade participants produced more logarithmic than linear series of estimates across the 0-1,000 problems ($\log R^2 = .91 > \text{lin } R^2 = .82$), the 0-10,000 problems ($\log R^2 = .85 > \text{lin } R^2 = .59$), and on the 0-100,000 problems ($\log R^2 = .84 > \text{lin } R^2 = .66$).
Figure 2.2 Lin and log best-fits of median estimates. Scales: 0-1,000, 0-10,000, and 0-100,000; Ages: second graders, sixth graders, and adults.

Next, I examined accuracy of numerical estimates across the orders of magnitude. To measure accuracy, I converted the magnitude estimate for each number (the participant’s hatch mark) to a numeric value (the linear distance from the “0” mark to the participant’s hatch mark), and then I divided the result by the total length of the line (e.g., 1,000, 10,000, or 100,000). The magnitude of each participant’s error \((0 - 1)\) was calculated by taking the mean absolute difference between each of the participants’ estimated values and the actual values. Accuracy was found by taking the inverse of the mean absolute error.
First, I was interested in determining whether there was an age x scale interaction, so I conducted a 3 (age: 2nd, 6th, college) x 3 (scale: 1,000, 10,000, 100,000) ANOVA on the dependent variable of accuracy scores. There was a main effect of age, \( F(2, 72) = 46.72, p < .0001 \), and a main effect of scale, \( F(2, 72) = 3.32, p < .05 \). These main effects were qualified by a significant age x scale interaction, \( F(4, 72) = 2.71, p < .05 \) (see Table 2.1). The adults and sixth graders outperformed the second graders in the 0-1,000 and 0-10,000 scales. In the 0-100,000 scales, adults outperformed both the sixth graders and second graders, who did not differ from one another in terms of accuracy on these scales.

<table>
<thead>
<tr>
<th></th>
<th>0-1,000</th>
<th>0-10,000</th>
<th>0-100,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( F(2, 23) = 19.42, p &lt; .0001 )</td>
<td>( F(2, 23) = 53.39, p &lt; .0001 )</td>
<td>( F(2, 23) = 7.97, p &lt; .01 )</td>
</tr>
<tr>
<td></td>
<td>( M_{adults} 96.2% &gt; M_{sixth} 91% &gt; M_{second} 80.3% )</td>
<td>( M_{adults} 96.7% &gt; M_{sixth} 89.2% &gt; M_{second} 67.9% )</td>
<td>( M_{adults} 96.3% &gt; M_{sixth} 80.4%, M_{second} 73.8% )</td>
</tr>
</tbody>
</table>

Table 2.1 Age x scale interaction.

In an attempt to determine if the pattern of results at the group level held at the individual participant level, I determined the percentage of participants who were best fit by the linear function as compared to the logarithmic function across ages (e.g., 2nd grade, 6th grade, and adults; see Figure 2.3). A 2 (best-fitting function: linear, logarithmic) x 3 (age: 2nd, 6th, college-aged) Fisher’s Exact Probabilities test indicated that the three groups differed reliably when they were asked to make estimates in the 0-1,000, 0-10,000, and 0-100,000 contexts (\( p’s < .01 \)). In the 0-1,000 and 0-10,000 conditions, 25% of second graders, 87.5% of sixth
graders, and 100% of adults were best-fit by the linear function. When participants were asked to make estimates in the 0-100,000 context, 25% of second graders, 50% of sixth graders, and 100% of adults were best-fit by the linear function.

![Graph showing percentage of participants best-fit by linear function](image)

Figure 2.3 Percentage of participants best-fit by linear function.

In summary, there were improvements in participants’ numerical estimates. For instance, the current experiment indicated that adults made more accurate estimates than did second grade and sixth grade students on 0-1,000, 0-10,000, and 0-100,000 estimation problems. Likewise, sixth graders outperformed second graders when they made estimates in the 0-1,000 and 0-10,000 contexts. Converging evidence indicative of the logarithmic-linear shift became apparent when 1) participants’ median estimates were compared against the actual values of the to-be-estimated numbers, and 2) when the percentage of participants whose estimates were best fit by the linear function, as opposed to the logarithmic function, were investigated.
Experiment 1b: Effect of Progressive Alignment on the Breadth of Transfer

Findings of age-differences in Experiment 1a motivated Experiment 1b. Specifically, results from Experiment 1a revealed that adults and older children produced a linear series of estimates in large numerical magnitude contexts, but second graders produced a logarithmic series of estimates across the 0-1,000, 0-10,000, and 0-100,000 contexts. I hypothesized that second graders produced a logarithmic series of estimates in these large numerical contexts because of their lack of familiarity with the numeric contexts and their failure to draw relevant analogies between these orders of magnitude and a context (0-100) in which they already thought about numbers in a linear fashion. Therefore, second graders were prime candidates for a training intervention that highlighted the perceptual similarity between the 0-100 context and the larger numerical magnitude contexts used in Experiment 1a. The intervention made participants aware that their automatized linear representation was appropriate not only for the 0-100 context, but also for the 0-1,000, 0-10,000, and 0-100,000 contexts as well. Thus, when the similarity between the training and transfer contexts was highlighted, according to the representational alignment hypothesis, transfer should be facilitated.

The present study addressed two novel questions about broad and rapid changes in the logarithmic-linear shift with regards to numerical estimation: 1) How widely can children generalize linear representations of numerical magnitudes? and 2) Is analogy a mechanism of representational change that is sufficient to produce generalization to larger numerical contexts? In an attempt
to investigate these questions, I brought number lines for high scales (0-1,000, 0-
10,000, and 0-100,000) into progressive alignment with scales that children
already represent linearly (0-100).

2.7 Method

2.7.1 Participants

Participants were 46 second graders (mean age = 7.78, SD = .41) enrolled in
public elementary schools in upper-middle class suburbs of a large city in the
Midwestern United States. There were 20 girls and 26 boys that participated in
return for a small prize (e.g., a sticker). One female graduate student served as
experimenter.

2.7.2 Design and Procedure

Children estimated the placement of numbers on number lines over three
phases of the experiment: training, generalization, and posttest (see Figure 2.4).
The number line task presented children with a line flanked by two hatch marks,
the left hatch mark was labeled “0,” and the right hatch mark was labeled with
either “100,” “1000,” “10000,” or “100000.” On each trial, children were asked to
estimate the position of a number (one per number line) by making a hatch mark
through the line. The numbers to be estimated (2, 5, 8, 11, 15, 25, 49, 61, 73, 94,
and multiples thereof) were chosen to reduce the influence of specific knowledge
(e.g., that 50 is half of 100) and to over-sample at the low end of the range to
maximize the discriminability of the logarithmic and linear functions.
Figure 2.4 Illustration of experimental conditions: a) no alignment, b) full alignment, and c) focused alignment.

In the training phase (Figure 2.4, left column), participants were given 40 0-100 number line problems, and they received corrective feedback on their estimates. For instance, children were told, “These two lines are rather close,” if their hatch marks deviated less than 10% from the actual placement of the to-be-
estimated number. Likewise, children were told, “These two lines are rather far apart,” if their hatch marks deviated more than 10% from the actual placement of the to-be-estimated number (for a more detailed description of the feedback procedure, see Opfer & Siegler, 2007). All training problems specified units (i.e., pears, cherries, and carrots), and after completing these problems, children were shown that their estimates did not differ much over different units. In the generalization phase (Figure 2.4, middle column), I highlighted similarity of generalization and training problems (that the color of units matched the color of zeros denoting order of magnitude) by asking participants “to try some more problems just like the ones you just finished.” Then, participants were asked to make estimates for 40 new number lines (0-100, 0-1,000, 0-10,000) without corrective feedback. In the posttest phase (Figure 2.4, right column), participants were asked to estimate the position of 10 numbers on 0-1,000 number lines without receiving feedback and without perceptual support (e.g., icons and colored zeros).

To test the effect of progressive alignment on transfer from training to generalization and posttest problems, participants were randomly assigned to one of three experimental conditions: no alignment, full alignment, and focused alignment (cf. Figure 2.4a and 2.4b and 2.4c). In the no alignment condition, participants received training and generalization problems one at a time; thus, participants were not able to directly compare new problems to old problems. In the alignment conditions, participants received identical training problems, but generalization problems were presented alongside previously solved training problems, thereby allowing children in the alignment group to compare
generalization problems to training problems (i.e., to compare units to orders of magnitude). Participants in the focused alignment group were only able to compare one training-generalization pair at a time, whereas participants in the full alignment condition were able to see all previous answers since all training and generalization problems were presented for them on one page. The focused alignment condition was used to ensure that participants in the alignment condition did not outperform participants in the no alignment condition simply because these participants were permitted to employ a “dumb” strategy of marking a single hatch mark down the page without paying attention to the context of the numerical magnitudes. In the partial alignment condition, participants only saw the relevant training and generalization problems aligned (e.g., pear training problem and 1 green zero generalization problem), so they were forced to more closely attend to the orders of magnitude in the problems they were attempting to solve; therefore, the analogy between the training and generalization problems was likely the most salient for this experimental group. Thus, in all three conditions, children were told that generalization problems were “just like” the training problems, but only the alignment conditions allowed perceptual comparison of training and generalization problems.

2.8 Results

In the following sections I will analyze children’s number line estimates during the training phase, generalization phase, and at posttest to determine whether children were capable of using an analogy to “scale up” their learned linear representation to larger and larger numerical magnitude contexts.
2.8.1 Training

I first sought to characterize the representations of the analogical base, the 0-100 context. Specifically, I wanted to determine that all three experimental groups possessed a linear representation of numerical magnitudes in this range. To investigate this issue, I compared the fit of the best fitting linear and logarithmic functions to the median numerical estimates across the 3 experimental conditions. Across all of the training problems, participants in the full alignment condition produced more linear than logarithmic series of estimates (Blank $R^2 = .99$; Pear $R^2 = 1$; Cherry $R^2 = 1$; Carrot $R^2 = 1$), as did the focused alignment condition (Blank $R^2 = .87$; Pear $R^2 = 1$; Cherry $R^2 = 1$; Carrot $R^2 = 1$), and the no alignment condition (Blank $R^2 = .99$; Pear $R^2 = .99$; Cherry $R^2 = 1$; Carrot $R^2 = 1$). As a reminder, these results were not surprising since I anticipated that these children would produce a more linear than logarithmic series of estimates in the 0-100 context as was found in previous research (Siegler & Opfer, 2003).

Next, I examined accuracy of numerical estimates across the training phase, which was calculated as in Experiment 1a. As expected, accuracy did not differ significantly across experimental conditions for the four training problem-types (e.g., blank number line, pear line, cherry line, and carrot line). Participants’ accuracy was at ceiling levels. These results replicated previous findings (Siegler & Opfer, 2003) that same-aged children generate both logarithmic (e.g., 0-1,000 context in Experiment 1a) and linear estimates (e.g., 0-100 context in Experiment 1b), depending on the numerical scale with which they are presented.
In an attempt to determine the percentage of children best-fit by the linear or logarithmic function across the training problems, tests of Fisher’s Exact Probabilities were conducted. More children were better-fit by the linear function than the logarithmic function across all of the training problems in the no alignment condition ($M_{\text{Blank}} = 86\%$, $M_{\text{Pears}} = 93\%$, $M_{\text{Cherries}} = 100\%$, $M_{\text{Carrots}} = 93\%$), the full alignment condition ($M_{\text{Blank}} = 63\%$, $M_{\text{Pears}} = 94\%$, $M_{\text{Cherries}} = 94\%$, $M_{\text{Carrots}} = 100\%$), and the focused alignment condition ($M_{\text{Blank}} = 50\%$, $M_{\text{Pears}} = 100\%$, $M_{\text{Cherries}} = 100\%$, $M_{\text{Carrots}} = 94\%$). These statistical analyses indicated that the percentage of participants who were best fit by the linear as compared to the logarithmic function did not differ significantly by experimental condition (e.g., no alignment, full alignment, focused alignment) across these training problems, $p$’s $>.05$, $ns$.

Thus, all three experimental groups appeared to possess a linear representation in the 0-100 range. In the next section, I explored whether these groups successfully generalized their linear representation to larger numerical contexts.

2.8.2 Generalization

Once again, I compared the fit of the best fitting linear and logarithmic functions to the median numerical estimates across the three experimental conditions in the generalization phase to assess the logarithmic-linear shift. Similar to the results found for the 0-100 problems during the training phase, the linear function was the best-fitting function for the experimental conditions on the 0-100 generalization problems (full alignment, $R^2 = .99$; focused alignment, $R^2 = .99$; no alignment, $R^2 = 1.0$).
The experimental groups differed in whether the best-fitting function for their median estimates across the other generalization problems (0-1,000, 0-10,000, and 0-100,000) was the logarithmic or the linear function (see Figure 2.5). For the no alignment group, the best-fitting function was the logarithmic one across the 0-1,000 and 0-10,000 generalization problems ($R^2 = .9$ and .85, respectively). On the 0-100,000 generalization problems, both the linear ($R^2 = .46$) and logarithmic function ($R^2 = .39$) provided a poor fit of the children’s estimates in the no alignment condition. For the full alignment group, the best-fitting function was the linear one across the 0-1,000 and 0-10,000 generalization problems ($R^2 = .96$ and .84, respectively), whereas the logarithmic function was the best-fitting function across the 0-100,000 generalization problems ($R^2 = .86$). Across the 0-1,000, 0-10,000, and 0-100,000 generalization problems, the focused alignment group was best fit by the linear function ($R^2 = 1, .98, and .98$, respectively), indicating that the focused alignment group was successfully applying their learned linear representation of number across larger orders of magnitude.
Figure 2.5 Lin and log best-fits of median estimates at generalization.
Experimental conditions: no alignment, full alignment, and focused alignment.

Recall that one potential explanation for why the focused alignment group outperformed the full alignment group was the fact that participants in the full alignment group were not prevented from employing a “dumb” strategy of simply drawing a single line down the page to mark their estimates for each of the generalization problems, a strategy that would also result in a better fit of the linear as opposed to logarithmic function. Participants in the focused alignment group were unable to employ this strategy, and were likely forced to consider each context (e.g., 0-1,000, 0-10,000, 0-100,000) as they were asked to solve individual generalization problems.
The results illustrated in Figure 2.5 clearly point to the fact that participants’ estimates in the alignment conditions were becoming more “linearized.” That is, the estimates were not necessarily becoming less better-fit by the logarithmic function, but rather, the linear fit was becoming much better with progressive alignment indicating that the children were honing in on the analogy of the interrelations of numbers within the decimal system.

As in the training phase, I analyzed participants’ accuracy on the generalization problems. To determine whether there was an interaction between experimental condition and scale, I conducted a 3 (condition: full alignment, focused alignment, no alignment) x 4 (scale: 0-100, 0-1,000, 0-10,000, 0-100,000) repeated measures ANOVA on accuracy scores on the generalization problems. There was a main effect of scale, $F(3, 41) = 28.73, p < .0001$, and a main effect of condition, $F(2, 43) = 10.44, p < .001$, though these main effects were qualified by a significant condition x scale interaction, $F(6, 84) = 2.99, p < .01$.

I wanted to further investigate the significant condition differences in levels of accuracy across the generalization problems (see Figure 2.6), so I conducted post-hoc ANOVAs to determine how differing amounts of progressive alignment (e.g., full, focused, no alignment) impacted accuracy of children’s estimates.
As anticipated, there were no differences in accuracy across experimental conditions for the 0-100 blank number line problems since participants already possessed a linear representation on these numbers. For the 0-1,000, 0-10,000, and 0-100,000 generalization problems, full and focused progressive alignment allowed children to produce more accurate series of number line estimates in comparison to receiving no progressive alignment (see Table 2.2).

<table>
<thead>
<tr>
<th>Condition</th>
<th>0-100</th>
<th>0-1,000</th>
<th>0-10,000</th>
<th>0-100,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F(2,45) = 4.41, p &lt; .05, M_{rw} 95.9%, M_{at} 96.3% &gt; M_{rw} 95.4%$</td>
<td>$F(2,45) = 8.38, p &lt; .001, M_{rw} 90%, M_{at} 88% &gt; M_{rw} 78%$</td>
<td>$F(2,45) = 6.06, p &lt; .01, M_{rw} 83%, M_{at} 84% &gt; M_{rw} 63%$</td>
<td>$F(2,45) = 10.89, p &lt; .001, M_{rw} 85%, M_{at} 82% &gt; M_{at} 63%$</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.2 Condition x scale interaction.
To assess whether these changes in accuracy were caused by the logarithmic-to-linear shift, I next analyzed the percentage of children best fit by the linear function as compared to the logarithmic function across the three experimental conditions during the generalization phase (see Figure 2.7). Fisher’s Exact Probabilities tests indicated that there were no statistically significant differences in the percentage of children best-fit by the linear or logarithmic function across the three conditions in the 0-100 or 0-1,000 contexts at generalization ($p’s > .05$, $ns$). The linear function provided the best fit for the majority of participants across experimental conditions in the 0-100 context and 0-1,000 contexts (no alignment: $M_{100} = 93\%, M_{1,000} = 57\%$; full alignment: $M_{100} = 100\%, M_{1,000} = 69\%$; focused alignment: $M_{100} = 100\%, M_{1,000} = 81\%$). There were statistically significant differences in the 0-10,000 and 0-100,000 contexts, $p’s < .01$, (no alignment: $M_{10,000} = 50\%, M_{100,000} = 50\%$; full alignment: $M_{10,000} = 56\%, M_{100,000} = 50\%$; focused alignment: $M_{10,000} = 100\%, M_{100,000} = 100\%$).

The focused alignment group had a higher percentage of participants who were better fit by the linear than logarithmic function across all orders of magnitude; the full alignment group had a higher percentage of participants who were better fit by the linear than logarithmic function across the 0-100, 0-1,000 and 0-10,000 contexts; the no alignment group had a higher percentage of participants who were better fit by the linear than logarithmic function in the 0-100 and 0-1,000 contexts only.
Figure 2.7 Percentage of participants best-fit by linear at generalization.

Conditions: no alignment, full alignment, and focused alignment.

In summary, these findings, taken together, indicate that across the 0-1,000, 0-10,000, and 0-100,000 generalization problems the full alignment and focused alignment conditions significantly outperformed the no alignment condition indicating that progressive alignment provided a powerful mechanism of representational change that allowed children to “scale up” their learned linear representation without much explicit instruction to do so. Simply providing children feedback in the 0-100 range and highlighting the perceptual similarity between the training and generalization problems did not allow participants in the no alignment condition to “scale up” their learned linear representation to larger and larger numerical magnitude contexts.

Further, the evidence for representational changes—as evidenced by the fit of the linear regression function—was broader in the focused alignment condition than in the full alignment condition. This finding is important because it suggests that the critical comparison for representational change to take place was that provided in the focused alignment condition, namely the comparison of units (e.g., 15 cherries) to orders of magnitude (e.g., 1500).
2.8.3 Posttest

In the final experimental phase, I wanted to determine whether participants in the full alignment and focused alignment groups were really capable of transferring their learned linear representation of number to an “unsupported” posttest context. This posttest was a blank 0-1,000 number line that did not capitalize on the use of progressive alignment. If children really did have a grasp on the analogy that I tried to implicitly teach them, then they should accurately place their estimates on these unsupported number lines.

Figure 2.8 indicates that participants in the no alignment group ($R^2 = .91$) and the full alignment group ($R^2 = .83$) were better fit by the logarithmic than the linear function at posttest, whereas participants in the focused alignment group were best fit by the linear function ($R^2 = .95$) at posttest. Not only were the focused alignment group best fit by the linear function across all generalization problems, this trend continued at posttest indicating that this experimental condition provided the most supportive condition in which participants utilized the progressive alignment information to “scale up” their learned linear representation. This finding also supported my claim that the reason the focused alignment condition outperformed the full alignment condition on the generalization problems was because the children in the full alignment condition failed to attend to the particular context provided in the generalization problems (e.g., 0-1,000, 0-10,000, 0-100,000), but rather employed the dumb strategy of marking a straight line down the right side of the page, a strategy that was implemented without cost on the training problems where units (e.g., pears, cherries, carrots) should not necessarily impact the location of participants’ estimates on the number lines. If participants were not cognizant of the true
structural alignment of the 0-100 training problems and the 0-1,000, 0-10,000, and 0-100,000 generalization problems, then I would expect that they would not be aware of the analogy that made the decimal system more transparent. Thus, these participants would perform poorly on an unsupported 0-1,000 number line posttest.

Figure 2.8 Lin and log best-fits of median estimates at posttest. Conditions: no alignment, full alignment, and focused alignment conditions

Further, analyses indicated that accuracy across the unsupported context of the 0-1,000 posttest (e.g., blank number line without fruit or vegetable icons and colored zeros) did not differ significantly ($F[2, 45] = .54, p > .05, ns$) across experimental conditions (full alignment, $M = 76\%$; focused alignment, $M = 79\%$; no alignment, $M = 74\%$).

Finally, I investigated the percentage of children best fit by the linear as opposed to the logarithmic function across the three experimental conditions during the posttest phase (Figure 2.9). Fisher’s Exact Probabilities tests indicated that there was a trend toward a statistically significant difference in the percentage of participants who were best fit by the linear function ($p = .07$) across
the no alignment ($M = 29\%$), full alignment ($M = 25\%$), and focused alignment conditions ($M = 63\%$).

Figure 2.9 Percentage of participants best fit by linear at posttest.

Experiment 1b’s methodology is akin to the technique known as analogical encoding whereby participants are asked to compare examples to facilitate extraction of underlying structure common to disparate examples (Gentner, Loewenstein, & Thompson, 2003; Loewenstein, Thompson, & Gentner, 2003). Analogical encoding was the mechanism of transfer investigated in recent experiments where novice negotiators were asked to learn about negotiation strategies. Novice negotiators who compared two written case studies detailing negotiation strategies were better able to abstract general problem-solving schemas which they later applied to future negotiation situations, like face-to-
face negotiations, than were negotiators who were not explicitly told to directly compare the cases or who received no case-based training whatsoever (Gentner, Loewenstein, & Thompson, 2003; Loewenstein, Thompson, & Gentner, 2003).

I believe analogical encoding is an appropriate means of interpreting my findings in Experiment 1b. Children in the alignment conditions were given the opportunity to compare across multiple training problems. This comparison process facilitated participants’ ability to extract the common structure of the decimal system that was present in these problems and subsequently transfer the analogy of the decimal system to the generalization problems. Participants in the no alignment condition did not have the ability to directly compare the training examples, and these participants were unable to extract the common structure of the training problems. Therefore the linear representation was not transferred to the generalization problems in the no alignment condition.

2.9 General Discussion

The ability to carry out effortless structural alignment is a hallmark of human cognitive processing (Gentner & Markman, 1997). In the present series of experiments, children were asked to align the familiar ratio structure of units (e.g., how 10 apples are more similar to 1 apple than to 100 apples) with the ratio structure of the decimal system (e.g., how 10 is more similar to 1 than to 100) in an effort to change how children represent the magnitudes of larger numbers, such as those between 0-1,000, 0-10,000, and 0-100,000. Results supported the assumption that progressive alignment was necessary to generalize a linear representation of 0-100 to higher numbers. Specifically, although I found that feedback on 0-100 problems caused all children to generate linear estimates on 0-100 “training problems,” only children in the alignment conditions generalized
their linear representation to 0-1,000, 0-10,000, and 0-100,000 “generalization problems,” and only children in the alignment groups continued to generate linear estimates on problems without perceptual support, alignment, and feedback (“posttest”).

It is important to note that I aligned contexts (e.g., 0-100) in which children were familiar with “new” and larger numeric contexts (e.g., 0-1,000 to 0-100,000). I think this is integral to my findings because I surmise that children were bootstrapping the linear representation from a context in which they were already familiar (0-100) to less familiar contexts (0-1,000 through 0-100,000). It is an open question as to whether children would be able to progressively align number lines in a context in which they were just taught a linear representation (e.g., 0-1,000) and then extend the analogy to higher orders of magnitude (e.g., 0-10,000 to 0-100,000). I hypothesize that this type of progressive alignment would not be as fruitful as alignment from a more familiar context in which children have had an abundance of experience.

Why are the results obtained in the present series of experiments important? These results provide the first behavioral evidence that a known cognitive mechanism—analogy—allows a pre-existing numerical representation to be extended to represent the magnitudes of a potentially infinite range of new numbers. Second, results are important developmentally because they indicate both the conditions that broad and abrupt cognitive changes take place over the lifespan (with increasing age in Experiment 1a and with progressive alignment in Experiment 1b), as well as the conditions under which cognitive development is slow and gradual (no progressive alignment condition in Experiment 1b). Currently, these two types of cognitive changes figure strongly in competing
theories of cognitive development, but the theory of representational change depicted here points to a way of reconciling them within an architecture that is capable of relational reasoning (e.g., Doumas, Hummel, & Sandhofer, 2008; Gentner, 1983; Hummel & Holyoak, 2003). My results indicate that even sixth graders are still “working out” the interrelations among the orders of magnitude in the decimal system, though their estimation performance was superior to the second graders in the experiments reported here.

Finally, in Experiment 1b, I showed that progressive alignment leads young children to adopt a linear representation of number and abandon a logarithmic one, which lends support to the important educational implications of my results. It seems that the methodology used in this experiment could be implemented as a simple classroom intervention that may allow children to being thinking about numbers in a more adult-like, linear fashion months and maybe even years before classical classroom instruction or life experiences would lead them to do so. To my knowledge, highlighting the properties of the decimal system by progressively aligning number lines is not a technique commonly used by elementary school teachers, but based on previous evidence of causal and correlational links between numerical estimation and other mathematical skills (Booth & Siegler, 2006; in press), I believe this intervention is likely to be easily and effectively implemented.
CHAPTER 3
NON-STRATEGIC GENERALIZATION

Experiment 2a: Costs and Benefits of Representational Change

Taken together, results from Experiment 1a and 1b indicated that there was a logarithmic-to-linear shift in numerical representation that occurred across the lifespan such that prior to training, second graders produced a logarithmic series of estimates in 0-1,000, 0-10,000, and 0-100,000 contexts, whereas adults produced a linear series of estimates in all of these abovementioned contexts. By simply highlighting the perceptual similarity of number lines in the 0-100 context, a context in which second graders were known to think about numbers in an adult-like linear fashion, to larger numerical magnitude contexts, the results indicated that progressive alignment provided second graders with an analogy that allowed them to “scale up” or automatically transfer their learned linear representation to other increasingly larger magnitudes.

The previous set of experiments offered a situation in which transfer of an automatized linear representation from one numerical context to another is advantageous since the linear representation is an appropriate representation for making estimates in the 0-100 as well as 0-1,000, 0-10,000, and 0-100,000 contexts. In the current experiments, I examined one interesting implication of the representational alignment hypothesis: because transfer is an automatic effect of abstract representation, representational changes can impose costs as well as
benefits, leading to unavoidable setbacks in the course of learning that can persist for many years and across many contexts.

Evidence for the benefits of representational change is widespread. For example, when children are given corrective feedback on where to place a few numbers on a line flanked by 0 and 1,000 and no numbers in between (number line estimation), accuracy improves greatly for numbers in the initial training set; learning transfers to numbers outside the training set; and there is robust transfer of learning to related numerical tasks (such as categorizing numbers as “small” or “large;” Laski & Siegler, 2007), with magnitude estimates on the transfer task being almost identical to estimates on the training task (Opfer & Siegler, 2007). Moreover, real world tasks that involve similar kinds of experiences (such as playing board games) also result in transfer to educationally important outcomes, such as preschoolers’ ability to compare numeric value and perform arithmetic (Siegler & Ramani, in press; Ramani & Siegler, 2008; Griffin, Case, & Siegler, 1994).

Evidence for the costs of transfer, however, is rare and indirect. One type of evidence for the costs of transfer comes from research on cognitive illusions in adults (Kahneman & Tversky, 1996), who make grossly mistaken comparisons of risk when framed in a manner that invites inappropriate transfer of numerical representations. As a real life example, genetic counselors often attempt to simplify the risks reported in epidemiological studies by reporting rates of disease in terms of simple frequencies (e.g., 1 in 333) rather than in scientific format, which reports rates of disease per unit of population exposed to the risk (e.g., 3 per 1,000 persons) (Burkell, 2004; Grimes & Snively, 1999; Walker, 1997). Although this simplification is well meaning, research on patients’
understanding of medical risks has shown that the simplification has the
unfortunate consequence of leading patients to make inaccurate comparisons,
such as judging a disease with a rate of 1 in 384 persons as being higher than a
disease with a rate of 1 in 112 persons (Grimes & Snively, 1999). Unlike the
beneficial effects of transferring from the spacing of numbers in board games to
spacing of numbers on number lines, transferring from the number line to
assessments of risk is costly in this case because, unlike the linear increase in
spacing of numbers on number lines, the average risk of disease per unit of
population (e.g., 1, .01, or .001 cases per person) increases as a power function of
the population base in the simplified frequencies (e.g., 1 in 1, 1 in 100, or 1 in
1,000). Further, it is exceedingly difficult to inhibit the automatic activation of the
magnitudes associated with the whole number integers (Besner & Coltheart,
1979) which leads participants to the obvious conclusion that 384 is larger than
112, and therefore a more common disease with which one is likely to be
infected.

To test my idea about the costs of representational change in an
experimental setting, I examined short- and long-term changes in children’s
estimates of numerical and fractional magnitude and the relation between
accuracy across the two tasks. Like the relation between the magnitude of a risk
and the population base, estimates of fractional magnitudes are interesting
because the value of a fraction also increases as a power function of its
denominator. Moreover, previous research has shown that this property of
fractions poses tremendous difficulties for adults who are asked to estimate the
value of salaries (Opfer & DeVries, in press). This led to my hypothesis that
inaccurate estimates of fractional value might stem from learners automatically
transferring their representations of numerical value to the fractional context regardless of the superficial differences across the two contexts and regardless of the gross inaccuracy of such transfer. To examine this issue directly, I gave children feedback as they placed numbers on a number line and then examined whether their subsequent estimates of fractional value were unchanged, better, or worse.

I present evidence that the nature of children’s initial numerical representations are better suited for comparing fractional values than values of whole numbers. In the section, “Issues Examined in Present Studies,” I briefly review evidence on how children’s representations of whole numbers vary as a function of experience and age, and I present a theoretical analysis that links how these same variables should influence accuracy of fractional magnitude estimates. In the same section, I describe how I tested these predictions in Experiments 2a and 2b.

Although automatic activations of linear magnitude representations are beneficial for estimating the value of whole numbers, such representations can be inappropriate for estimating the value of fractions. Specifically, by automatizing that 150 is closer to 1 than to 1,000, adults are subject to a powerful cognitive illusion in which 1/150 seems closer to 1/1 than to 1/1,000. In contrast, children’s belief that 150 is closer to 1,000 than to 1 appears to protect them from this cognitive illusion (Opfer & DeVries, in press). The costs of automatic magnitude representations can also be seen in how adults approach the problem of estimating the sum of fractions, where the magnitude of the denominator and numerator—but not the magnitude of the fraction itself—is represented automatically. Thus, when estimating the answer to 12/13 + 7/8 on a National
Assessment of Educational Progress, for example, fewer than a third of 13- and 17-year-olds correctly chose “2” from the options “1,” “2,” “19,” and “21” (Carpenter et al., 1981). Had students represented the magnitude of 12/13 and 7/8 as each being about equal to 1, the answer would have been easy to solve (~1 + ~1 = ~2). Instead, about half of students answered 19 or 21, indicating that they focused exclusively on numerators (12 + 7 = 19) or denominators (13 + 8 = 21).

This interpretation of adults’ approach to fractions led me to make two predictions about how older children would estimate the value of quantities expressed in fractional notation, such as when estimating the placement of a salary (e.g., $1/60 minutes) on a line that begins with one salary (e.g., $1/minute) and ends with another (e.g., $1/1,440 minutes). The first prediction is that if older children also compare only the value of the denominators in the salary, their linear representation of number will lead to radically inaccurate estimates. This inaccuracy is predicted by the fact that the relation between the numeral expressed in the denominator and the magnitude denoted by the whole fraction is provided by a power function rather than by a linear function. For example, although 60 is closer to 1 than to 1,440, k/60 is closer to k/1,440 than it is to k/1.

The second prediction is that if younger children also compare denominators, their logarithmic representation of numbers should have a correcting effect and thereby lead to more accurate estimates than those of older children. This prediction stems from the fact that the power relation between the value of the denominator and the magnitude of the fraction is somewhat similar to that of a logarithmic function. Thus, for example, the natural logarithm of 60
(4.09) is closer to the natural logarithm of 1,440 (7.27) than to the natural logarithm of 1 (0), much as \( k/60 \) is closer to \( k/1,440 \) than it is \( k/1 \).

### 3.1 Relation between Estimation Tasks

To test the predicted costs and benefits of representational change, I examined the accuracy of symbolic magnitude estimation across two contexts—whole numbers and fractions—and how these two contexts affected the relation between age and accuracy. Will possessing a linear representation of numerical magnitude interfere with estimates of fractional magnitude? To answer this question, I obtained both correlational and causal data.

Correlational evidence was obtained in Experiment 2a by examining whether accuracy of numerical magnitude estimates were inversely related to accuracy of fractional magnitude estimates. In this experiment, I investigated (1) whether a positive relation between age and accuracy in numerical magnitude estimation might co-exist with a negative relation between age and accuracy in fractional magnitude estimation, and (2) whether within each age group, individual differences in accuracy of numerical magnitude estimates might be negatively correlated with accuracy of fractional magnitude estimates.

Causal evidence was obtained in Experiment 2b when I examined transfer of learning from the numerical magnitude estimation context to the fractional magnitude estimation context. Transfer of learning from one context to another is notoriously difficult to elicit (see Barnett & Ceci, 2002, for a recent review), but previous evidence of broad and robust transfer of numerical representations (Laski & Siegler, 2007) lends support to the idea that transfer of numerical representations is automatic. To examine this issue directly, I examined children’s transfer of numerical representations to the fractional
magnitude context, where linear representations would generate inaccurate task performance. In Experiment 2b, I attempted to determine if children who learned to adopt the linear representation after receiving training on the number line task would transfer the more mature representation to their estimation performance on a fractional units task.

Experiment 2a had three major purposes. One purpose was to replicate the finding that children initially generate estimates that increase logarithmically with numerical value. This goal was important because these children will subsequently participate in a microgenetic study of the transition from use of a logarithmic representation to use of a linear representation in numerical estimation (Experiment 2b). The second goal was to test whether the greatest improvement between first and third grade occurs for numbers around 150. This goal was important both because of the theoretical prediction that the greatest improvement with age should come in this area, where the logarithmic and linear functions are most discrepant, and because I later provided feedback for estimates in this region. The third purpose of Experiment 2a was to examine whether increasing accuracy in numerical estimation was accompanied by decreasing accuracy in fractional estimation. This test was the most important because evidence to the contrary would disconfirm my hypotheses about the relation between numerical and fractional magnitude estimation.

3.2 Method

3.2.1 Participants

Participants were 64 first through third graders (mean age = 8.41 years, SD = 0.75; 37 girls and 27 boys) who attended neighborhood schools in largely
European-American, middle class suburbs surrounding a Midwestern city. One of two female research assistants served as experimenter.

3.2.2 Number Line Estimation Task

All number line problems (see Figure 3.1a) consisted of a 20 cm line where the left endpoint was labeled “0,” the right endpoint was labeled “1000,” and the number to be estimated appeared 2 cm above the midpoint of the number line. The experimenter instructed,

Today we’re going to play a game with number lines. We use number lines to help us with math. It looks just like a line with numbers at each end. It shows us where all the numbers in between go. Different numbers go in different places on a number line. In this game there will be a number in a circle up here. Your job is to show me where that number goes on a number line like this one. Each number line will have a 0 at one end and 1,000 at the other end. When you decide where the number goes, I want you to make a mark through the line like this.

Before the first item, the experimenter said, “This number line goes from 0 at this end to 1,000 at this end. If this is 0 and this is 1,000, where would you put N?" Participants were then asked to place numbers (2, 5, 18, 27, 34, 42, 56, 78, 100, 111, 122, 133, 147, 150, 156, 162, 163, 172, 179, 187, 246, 306, 366, 426, 486, 546, 606, 666, 722, 725, 738, 754, 818, 878, and 938) on a number line by making a hatch mark. Each number line problem appeared on its own page. These numbers maximized the discriminability of logarithmic and linear functions by over-sampling the low end of the range, minimized the influence of specific knowledge (e.g., that 500 is halfway between 0 and 1,000), and tested predictions about the range of numbers where estimates would most differ with age.

3.2.3 Fraction Line Estimation Task

Participants were asked to estimate the total amount of money a person might make at a given salary (e.g., $1/60 minutes) by placing a mark on a 20 cm
“money line” (see Figure 3.1b). Both the value of the salaries to be estimated (e.g., $1/60 minutes) and the endpoints of the money line ($1/1 minute, $1/1,440 minutes) were expressed in fractional units. For this task, the experimenter instructed,

Today we’re going to play a game with money lines. We use money lines to tell us how much money a person makes. It looks just like a line with amounts of money at each end. It shows us where all the amounts of money in between go. Different amounts of money go in different places on a money line. In this game there will be an amount of money that a person might make up here (pointing to the top of the blank money line data sheet where the amount of money should be). Your job is to show me where that amount of money goes on a money line like this one. Each money line will have $1 every minute at one end and $1 every 1,440 minutes at the other end. When you decide where the amount of money goes, I want you to make a mark through the money line like this.

Children were asked to estimate the location of the following “salaries” on the “money line:” $1/2 minutes, $1/8 minutes, $1/9 minutes, $1/60 minutes, $1/120 minutes, $1/240 minutes, $1/360 minutes, $1/480 minutes, $1/540 minutes, $1/720 minutes. Again, every problem appeared on its own page. All fractions were read aloud to the children (e.g., “Where would you put one dollar every two minutes?”).
Figure 3.1 Experimental stimuli: a) sample number line estimation problem, b) sample fraction line estimation problem.

3.2.4 Design and Procedure

To ensure that participation in the fractional units task did not affect subsequent performance on number-line problems or the fractional units post-test in Experiment 2b, children were randomly assigned to two groups. One group received the fractional units task first, whereas the other group did not receive the task at all. All participants then received 22 number line estimation problems. Participants were tested in a single session. The items within each scale were randomly ordered, separately for each child, and were presented in small workbooks, one problem per page.
3.3 Results

3.3.1 Age Group Differences in Numerical Estimation

I first examined age differences in the accuracy of numerical estimates. To measure accuracy, I converted the magnitude estimate for each number (the child’s hatch mark) to a numeric value (the linear distance from the “0” mark to the child’s hatch mark), then divided the result by the total length of the line, and multiplied the result by 1,000. The magnitude of each child’s error was calculated by taking the mean absolute difference between each of the child’s estimated values and the actual values. As expected, the mean absolute error declined with increasing age, $r(63) = -.56, p < .001$, decreasing from 25% for the younger half of the sample (7- to 8.49-year olds, $n = 30$) to 11% for the older half of the sample (8.5- to 9.5-year-olds, $n = 34$), $F(1, 63) = 41.75, p < .0001, d = 1.64$.

To determine whether this improvement in accuracy was associated with the hypothesized logarithmic-to-linear shift, I next compared the fit of the best fitting linear and logarithmic functions to the median numerical estimates of the younger (7- to 8.49-year-olds, $n = 30$; 21 girls, 9 boys) and older children (8.5- to 9.5-year-olds, $n = 34$; 16 girls, 18 boys). As in previous studies of this age range (Opfer & Siegler, 2007; Siegler & Opfer, 2003), the fit of the linear function to children’s estimates increased, whereas the fit of the logarithmic function decreased. The median estimates of the younger group were better fit by the logarithmic function ($R^2 = .95$) than by the linear function ($R^2 = .63$), whereas the median estimates of the older children were better fit by the linear ($R^2 = .98$) than by the logarithmic ($R^2 = .74$) function (Figure 3.2).
Figure 3.2 Long-term changes in numerical magnitude estimation:
Estimates of 7- to 8.49-year-olds were better fit by a logarithmic function than by a linear function, whereas the estimates of 8.5- to 9.5-year-olds were better fit by a linear function than by a logarithmic function.

I next tested whether the logarithmic/linear characterization of median estimates reflected individual children’s performance by comparing individuals’ estimates on each task against the predictions of the best fitting linear and logarithmic functions. I assigned a 1 to participants when the linear model provided the best fitting function to his or her estimates and a 0 to participants when the logarithmic model provided the best fitting function. (Because the degrees of freedom were identical for these two models, the simple comparison of R² values was appropriate.) The linear function provided the better fit for 17% of the younger group and 79% of the older groups (see Figure 3.3), whereas the
logarithmic function provided the better fit for 83% of younger children and 21% of older children. To test the association of age with generation of linear estimates more precisely, I used a logistic regression model to test for the effect of age on the odds of generating linear estimates, where age was entered as a continuous variable (range: 7.15 – 9.64 years). The test indicated that there was a significant positive effect of age, with children in this sample being 5.46 times more likely to generate linear estimates with each year of age, \( \hat{\beta} = 1.87, z = 3.99, \) Wald (1, \( N = 64 \)) = 22.40, \( p < .0001 \). Thus, data from multiple levels of analysis—individual children, and younger as compared to older children—indicated a logarithmic-to-linear shift in the 0-1,000 numerical context.

Figure 3.3 Linear best-fit: Proportion of children whose estimates were best fit by a linear function increased substantially between ages 7.5 and 8.5.
3.3.2 Breadth of Age Group Differences

I next examined whether the greatest improvements in estimates occurred on numbers around 150, where the discrepancy between the logarithmic and linear functions was greatest (see Figure 1.1) and where children would later receive feedback in Study 2b. This analysis was important because it is possible, for example, for children to have a linear representation with a very high or low slope, thereby affecting where the maximum discrepancy would actually occur. To examine improvement with age over the numbers tested, I used the absolute age differences between each age group’s median estimate for each number and the correct value for the number.

From these, I correlated the absolute numerical distance of each to-be-estimated number from 150 with the absolute difference in estimates between younger and older children on that number. Improvement in estimation accuracy proved to be highly correlated with distance from 150: \( r(21) = -.71, p < .001; \) the closer the number to 150, the greater the improvement with age. I then examined estimates for a fixed numerical range around three anchors of interest—150 (where the discrepancy in estimates is greatest between the logarithmic and linear functions), 725 (where the discrepancy is 40% of the discrepancy at 150) and 5 (where the discrepancy is also 40% of the discrepancy at 150). The stimulus set included four numbers in each of these three numerical ranges. As anticipated by the log discrepancy hypothesis, a one-way ANOVA indicated differences between these three ranges, \( F(2, 11) = 25.55, p < .001. \) Post-hoc analyses indicated that age-related improvements in estimation accuracy was greater for the four numbers around 150 than for either the four numbers around 725 (\( t[3] = 10.72, p < .01, d = 7.06 \)) or for the four numbers around 5 (\( t[3] = 3.35, p \))
In contrast, younger and older students’ estimates around 5 and 725 did not differ significantly from each other ($t[3] = 2.39$, ns).

3.3.4 Relation between Estimation Tasks

I next examined whether younger children provided more accurate estimates of fractional magnitude than older children, and if so, whether this change was associated with performance on the numerical estimation task. To measure accuracy on the fraction line task, I converted each estimate to a numerical value by measuring the distance between the child’s estimate and the origin of the scale (0 – 20 cm) divided by the length of the scale (20 cm). The magnitude of error for each estimate (0 - 1) was obtained by taking its absolute difference from the correct placement of the fraction on the scale (0 – 1), and accuracy was obtained by subtracting the error from 1. As predicted by the representational change hypothesis (Opfer & Siegler, 2007), age was negatively related to accuracy, $r(63) = -.43$, $p < .001$, with younger children’s accuracy (48%) being significantly higher than older children’s (38%), $F(1, 63) = 9.81$, $p < .01$, $d = .73$.

I next performed a linear regression to examine whether each child’s accuracy on the number line task (0-100%) predicted accuracy on the fraction line task (0-100%). The relation between the two tasks was very strong and negative, $r = -.80$, $F(1, 34) = 57.93$, $p < .0001$, indicating that 63.7% of variation in accuracy of fractional magnitude estimation was accounted for by inaccuracy in numerical magnitude estimation (see Figure 3.4). Because age alone accounted for a marginally significant amount of variation in accuracy of fractional magnitude estimates ($R^2 = .10$, $F[1, 33] = 3.80$, $p = .06$), I next entered the age variable in the regression model, and I found that the combination of age and accuracy on the
number line task accounted for only an additional 1.3% of variance ($R^2 = .65$), which was not a significant addition to the model ($F$-change = 1.19, ns). The full regression model with age, accuracy of numerical magnitude estimation, and interaction between the other two variables accounted for 69.5% of variance in accuracy of fractional magnitude estimation.

Figure 3.4 Relation between accuracy on number line (NL) and fraction line (FL) task: Darkness of circles depicts age group of participants (dark = 7 to 8.49 years old; light = 8.5 to 9.5 years old).

Why might inaccurate numerical magnitude estimation reliably predict accurate fractional magnitude estimation? According to the representational change hypothesis (Opfer & Siegler, 2007), this relation stems from the particular
pattern of errors that children are likely to make when estimating numerical value. That is, children’s errors in numerical estimation are not random, but are generated by their reliance on a logarithmic representation, which generates estimates that are somewhat similar to the power function relating the value of a fraction to its denominator. The similarity of the two functions is apparent in Figure 3.5, which depicts children’s fractional magnitude estimates against the denominator. Ideally, estimates of the value of the fraction (y) should initially decrease dramatically as the denominator increases in magnitude (i.e., y = 1/x). Younger children’s representation of the denominator also leads them to generate estimates in a way that is somewhat similar to this pattern, with their estimates of fractional value decreasing logarithmically with the value of the denominator (log R² = .87). In contrast, older children’s numeric representation leads them to generate a less accurate pattern, with their estimates of fractional value decreasing linearly with the value of the denominator (lin R² = .96).
Figure 3.5 Long-term changes in fractional magnitude estimation: Estimates of 7- to 8.49-year-olds (dark circles) were better fit by a logarithmic function than by a linear function, whereas the estimates of 8.5- to 9.5-year-olds (light circles) were better fit by a linear function than by a logarithmic function. Ideal pattern of performance \( (y = 1/x) \) is depicted in grey.

To test this explanation more directly, I next examined accuracy in fractional magnitude estimation as a function of the logarithmicity and linearity of children’s numerical magnitude estimation. The first relevant evidence came from the fit of the linear and logarithmic R\(^2\) values associated with each child’s numerical magnitude estimates. As hypothesized by the representational change hypothesis, each of these variables accounted for a significant amount of variation in children’s fractional magnitude estimates. As children’s numerical estimates grew more logarithmic, their fractional estimates increased in accuracy, \( r = .45, F(1, 33) = 8.43, p < .0001 \). As children’s numerical estimates grew more linear, their fractional estimates decreased in accuracy, \( r = .76, F(1, 33) = 44.25, p < .0001 \).
Experiment 2b: Short-Term Changes in Numerical and Fractional Magnitude Estimation

Experiment 2b provides a more direct test of the representational change hypothesis (Opfer & Siegler, 2007) regarding accuracy of numerical and fractional magnitude estimation. I predicted that feedback on numbers around 150 would elicit a large change in the proportion of children best fit by the logarithmic function (because this is the area of maximum discrepancy between logarithmic and linear representations) and that the change will involve a broad range of numbers and will occur abruptly rather than gradually (because the change involves a choice of a different representation, rather than a local repair to the original representation). Further, I tested the hypothesis that feedback on accuracy of numerical magnitude estimates leads to increasing accuracy of numerical magnitude estimates but decreasing accuracy of fractional magnitude estimates (again because change is hypothesized to involve substituting linear representations of numbers for logarithmic ones).

The methodology used in Experiment 2b is known as the microgenetic method (Siegler, 1996; 2006). This design allows researchers to observe the progressions and regressions of learning from one context to the other as well as the highlighting of the transitional states that often only last a fleeting amount of time but are integral to the change process. Microgenetic studies are most effective for characterizing how change occurs over time. These studies have three defining characteristics: 1) observations must span the entirety of the change process, 2) a dense sampling of observations are taken as compared to the rapidity of the change process (e.g., high temporal resolution), and 3) the
observations are analyzed with an eye towards the representations and processes that give rise to the change process (Siegler, 1996; 2006).

The microgenetic method characterizes five important dimensions of change. The source of change highlights the things, like providing feedback, or some other type of experimental intervention, that set the change in question into motion. The rate of change focuses on the amount of time that passes between initial usage of a new strategy and consistent use of that strategy over time (e.g., first occurrence of linear function being the best fit for children’s numerical estimates). Next, the path of change can be investigated, or the various knowledge states that children pass through before they gain competence in a particular domain (e.g., did participants gradually switch from use of a logarithmic to a linear representation, or was there a more abrupt switching of these numerical representations). Breadth of change is the analogous dimension to transfer of learning, or how widely the participant generalizes their newly acquired knowledge to novel problems or contexts. Finally, variability is a dimension that sheds light on the individual differences within a learning context (e.g., differences in numerical estimates between males and females or younger and older children).

3.4 Method

3.4.1 Participants

Participants were the same children who participated in Experiment 2a. For convenience, I divided them into “younger” (7- to 8.49-year-olds, n = 30; 21 girls, 9 boys) and “older children” (8.5- to 9.5-year-olds, n = 34; 16 girls, 18 boys), as I had done in Experiment 2a. One of two female research assistants served as experimenter.
3.4.2 Tasks

The number line and fraction line estimation tasks that were described in Experiment 2a were also used in Experiment 2b.

3.4.3 Design and Procedure

Immediately after participation in Experiment 2a, children were randomly assigned to one of two groups: one group received feedback during the training phase (treatment group, \( n = 32 \)), whereas the other group did not receive feedback (control group, \( n = 32 \)).

As shown in the outline of the procedure in Table 3.1, children in both groups completed the number line estimation task for three training trial blocks, and a posttest. The purpose of these phases (training trial blocks, and posttest) was to examine the course of learning prior to posttest and to ensure that learning had occurred prior to the transfer task (i.e., the fractional estimation posttest). On the number line pretest (from Experiment 2a) and posttest, children in the treatment and control groups were presented the same 22 problems without feedback (i.e., without treatment). For children in the treatment group, each training trial block included a feedback phase and a test phase. As shown in Table 3.1, the feedback phase of each training trial block included either one item on which children received feedback (Trial Block 1) or three items on which they received feedback (Trial Blocks 2 and 3). The test phase in all three training trial blocks included 10 items on which children did not receive feedback; this test phase occurred immediately after the feedback phase in each training trial block. Children in the control group received the same number of estimation problems, but they never received treatment. On the posttest, children in both groups were presented the same 22 number line estimation problems without feedback as
were presented in Experiment 2a. The children’s estimates in Experiment 2a provided pretest data, which was used as a point of comparison for their subsequent performance on training and posttest trials and was elicited during the same session.

<table>
<thead>
<tr>
<th>Experimental Groups</th>
<th>Training Phase Number line</th>
<th>Posttest Number line</th>
<th>Fraction line</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment Group</td>
<td>Feed 150 10 #s (0-1000)</td>
<td>Feed 3 #s (147-187) 10 #s (0-1000)</td>
<td>Feed 3 #s (147-187) 10 #s (0-1000)</td>
</tr>
<tr>
<td>Control Group</td>
<td>None 10 #s (0-1000)</td>
<td>None 10 #s (0-1000)</td>
<td>None 10 #s (0-1000)</td>
</tr>
</tbody>
</table>

Table 3.1. Outline of procedures in Experiment 2b.

Feedback was administered to the treatment group following the same procedure used in Opfer and Siegler (2007). The treatment procedure was conducted as follows. On the first feedback problem, children were told, “After you mark where you think the number goes, I’ll show you where it really goes, so you can see how close you were.” After the child answered, the experimenter took the page from the child and superimposed on the number line a 20 cm ruler (hidden from the child) that indicated the location of every 10th number from 0-1,000. Then the experimenter wrote the number corresponding to the child’s mark ($N_{estimate}$) above the mark, and indicated the correct location of the number that was presented ($N$) with a hatch mark. For example, if a child was asked to mark the location for 150 (i.e., $N$) and his estimate corresponded to the actual
location of 600 (i.e., $N_{\text{estimate}}$), the experimenter wrote the number 600 above the child’s mark and marked where 150 would go on the number line. After this, the experimenter showed the corrected number line to the child. Pointing to the child’s mark, she said, “You told me that $N$ would go here. Actually, this is where $N$ goes (pointing). The line that you marked is where $N_{\text{estimate}}$ actually goes.” When children’s answers deviated from the correct answer by more than 10%, the experimenter said, “That’s quite a bit too high/too low. You can see these two lines [the child’s and experimenter’s hatch marks] are really quite far from each other.” Finally, for both high and low deviation estimates, the experimenter asked children to explain the feedback given (a feedback strategy that often has large effects on changes in accuracy; see Chi et al., 1989; Siegler, 2002).

3.5 Results

I organized my results into two sections. Within the first section (Process of Change in Numerical Estimation), I report on the conditions that led to changes in numerical estimation (Source of change), how quickly those changes occurred (Rate of change), and approaches that children used up to and following the use of mature approaches (Path of change). Within the second section (Transfer of Learning to Fractional Magnitudes), I report results testing my hypotheses about transfer of learning to fractional estimation.

3.5.1 Process of Change in Numerical Estimation

3.5.1.1 Source of change. I first examined the source of change in estimation performance on the number line task. Specifically, I wanted to test whether the experiences that children received during the training phase of the experiment improved their estimation accuracy on posttest and influenced the degree to
which those estimates came to follow a linear function. To find out, I first examined posttest accuracy (0 to 1) as a function of age and condition (treatment, control). As expected, a linear regression indicated a main effect of age ($F[1, 60] = 29.4, p < .0001$), of treatment ($F[1, 60] = 5.14, p < .05$), and an age by treatment interaction, $F(3, 60) = 3.85, p = .05$. To examine this interaction more closely, I looked at changes in younger and older children separately.

Among younger children, I first examined the effect of treatment on estimation accuracy by calculating the mean absolute error for each child and then performing a 2 (condition: treatment, control) X 2 (test phase: number line pretest, number line posttest) repeated-measures ANOVA on the error scores. As expected, there was a main effect of test phase, $F(1, 28) = 19.73, p < .0001$, a trend toward a main effect of feedback, $F(1, 28) = 3.52, p = .07$, and a significant interaction between test phase and feedback, $F(1, 28) = 6.99, p < .01$. For the children in the treatment group, the mean absolute error declined from 24% ($SD = 9\%$) to 15% ($SD = 6\%$) ($t[31] = 4.45, p < .001, d = 1.18$), whereas for the children who were in the control groups, the mean absolute error did not differ by test phase (number line pretest, $M = 27\%, SD = 10\%$; number line posttest, $M = 24\%, SD = 11\%$). Finally, the treatment group’s posttest mean absolute error was significantly lower than the control groups’ posttest mean absolute error ($M = 24\%, SD = 9\%$), $F(1, 29) = 8.35, p < .01, d = 1.18$. Thus, among younger children, feedback had a large effect on short-term changes in the accuracy of numerical magnitude estimates.

I next examined whether younger children’s short-term changes in estimation accuracy were also accompanied by the hypothesized logarithmic-to-linear shift. On pretest, young children in both the treatment group and control
group initially provided median estimates for each number that were in fact fit better by the logarithmic regression function than by the linear one (see Figure 3.6). The precision of the fit of the logarithmic function, and the degree of superiority of that function to the linear function, was similar across the treatment group (log $R^2 = .92$; lin $R^2 = .58$) and control group (log $R^2 = .93$, lin $R^2 = .65$). In contrast, the treatment and control groups differed considerably in their number line posttest estimation patterns (see Figure 3.6). Children in the control group continued to generate estimates that fit the logarithmic function better than the linear one (log $R^2 = .91$, lin $R^2 = .78$). In contrast, children in the treatment group generated posttest estimates that fit the linear function substantially better than the logarithmic one (lin $R^2 = .91$, log $R^2 = .75$).

Figure 3.6 Short-term changes in numerical magnitude estimation by condition: Children in the treatment group provided estimates that were better fit by a logarithmic than linear function on pretest, and they provided estimates that were better fit by a linear function than logarithmic on posttest.
To determine whether the fit of the two functions merely arose from aggregating data over individual estimates, I also performed the same analyses for each individual participant’s set of estimates. As expected, before children received any training, the majority of children (83%) provided estimates that were better fit by the logarithmic function than by the linear one, regardless of whether they later received treatment (88% of children generated logarithmic estimates in the treatment group, 79% of children generated logarithmic estimates in the control group, Fisher’s Exact Probability Test: \( p = .62, ns \)). Furthermore, posttest estimates also indicated that treatment led to more children providing linear estimates: 69% of children who were in the treatment groups provided more linear than logarithmic estimates, whereas only 29% of children who were in the control groups provided more linear than logarithmic series of estimates (\( \chi^2 [1] = 4.82, p < .05 \)). Thus, as in Opfer and Siegler (2007), feedback on a very small (but strategic) set of estimation problems led to large changes in estimation accuracy.

I next examined changes in older children’s estimation performance by performing a 2 (test phase: pretest, posttest) X 2 (condition: treatment, control) repeated-measures ANOVA on the error scores. I found a main effect of test phase, \( F (1, 32) = 7.86, p < .01 \), and a significant test phase by condition interaction, \( F (1, 32) = 6.50, p < .05 \). For children in the control group, a paired t-test indicated that error scores did not differ from pretest (\( M = 11\%, SD = 8\% \)) to posttest (\( M = 11\%, SD = 9\% \)), \( t(17) = .35, ns \). For children in the treatment group, however, error scores declined significantly from pretest (\( M = 12\%, SD = 8\% \)) to posttest (\( M = 7\%, SD = 5\% \)), \( t(15) = 2.73, p < .05, d = .75 \). Unlike younger
children’s estimates, older children’s improvement in accuracy did not stem from their estimates becoming better fit by the linear function. For the treatment group, results of a paired t-test indicated that linearity of individual children’s pretest estimates ($M = .83$) did not differ from post-test estimates ($M = .90$), $t(15) = 1.42, ns$; for the control group, linearity of pretest ($M = .83$) and post-test ($M = .83$) estimates also did not differ.

In summary, short-term gains in accuracy of children’s numerical estimates came from feedback on a small set of estimates, which had a larger effect on younger children’s (logarithmic) estimates than on older children’s (linear) estimates. The short-term gains of younger children in response to treatment were also quite impressive: by post-test, treatment had boosted younger children’s accuracy to levels comparable to the accuracy of older children who received no treatment (15% versus 11%, $F[1, 33] = 2.8, ns$).

3.5.1.2 Rate of change. To address the rate of change in numerical estimation, I used logistic regression to examine the relation between generation of more linear than logarithmic patterns of estimates (linear model fitting best or not) and number of trial blocks of treatment (0-4), where 0 corresponded to the trial block prior to the administration of the treatment and thus 0 trials of treatment (feedback). First, I examined the effect of trial block for the treatment group of younger children. There was a significant positive effect of trial block for the treatment group, indicating that with each additional trial block the likelihood of generating linear estimates was 1.44 times as likely as the previous one, $\hat{\beta} = .37, z = 2.18, \text{Wald}(1, N = 80) = 5.02, p < .05$. A similar analysis found no significant effect of trial block for the younger children in the control group, indicating that time on task did not elicit change, nor did we find significant
effects for older children in either group (largely because they were already very likely to generate linear estimates).

To put this rate of change in context, it is useful to compare the average incremental change in a single trial block of training (x 1.44) to the average incremental change found with a year of real life experience in Experiment 2a (x 5.46). Taken literally, the comparison suggests that four trial blocks of training accomplished as much as nearly 11 months of real world experience. Of course, the caveat that has to be raised is that it is not at all clear whether the change occurred at a constant rate. Suggesting that the rate was not constant (at least over trial blocks), I observed a 6-fold increase in the proportion of younger children in the treatment group best fit by the linear function from Trial Block 0 (13%) to Trial Block 1 (81%), which is consistent with very rapid and abrupt learning. In the next section, I examined the abruptness of change more directly.

3.5.1.3 Path of change. Younger children could have moved from a logarithmic to a linear representation via several paths. To examine which path(s) they actually took, I examined the fit of the linear regression function to each individual child’s numerical estimates as a function of the number of trials that elapsed since the linear function provided a better fit than the logarithmic (i.e., when the logarithmic-to-linear shift was thought to occur). To measure this, I identified the first trial block on which the linear function provided the best fit to a given child’s estimates, and I labeled it “trial block 0.” The trial block immediately before each child’s trial block 0 was that child’s “trial block –1,” the trial block before that was the child’s “trial block -2,” and so on.

These assessments of the trial block on which children’s estimates first fit the linear function made possible a backward-trials analysis that allowed me to
test alternative hypotheses about the path of change from a logarithmic to a linear representation. One hypothesis, suggested by incremental theories of representational change (Brainerd, 1983), was that the path of change entailed gradual, continuous improvements in the linearity of estimates (and thus the fit of the linear regression function to their estimates). According to this hypothesis, the fit of the linear model would have gradually increased, from Trial Block -3 to Trial Block +3. In this scenario, Trial Block 0 — the first trial block in which the linear model provided the better fit — would simply mark an arbitrary point along a continuum of gradual, trial block-to-trial block improvement, rather than the point at which children first chose a different representation.

A second hypothesis was that the path of change involved a discontinuous switch from a logarithmic to a linear representation, with no intermediate state. This would have entailed no change in the fit of the linear model from Trial Block -3 to -1, a large change from Trial Block -1 to Trial Block 0, and no further change after Trial Block 0. This second hypothesis clearly fit the data. As illustrated in Figure 3.7, from Trial Block -3 to -1, a one-way ANOVA on the linear regression function, $F(4, 56)=.74, p > .05, ns$, indicated that there was no change in the fit of the linear function across these trial blocks. There also was no change from Trial Block 0 to Trial Block 3 in the fit of the linear function, $F(4, 92) = .8, p > .05, ns$. However, from Trial Block -1 to Trial Block 0, there was a large increase in the fit of the linear function to individual children’s estimates, from a median $R^2 = .43$ to a median $R^2 = .74$, $F(1, 48) = 8.85, p < .01, d = .85$. Thus, rather than Trial Block 0 reflecting an arbitrary point along a continuous path of improvement, it seemed to mark the point at which children switched from a logarithmic representation to a linear one.
Figure 3.7 Trial block-to-trial block changes: Backward trials graph of fit of linear model to children’s estimates on numerical magnitude estimation task. The 0 trial block is the block on which the linear function first provided a better fit to each child’s estimates; the -1 trial block is the block before that, and so on.

3.5.2 Transfer of Learning to Fractions

In the previous section, I observed that feedback on a small set of numerosities around 150 induced a large and broad change in both the accuracy and linearity of younger children’s numerical estimates. If this improved accuracy resulted from a representational change, I reasoned, learning should transfer to a superficially different context—the fraction line task—regardless of its costs in accuracy.
To examine transfer to the fractional magnitude context, I first calculated the mean absolute error for each child on the fractional magnitude estimation task (\(|\text{actual-estimate}| / \text{range of scale}\)) and regressed condition (1 = treatment, 0 = control) and age (7.15 – 9.64) against the mean absolute error scores. As expected, there was a significant interaction between age and condition, $F(3, 60) = 12.58, p < .001$ (see Table 3.2). For younger children (who had mostly generated logarithmic estimates in the numerical context), treatment had a large and negative effect on accuracy of fractional magnitude estimates, with the mean absolute error on the fraction-line task being larger for the treatment group ($M = 61\%, SD = 11\%$) than for the control group ($M = 41\%, SD = 16\%$), $F(1, 29) = 15.75, p < .001, d = 1.46$. In contrast, for older children (who had mostly generated linear estimates in the numerical context), treatment did not affect accuracy of fractional magnitude estimates (treatment, $M = 62\%, SD = 12\%$; control, $M = 61\%, SD = 11\%$), $F(1, 33) = .06, ns$. Thus, correcting younger children’s numerical estimates imposed a large cost on the accuracy of their fractional magnitude estimates.

<table>
<thead>
<tr>
<th>Number Line Accuracy</th>
<th>Fraction Line Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>$F_{3,60} = 12.58, p = .00001$</td>
</tr>
<tr>
<td></td>
<td>$t = 4.98, p &lt; .0001$</td>
</tr>
<tr>
<td>Older, $M = 91% &gt; $</td>
<td></td>
</tr>
<tr>
<td>Younger, $M = 80%$</td>
<td></td>
</tr>
<tr>
<td><strong>Treatment</strong></td>
<td>$t = 2.16, p = .01$</td>
</tr>
<tr>
<td>Treatment, $M = 89% &gt; $</td>
<td></td>
</tr>
<tr>
<td>Control, $M = 83%$</td>
<td></td>
</tr>
<tr>
<td><strong>Age x Treatment</strong></td>
<td>$t = 1.96, p = .05$</td>
</tr>
<tr>
<td>Younger: Treatment, $M = 85% &gt; $</td>
<td></td>
</tr>
<tr>
<td>Control, $M = 76%; p &lt; .007$</td>
<td></td>
</tr>
<tr>
<td>Older: Treatment, $M = 93%; Control, M = 89%; ns$</td>
<td></td>
</tr>
<tr>
<td><strong>Fraction Line Accuracy</strong></td>
<td>$F_{3,59} = 12.98, p &lt; .0001$</td>
</tr>
<tr>
<td></td>
<td>$t = 5.36, p &lt; .0001$</td>
</tr>
<tr>
<td>Younger, $M = 48% &gt; $</td>
<td></td>
</tr>
<tr>
<td>Older, $M = 38%$</td>
<td></td>
</tr>
<tr>
<td><strong>Treatment</strong></td>
<td>$t = 3.82, p = .0003$</td>
</tr>
<tr>
<td>Control, $M = 47% &gt; $</td>
<td></td>
</tr>
<tr>
<td>Treatment, $M = 38%$</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.2 Analysis of age differences at posttest.
To check whether this differing pattern of performance might be related to how younger children interpreted the meaning of the denominator, I next regressed their median estimate for each fractional value against the denominator of the fraction (see Figure 3.8). Specifically, I wanted to test my assumption that children’s estimates across the two tasks were quite similar, in spite of the superficial differences between the tasks and the costs this entailed for accuracy. As predicted, children in the control group, who were typically better fit by the logarithmic function at number line posttest, also provided a series of fractional estimates that were better fit by the logarithmic function ($R^2 = .89$) than by the linear function ($R^2 = .67$). In contrast, children in the treatment group, whose number line posttest estimates were typically better fit by the linear function, provided a series of fractional magnitude estimates that were better fit by the linear function ($R^2 = .87$) than were the estimates made by participants in the control group. Thus, children did appear to transfer their understanding of numeric value to the unfamiliar fractional context regardless of its cost in accuracy.
Figure 3.8 Short-term changes in fractional magnitude estimation:

Estimates of children in the control group (light circles) were better fit by a logarithmic function than by a linear function, whereas the estimates of children in the treatment group (dark circles) were better fit by a linear function than by a logarithmic function. Ideal pattern of performance ($y = 1/x$) is depicted in grey.

Finally, to determine whether linearity of performance on the number line posttest was correlated with performance on the fraction line posttest, I regressed individual participants’ linear $R^2$ values for each task. As expected, there was a high correlation between the two variables, $r = .62$, $F(1, 29) = 17.78$, $p < .001$. When I analyzed the two groups’ $R^2$ values on these tasks separately, I found that linear performance on the number line estimation task predicted linear performance on the fractional units task for the control group, $r = .67$, $F(1, 13) = 9.91$, $p < .01$, as well as for the treatment group, $r = .6$, $F(1, 15) = 7.98$, $p < .01$. Consistent with this finding, the degree of inaccuracy in the fractional magnitude context was strongly predicted by the degree of accuracy in the numerical magnitude context, $r = -.77$, $F(1, 29) = 42.44$, $p < .001$ (see Figure 3.9).
3.6 Discussion

Representational changes were hypothesized to lead to automatic transfer of learning, leading to gains in accuracy in one context coming at the expense of accuracy in another context. To test this hypothesis, I examined the accuracy of symbolic magnitude estimation across two contexts—whole numbers and fractions—and how these two contexts affected the relation between age and accuracy. In the next sections, I highlight the specific findings of my tests, and I remark on the broader implications of findings for future research on age differences in mathematical thinking, as well as implications of findings for
prevention of “cognitive illusions.”

3.6.1 Effect of Context on Age Differences

Across a broad range of contexts, including estimation of distance (Cohen, Weatherford, Lomenick, & Koeller, 1979), amount of money (Sowder & Wheeler, 1989), number of discrete objects (Hecox & Hagen, 1971), answers to arithmetic problems (LeFevre, Greenham, & Naheed, 1993), or locations of numbers on number lines (Siegler & Opfer, 2003), children’s estimates are highly inaccurate (see Siegler & Booth, 2005, for review). According to the representational change hypothesis, children’s inaccurate estimation across these many contexts stems from initial reliance on logarithmic representations of numerical value (Siegler & Opfer, 2003), a representation of numerical value that is consistent with Fechner’s Law and that is widespread among species, human infants, and time-pressured adults (see Dehaene et al., 1998, for review). In contrast, older children and adults are thought to have learned a linear representation of numerical value from encountering experiences in school and daily life that provide log discrepant information, which typically induces rapid switching from logarithmic to linear estimation patterns (Opfer & Siegler, 2007), and learned linear representations appear to generalize robustly to novel contexts (Laski & Siegler, 2007; Ramani & Siegler, 2008).

Although evidence of young children’s initially poor estimation skills and logarithmic representation of numeric value and evidence of older children’s good estimation skills and linear representations of numeric value has been drawn from a wide range of contexts, these contexts share an important property—accuracy could be attained either from representational changes (Joram, Subrahmanyam, & Gelman, 1998; Siegler & Opfer, 2003) or from
improving mathematical skills (Dowker et al., 1996; Hiebert & Wearne, 1986). To address this issue, I sought to provide a particularly strong test for the representational change hypothesis by examining short- and long-term changes on an estimation task—fractional magnitude estimation—that favors the logarithmic representation at the expense of the linear one. My reasoning was that if children learn to make better estimates by automatizing use of a linear representation, estimation of fractional magnitudes should suffer with age and experience.

The results of the present studies provided both correlational and experimental evidence supporting the hypothesis that numerical representations have a powerful effect on estimates of fractional magnitude. Correlational evidence was provided in Experiment 2a, where older children provided more accurate estimates of numerical magnitude than younger children, but younger children provided more accurate estimates of fractional magnitude than older children. Overall, inaccuracy of numerical estimates accounted for 64% of the variance in accuracy of fractional magnitude estimates, even when controlling for age. Experimental evidence from Experiment 2b indicated that this correlation reflected a causal link between numerical and fractional magnitude estimation. Specifically, children who were given training in numerical estimation subsequently provided more accurate estimates of numerical magnitude than did a control group that had received no training, but children in the treatment group also provided less accurate estimates of fractional magnitude than did the control group.

I believe that these findings of context effects provide particularly strong evidence for the hypothesis that numerical representations are automatic and can
thereby impose costs as well as benefits for accuracy of estimation. This idea is an important one because it simultaneously explains (1) high correlations among individual estimation tasks, (2) the breadth of transfer that is typically observed in training studies of numerical estimation, and (3) why adults are subject to certain cognitive illusions involving fractions, including incorrect comparisons of salaries (Opfer & DeVries, in press) and incorrect evaluation of medical risks (Burkell, 2004).

Finally, I believe that my results have a number of important educational implications. The most important implication is that efforts at improving estimation should not stop at improving numerical estimation. Rather, efforts at improving fractional magnitude estimation are likely to help the development of representations of fractional value, which—following the literature on development of numerical representations (Booth & Siegler, 2006; Laski & Siegler, 2007)—are likely to transfer to comparison of fractional values, to categorization of fractions as “small” or “large,” and to accurate similarity judgments of fractional magnitudes. Discoveries about educational interventions that lead to improvement of children’s numerical estimation (e.g., Opfer & Siegler, 2007; Ramani & Siegler, 2008; Siegler & Ramani, in press) suggest a number of interventions that could improve children’s representations of fractional magnitude. I believe these interventions are important given the inappropriateness of relying exclusively on linear representations of symbolic magnitude, and I believe the failure to provide children with effective intuitions about fractional value is lasting and may contribute to adults’ known difficulties with the fractions encountered in everyday life, including understanding of
statistics (Evans et al., 2002), addition of discounts (Chen & Rao, 2007), comparison of salaries (Opfer & DeVries, in press), and evaluation of medical risks (Burkell, 2004).
CHAPTER 4

WHY TRANSFER IS RARE IN THE LITERATURE

Experiment 3: The Trouble with Transfer: Insights from Microgenetic Changes in the Representation of Numerical Magnitude

Across the previous two sets of experiments, I set out to indicate that numerical representations are deployed automatically and this leads to benefits in estimation accuracy when the representation that participants use is appropriate to the task at hand (e.g., when scaling up the linear representation from the 0-100 context to larger numerical contexts with the help of a progressive alignment methodology), but can also lead to costs when the representation that is automatically deployed, like a linear representation of number, is inappropriate for the transfer context (e.g., fractional magnitude estimation task).

Despite the fact that I have indicated that there are numerous situations in which participants automatically deploy a particular representation to a transfer context, why might it be the case that sometimes transfer of learning is so very difficult to elicit? The vast literature on transfer of learning has offered many specific theories as to what is the underlying cause for the narrowness of transfer. To add to this already daunting list, I propose a novel (yet somewhat more optimistic) explanation for why transfer is difficult to elicit in microgenetic studies in particular and possibly other studies more generally—proactive interference from previous practice. The basic premise of my account is that
when children have practice on a task without any feedback (e.g., when they complete a pretest on a transfer task), children must use some representation to complete that task, and the more they use that representation, the greater the strength of the representation, and the more likely children will continue using it in the future when that task is encountered again (Erdelyi & Becker, 1974; Roediger & Payne, 1982; Siegler & Shipley, 1995). Under circumstances in which the representations used on the task are already appropriate, pretests can facilitate transfer (“practice makes perfect”) (Roediger & Karpicke, 2006; Gick & Holyoak, 1983). However, when representations are inappropriate, practice on a pretest makes imperfect because practice merely strengthens the inappropriate representation used on the transfer task and thereby blocks transfer of the more optimal representations learned during training (Roediger & Marsh, 2005; Gick & Holyoak, 1980). If true, this explanation is not a trivial one: it suggests that the trouble with transfer is at least partly an experimental artifact of studies that pretest the treatment and control groups.

To test my practice interference hypothesis, I used Solomon and Lessac’s (1968) four-group design to assess the independent and interactive effects of treatment (feedback) and pretesting on children’s judgments of numerical magnitude. The tasks I used are ones where previous cross-sectional (Laski & Siegler, 2007; Siegler & Opfer, 2003) and microgenetic (Opfer & Siegler, 2007) studies had shown that young children use an inappropriate (logarithmic) representation before using an appropriate (linear) one (e.g., younger children judge 150 to be closer to 1,000 than to 1, whereas older children judge 150 to be closer to 1 than to 1,000). In the first task, children were asked to make estimates of numerical quantity (e.g., where 150 would fall on a line flanked by 0 and
and they were given feedback on their estimates so they would learn to
use a linear representation (as in Opfer & Siegler, 2007). In the second task,
children were asked to categorize numerals by their magnitude (e.g., whether
150 is a small or big number in the context of the 0-1,000 range); this is the task
where I hoped children would transfer their learning. On the assumption that the
two tasks tap a common representation, I predicted that children’s learning of
the linear representation on the number line estimation task would transfer to
their category judgments (at least when they were not given a categorization
pretest). Further, based on my practice interference hypothesis, I predicted that
pretesting on the categorization task would also strengthen the logarithmic
representation and thus block spontaneous transfer of learning.

In the next paragraphs, I briefly examine the literature indicating why
transfer might be difficult for children, and then describe a microgenetic
experiment which I designed to test my explanation for children’s trouble with
transfer in this numeric domain and more broadly.

My proposal for why children have trouble with transfer in these
situations is inspired by Duncker’s (1945) classic study of problem-solving.
Duncker tested whether participants could apply novel functions to various
familiar objects (e.g., a matchbox, tacks, and candles). For example, to solve the
problem of placing three small candles on a door at eye-level, participants had to
conceive of a novel function for the matchbox (i.e., to serve as a platform).
Normally, about 86% of such problems were solved; however, if subjects were
given a pretest to determine their knowledge of the original functions of the
familiar objects, the rate of problem solving dropped to 58%. Duncker’s
explanation was that previous experience with the objects induced a “functional fixedness” that inhibited their novel solutions.

Might prior experience broadly induce a kind of ‘conceptual fixedness’ that will also prevent children from transferring their knowledge? For example, one reason that children may fail to transfer from school lessons to familiar real-world problems (yet succeed in transferring to novel problems like throwing darts at an underwater dartboard) is that their previous successes on real-world problems interferes with the application of novel school lessons, much like Duncker’s pretest interfered with his subjects’ ability to think of novel solutions. That is, previous experiences in school and real-world settings interfere with transfer because they lead children to think that symbolic operations are “for school” and not “for real-world problems,” much like Duncker’s subjects thought of matchboxes as “for holding matches” and not “for supporting candles.”

Although my application of Duncker’s findings to transfer is novel, it is also consistent with previous findings on the formation of associations between strategies and problems (Shrager & Siegler, 1998), the formation of undesirable memory traces through practice (Roediger & Payne, 1982), and with the induction of “mental sets” in analogy (Gick & Holyoak, 1980) and problem-solving (Luchins, 1942; Luchins & Luchins, 1950). Moreover, pretests can sometimes harm animals’ learning as well (Lessac & Solomon, 1969). In Solomon and Lessac’s (1965, 1968, 1969) studies, for example, beagle pups were reared in isolation or a kennel, and half were pretested across a range of tasks (conditioning, perceptual, motor) to assess the effects of pretesting. When pups were later tested after the isolation/kennel rearing, results indicated that
pretesting protected the isolated pups from many—but not all—deleterious effects of social isolation. In one spatial-reasoning task, for example, performance of pretested isolated pups decreased 10-fold from pretest to posttest, which was much worse than posttest performance of the isolated but unpretested pups.

The potential generality of findings of pretest effects is especially important for microgenetic studies. If true, the hypothesis that pretesting inhibits transfer performance could explain why children’s capacity for transfer is so seldom evident in microgenetic studies that repeatedly test children over many trials, sessions, or days. Moreover, this possibility is an important one because it suggests a real-world situation where transfer might be difficult (i.e., for problems where inappropriate approaches are practiced repeatedly), a situation where transfer might be easy (i.e., for problems where children do not practice inappropriate approaches repeatedly), and a very general methodological control for assessing the impact of unsupervised practice (i.e., manipulating the pretesting variable).

Since children’s estimates and numerical category judgments appear to tap a common representation that changes in a coherent fashion with age and experience, I anticipated a high potential for children transferring representations learned in an estimation context to performance in a categorization context. However, based on my practice interference hypothesis, I am also concerned that the pretest commonly used in studies of children’s learning might impede this transfer by strengthening the logarithmic representation.
To test my practice interference hypothesis, I conducted a microgenetic study in which I examined the process of change in number line estimation, the transfer of representations from an estimation context to a categorization context, and the impact of pretesting on the categorization task for later transfer. To examine the change process in number line estimation, I employed the basic design previously used by Opfer and Siegler (2007). To examine the transfer of representations from estimation to categorization, I additionally examined performance on a number categorization task that I adapted from Laski and Siegler (2007). Finally, to test the hypothesis that using representations interferes with the ability to change them, I structured the microgenetic study using the Solomon and Lessac (1968) four-group design rather than the traditional two-group design used in the 105 microgenetic studies reviewed by Siegler (2006). Specifically, the Solomon and Lessac (1968) design allowed me to examine the effect of pretesting the transfer (categorization) task on subsequent performance on that task.

Although the purpose and design of microgenetic studies have been widely discussed (see Siegler, 2006 for a recent review), the purpose and features of the Solomon and Lessac (1968) design require some explanation. The purpose of the design is to assess change from pretest to posttest while controlling for the effect of the pretest itself, which is often important in studies of isolation (e.g., Lessac & Solomon, 1969), sensory deprivation (Rubel, 1984; Conlee & Parks, 1981), and attachment (van den Boom, 1994, 1995). To accomplish this goal, the design calls for an administration of posttest to four groups that vary orthogonally in the administration of pretest and treatment: Group I received
both pretest and treatment; Group II received no pretest but did receive
treatment; Group III received a pretest but no treatment; and Group IV received
neither pretest nor treatment. This design provided two major benefits. First,
strategic comparisons among the four groups allowed me to anchor
interpretations of developmental change to absolute as well as to relative data.
Second, it allowed for the test of pretest effects and the interaction between pre-
treatment testing and treatment itself.

Although there are many possible interactions between pretest and
treatment, I illustrated two of the most interesting ones for my study in Figure
4.1: the case where pretesting the treatment group on the transfer task inhibits the
effect of treatment (i.e., transfer of learning from the training task to the transfer
task), and the case where pretesting the treatment group on the transfer task
enhances the effect of treatment. The first case directly corresponds to the
predictions of my pretest interference hypothesis, whereas the second case
corresponds to findings that would falsify my hypothesis.
Figure 4.1 Two interactions detectable by four-group design: In the first case (panel a), pretesting inhibits treatment effects. In second case (panel b), pretesting enhances treatment effects.

4.2 Method

4.2.1 Participants

Participants were 56 first and second grade students ($M = 7.85$, $SD = .65$; 29 girls, 27 boys). The children attended neighborhood schools in largely European-American, middle class suburbs surrounding a large metropolitan city in Midwestern United States. One of two female research assistants served as experimenter.

4.2.2 Number Line Task

The number line task was the same as reported in Experiment 2a and 2b.

4.2.3 Categorization Task

In the categorization task, children were asked to say how large numbers
were when compared to 0 (really small) and 1,000 (really big). To do this, children were given the following instructions (adapted from Laski and Siegler, 2007): “I’m going to ask you what you think about some of the numbers between 0 and 1,000. Some of these numbers are really small, some are small, some are medium, some are big, and some are really big. I’m going to say a number, and you need to tell me if you think the number is a ‘really small’ number, a ‘small’ number, a ‘medium’ number, a ‘big’ number, or a ‘really big’ number.” The experimenter will then tell the children that they may refer to boxes to help them remember all of their choices. Then the experimenter set out five labeled, identically-sized boxes, one at a time and from left to right, and read each label in turn, “really small,” “small,” “medium,” “big,” and “really big.”

To orient the participants to the endpoints and to ensure that they understood the task, children were first asked to categorize 0 and 1,000. On these two practice trials (and no others), the experimenter provided feedback if the participant did not categorize 0 as “really small” and 1,000 as “really big.” The other 10 numbers children were asked about comprise a subset of the numbers (2, 5, 78, 100, 150, 246, 486, 606, 725, and 938) used in the number line task, and these numbers were randomized for each participant.

4.2.4 Design and Procedure

Children were randomly assigned to one of four groups (see Table 4.1): a pretested treatment group (Group I), an unpretested treatment group (Group II), a pretested control group (Group III), or an unpretested control group (Group IV). The four groups differed in whether they were pretested on the transfer task (number categorization) and in whether they received treatment (i.e., feedback) during the training task (number line estimation). Participants in Group I
received the categorization task (where I hoped for transfer) immediately prior to and following their completion of number line problems, and they were also given feedback on number line problems. Participants in Group II received the categorization task only following their completion of number line problems, and they were also given feedback on these number line problems. Participants in Group III received the categorization task immediately prior to and following their completion of number line problems, but they were not given feedback on number line problems. Participants in Group IV received the categorization task only after their completion of number line problems, but they too were not given feedback on number line problems during training.

As shown in the outline of the procedure in Table 4.1, children in all four groups completed the number line estimation task for a pretest, three training trial blocks and a posttest. The purpose of these three phases (pretest, training trial blocks, and posttest) was to ensure that learning had occurred prior to the transfer task (i.e., on the number line posttest) and to examine the course of learning prior to posttest (i.e., to examine changes from the number line pretest through posttest). On the number line pretest and posttest, children in all four groups were presented the same 22 problems without feedback (i.e., without treatment). For children in the two treatment groups, each training trial block included a feedback phase and a test phase. As shown in Table 4.1, the feedback phase of each training trial block included either one item on which children received feedback (Trial Block 1) or three items on which they received feedback (Trial Blocks 2 and 3). The test phase in all three training trial blocks included 10 items on which children did not receive feedback; this test phase occurred immediately after the feedback phase in each training trial block. Children in the
two control groups received the same number of estimation problems, but they never received treatment.

Feedback was administered to the two treatment groups (Group I and Group II) following the same procedure used in Opfer and Siegler (2007); the treatment procedure was administered as in Experiment 2b.

Table 4.1 Experimental groups.

<table>
<thead>
<tr>
<th>Experimental Groups</th>
<th>Pretest</th>
<th>Number line Training</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Pretested Treatment Group</td>
<td>10 fs (0-1000)</td>
<td>22 fs (0-1000)</td>
<td>150 10 fs (0-1000) 3 fs (147-187) 10 fs (0-1000) 3 fs (147-187) 10 fs (0-1000) 22 fs (0-1000) 10 fs (0-1000)</td>
</tr>
<tr>
<td>II. Unpretested Treatment Group</td>
<td>None 22 fs (0-1000)</td>
<td>150 10 fs (0-1000) 3 fs (147-187) 10 fs (0-1000) 3 fs (147-187) 10 fs (0-1000) 22 fs (0-1000) 10 fs (0-1000)</td>
<td></td>
</tr>
<tr>
<td>III. Pretested Control Group</td>
<td>10 fs (0-1000) None 22 fs (0-1000) None 10 fs (0-1000) None 10 fs (0-1000) None 10 fs (0-1000) 22 fs (0-1000) 10 fs (0-1000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV. Unpretested Control Group</td>
<td>None 22 fs (0-1000) None 10 fs (0-1000) None 10 fs (0-1000) None 10 fs (0-1000) None 10 fs (0-1000) 22 fs (0-1000) 10 fs (0-1000)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.3 Results

I organized my results into two sections: (1) results concerning the process of change in numerical estimation, and (2) results concerning transfer of learning to numerical categorization. Within the first section, I report on the conditions that led to changes in numerical estimation (Source of change), how quickly those changes occurred (Rate of change), approaches that children used up to and following the use of mature approaches (Path of change), and individual differences in learning (Variability of change). Within the second section, I report
results testing my hypotheses about transfer (*Breadth of change*). These dimensions of change have proven useful in prior microgenetic studies and in characterizing cognitive change more broadly (Siegler, 1996, 2006).

4.3.1 *Process of Change in Estimation*

4.3.1.1 *Source of change.* I first examined the source of change in estimation performance on the number line task. Specifically, I wanted to test whether the experiences that children received during the training phase of the experiment improved their estimation accuracy and influenced the degree to which their estimates came to follow a linear function. To find out, I compared number-line pretest and posttest estimates of the treatment groups to the control groups. Throughout this section, I collapsed the two treatment groups and two control groups because there was never an effect of testing numerical categorization on children’s subsequent numerical estimates, $F_s < 1.25$.

To examine the effect of treatment on estimation accuracy, I first calculated the mean absolute error for each child ($|\text{actual-estimate}| / \text{range of scale}$) and then performed a 2 (condition: treatment, control) X 2 (test phase: number-line pretest, number-line posttest) repeated-measures ANOVA on the error scores. As expected, test phase interacted with condition, $F(1, 54) = 4.31, p < .05$. For the children in the treatment groups, the mean absolute error declined from 22% ($SD = 14\%$) to 17% ($SD = 1\%$) ($t [23] = 2.2, p < .05, d = .41$), whereas for the children who were in the control groups, the mean absolute error did not differ by test phase (number line pretest, $M = 19\%, SD = 6\%$; number line posttest, $M = 20\%, SD = 8\%$).

I next examined whether these changes in estimation accuracy were also accompanied by the hypothesized logarithmic-to-linear shift. As shown in Figure
4.2, on the number line pretest, children’s mean estimates for each number were better fit by the logarithmic regression function than by the linear one, regardless of experimental condition. The precision of the fit of the logarithmic function, and the degree of superiority of that function to the linear function, was similar across the treatment (log $R^2 = .95$; lin $R^2 = .80$) and control groups (log $R^2 = .93$, lin $R^2 = .82$). In contrast, the groups differed considerably in their posttest estimation patterns (see Figure 4.2). Children in the control groups continued to generate estimates that fit the logarithmic function better than the linear one (log $R^2 = .92$, lin $R^2 = .85$), with their median estimates being almost identical from pretest to posttest (hence the large overlap between the pretest and posttest series in Figure 4.2). In contrast, children in the treatment groups generated posttest estimates that fit the linear function substantially better than the logarithmic one (lin $R^2 = .96$, log $R^2 = .69$).
Figure 4.2 Source of change: Best fitting functions for mean estimates on number line (training) task at pretest (indicated in black) and posttest (indicated in grey) for treatment and control groups. Solid function lines indicate that the function fit the data significantly better than the alternative model did. Error bars indicate standard errors.

To determine whether the fit of the two functions merely arose from aggregating data over individual estimates, I also performed the same analyses for each individual participant’s set of estimates. As expected, before children received any training, the majority of children (66%) provided estimates that were better fit by the logarithmic function than the linear one, regardless of whether they later received treatment (63% of children generated logarithmic estimates in the treatment groups, 69% of children generated logarithmic estimates in the control groups, $\chi^2 [1] = .24, ns$). Furthermore, posttest estimates also indicated that receiving the treatment led to more children providing linear estimates: 83% of children who were in the treatment groups provided more linear than logarithmic estimates, whereas only 34% of children who were in the
control groups provided more linear than logarithmic series of estimates ($\chi^2 [1] = 13.30, p < .001$).

Thus, as in Opfer & Siegler (2007), feedback on a very small (but strategic) set of estimation problems led to large changes in estimation accuracy and in an overall pattern of estimates that was more characteristic of adults than children. In the next section, I examined just how small this set of feedback problems could be to produce the large changes observed in the experimental session.

4.3.1.2 Rate of change. To address the rate of change in numerical estimation, I used logistic regression to examine the relation between generation of more linear than logarithmic patterns of estimates (linear model fitting best or not) and number of trial blocks of treatment (0-4), where 0 corresponded to the trial block prior to the administration of feedback (i.e., pretest) and thus 0 trials of feedback. I also included experimental group (treatment versus control, as above) and the interaction of experimental group and trial block as predictor variables in the logistic regression model. My key prediction was that training group and trial block would interact, with the interaction occurring due to children in the treatment group generating linear estimates after fewer trial blocks than those in the control group.

As predicted, there was a significant interaction between number of trial blocks (0-4) and experimental group (treatment or control) on likelihood of generating linear patterns of estimates, when controlling for the other predictors in the model, $\hat{\beta} = .48$, $z = 2.48$, Wald $(3, N = 280) = 38.11$, $p < .05$. This interaction between experimental group and trial block (Figure 4.3) reflected different rates of learning in the treatment and control groups. For the treatment group, there
was a significant positive effect of trial block on the likelihood of generating linear estimates, $\hat{\beta} = .52$, $z = 3.34$, Wald (1, $N = 120$) = 12.49, $p < .001$, indicating that with each trial block children were 1.62 times as likely to generate linear estimates. For the control group, however, there was no effect of trial block on the likelihood of generating linear estimates, $\hat{\beta} = .04$, $z = 0.35$, Wald (1, $N = 160$) = .12, ns.

Figure 4.3 Rate of change: Trial block-to-trial block changes in percentage of children whose estimates were best fit by the linear function on the number line (training) task.
I next looked at how quickly children responded to treatment by using logistic regression to examine the effect of treatment on generating linear estimates in each trial block (see Table 4.2). On trial block 0, before feedback was given to the treatment group, the treatment and control groups did not differ in how often they generated linear estimates, $\hat{\beta} = .10$, $z = .17$, Wald $(1, N = 56) = .03$, ns. On trial block 1, after children in the treatment group were given feedback only on the magnitude of 150, the effect of treatment was immediately evident, $\hat{\beta} = 1.85$, $z = 2.97$, Wald $(1, N = 56) = 10.08$, $p < .005$, with children in the treatment group being 6.33 times as likely to generate linear estimates as children in the control group. Further, on trial blocks 2-4, children in the treatment groups continued to generate linear patterns of estimates more frequently than children in the control groups, $\hat{\beta} = 1.55$, $z = 4.48$, Wald $(1, N = 168) = 22.12$, $p < .0001$, with children in the treatment group being 4.7 times as likely to generate linear estimates on each trial block as children in the control group. What this meant was that treatment effects manifested and persisted after feedback on a single estimate.

<table>
<thead>
<tr>
<th>Trials of Treatment</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency of Linear Best</td>
<td>Odds of Linear Best</td>
<td>Frequency of Linear Best</td>
<td>Odds of Linear Best</td>
<td>Frequency of Linear Best</td>
</tr>
<tr>
<td>Treatment Groups</td>
<td>0.33 0.5</td>
<td>0.79 3.8</td>
<td>0.67 2</td>
<td>0.83 5</td>
<td>0.83 5</td>
</tr>
<tr>
<td>Control Groups</td>
<td>0.31 0.45</td>
<td>0.38 0.6</td>
<td>0.46 0.88</td>
<td>0.4 0.68</td>
<td>0.34 0.52</td>
</tr>
</tbody>
</table>

Table 4.2 Rate of change in training phase.
Two types of evidence were consistent with this portrait of fast and persistent learning. The first type of evidence came from the trial of discovery, that is, the number of trials before children used the linear representation for the first time. For this analysis, I necessarily excluded children (n = 13) whose estimates were never better fit by the linear function (indicating no discovery) and children whose estimates were already better fit by the linear function on pretest (n = 18) (indicating that the discovery had already occurred). Of the remaining 25 children (10 from control group, 15 from treatment group), the fastest learners, children whose estimates were better fit by the linear model on trial block 1, were assigned a score of 1; the slowest learners, children whose estimates were better fit by the linear model for the first time on the posttest, were assigned a score of 4. An ANOVA indicated a trend toward a difference in the rate of discovery, \( F(1, 24) = 3.48, p < .10, \, d = .70, \) with the first trial block on which the linear function provided a better fit occurring slightly earlier in the treatment groups than in the control groups \((M = 1.47 \text{ trial blocks, } SD = .70 \text{ versus } M = 2.20 \text{ trial blocks, } SD = 1.29).\) Thus, even when I excluded children who never used the linear representation (principally from the control group), treatment still had the effect of leading children to use the representation earlier (typically after feedback on a single numerosity) than they would have had they not received treatment.

The second type of evidence for faster learning in the treatment groups came from children’s trial block of last error, which was the trial block on which children’s estimates were last best fit by the logarithmic function. By this measure, too, treatment led to earlier adoption of the linear representation. For children in the control groups, the average trial block of last error \((M = 3.6, SD =\)
.15 of 4 trial blocks) occurred later than for children in the treatment groups ($M = 1.67, SD = .54$ of 4 trial blocks), $F(1, 24) = 15.29, p < .001, d = 4.87$. Thus, once children discovered the linear representation following treatment ($M = 1.47$ trial blocks), they generally continued to use it throughout the experimental session.

In summary, both the rapidity of the change in estimates and its stability once it was made suggest that the change was made at the level of the entire representation, rather than as a local repair, a conclusion that the data on the path and breadth of change (in the next sections) also supported.

4.3.1.3 Path of change. Children could have moved from a logarithmic to a linear representation via several paths. To examine which path(s) they actually took, I examined the fit of the linear regression function to each individual child’s numerical estimates as a function of the number of trials that elapsed since the linear function provided a better fit than the logarithmic (i.e., when the logarithmic-to-linear shift was thought to occur). To measure this, I identified the first trial block on which the linear function provided the best fit to a given child’s estimates, and I labeled it “trial block 0.” The trial block immediately before each child’s trial block 0 was that child’s “trial block –1,” the trial block before that was the child’s “trial block -2,” and so on.

These assessments of the trial block on which children’s estimates first fit the linear function made possible a backward-trials analysis as conducted in Experiment 2b. As illustrated in Figure 4.4, from Trial Block -3 to –1, a one-way ANOVA on the linear regression functions indicated that there was no change in the fit of the linear function ($F < 1, ns$). There also was no change from Trial Block 0 to Trial Block 3 in the fit of the linear function ($F < 1, ns$). However, from Trial Block –1 to Trial Block 0, there was a large increase in the fit of the linear function...
to individual children’s estimates, from an average $R^2 = .46$ to an average $R^2 = .68$, $F (1, 80) = 12.20, p < .001, d = 2.67$. Thus, rather than Trial Block 0 reflecting an arbitrary point along a continuous path of improvement, it marked the point at which children switched from a logarithmic representation to a linear one.

Figure 4.4 Path of change: Backward trials graph of fit of linear model to children’s estimates on number line (training) task. The 0 trial block is the block on which the linear function first provided a better fit to each child’s estimates; the -1 trial block is the block before that, and so on. The N’s indicate the number of children who contributed data at each trial block; thus, 43 children used the linear representation on a least one trial block and therefore contributed data to Trial Block 0, 41 of these children had at least one trial block after this point and therefore contributed data to Trial Block 1, and so on.
4.3.1.4 Variability of change. I last examined individual differences in responses to treatment. Previous work (Opfer & Siegler, 2007) found that children whose initial representations were consistently logarithmic responded to treatment by adopting representations that were more consistently linear than did children whose initial representations were less consistently logarithmic. In contrast, the fit of the linear function to children’s number-line pretest estimates did not predict the fit of the linear function to their number-line posttest estimates. The explanation previously offered for this finding was that the difference between the children’s estimates and the feedback they received was more dramatic, and thus more likely to motivate a shift to the alternative (linear) representation, among children whose initial estimates were most strongly logarithmic.

In an attempt to replicate this finding, I regressed the percent of variance in number-line pretest estimates accounted for by the logarithmic function against percent variance in number-line posttest estimates accounted for by the linear function for the children in the treatment groups. As previously found, the fit of the logarithmic model to each child’s number-line pretest estimates (mean log $R^2 = .72$, $SD = .20$) predicted the fit of the linear model to the child’s number-line posttest estimates (mean lin $R^2 = .71$, $SD = .34$) ($r = .5$, $F [1, 23] = 7.36$, $p < .05$). That is, the better the logarithmic model fit the children’s number-line pretest estimates, the better the linear model fit their number-line posttest estimates. Surprisingly, the fit of the linear function to children’s number-line pretest estimates (mean lin $R^2 = .62$, $SD = .29$) also predicted the fit of the linear function to their number-line posttest estimates, ($r = .8$, $F [1, 22] = 38.29$, $p < .01$), whereas it did not in Opfer and Siegler (2007). Thus, it seems that having consistent
representations of numeric value (whether accurate or not) produces the greatest learning.

4.3.2 Transfer of Learning to Numerical Categorization

Finally, to test my practice interference hypothesis, I examined the breadth of changes in children’s numerical judgments by examining whether children transferred learning on number line problems to their performance on the number categorization task and whether this transfer was impeded by previous practice on the categorization task.

To examine these issues, I first analyzed the relation between numerical value and categorization judgments at pretest. To do so, category labels were converted to a numeric code (i.e., “very small” = 0, “small” = 1, “medium” = 2, “big” = 3, and “very big” = 4), and then I examined the fit of the linear and logarithmic regression functions to the mean judgments. As on the number line task, children’s pretest magnitude judgments were again better fit by a logarithmic ($\log R^2 = .95$) than by a linear function ($\text{lin } R^2 = .69$). These fits of the logarithmic function did not result from aggregating over subjects. Of all the children who received a categorization pretest, 90% of individual children provided patterns of judgments that were better fit by the logarithmic than linear function, with the average fit of the logarithmic function to each child’s judgments ($\text{mean } \log R^2 = .78$, $SD = .03$) being significantly better than the average fit of the linear function ($\text{mean } \text{lin } R^2 = .59$, $SD = .05$), $t(30) = 5.71$, $p < .001$, $d = 4.61$. Furthermore, pretest performance on the two tasks was generally correlated. The more linear were the estimates on the number line task, the more linear were the judgments on the number categorization task ($r = .52$, $F [1, 30] = 10.92$, $p < .01$), and the more logarithmic were the estimates, the more logarithmic
were the judgments on the number categorization task ($r = .72, F [1, 30] = 30.41, p < .001$). Thus, it appeared that a common, logarithmic representation of numeric value influenced children’s categorization as well as estimation performance.

As an overall measure of transfer, I next examined whether individual differences in learning to generate linear estimates on the number line estimation task (as measured by the fit of the linear regression function to estimates on the number-line posttest) were associated with individual differences in the linearity of judgments on the number categorization task (also measured by the fit of the linear regression function). If children who had learned to generate linear estimates on the number line task failed to generalize their learning to the categorization task, I would expect no correlation between the linearity of judgments on the two tasks. This was not the case, however. Rather, the linearity of judgments across the two tasks were highly correlated, $r = .67, F (1, 55) = 44.42, p < .0001$. Could this correlation have arisen simply because the two tasks tapped a third factor unrelated to learning? If so, the correlations would be expected to be equally high in both the training and control groups. This was not the case, either. Rather, the correlation in performance across the two tasks was very high for the treatment groups ($r = .91$), as would be expected by transfer of learning, but very low for the control groups ($r = .20$), as would be expected with no transfer of learning to the categorization context.

I next examined the effect of categorization pretesting on transfer. As expected, category judgments on posttest varied substantially with the administration of treatment and categorization pretest (see Figure 4.5). At the group level, the linear function provided a better fit for the mean judgments of children in the unpretested treatment group ($\text{lin } R^2 = .84$) than in the pretested
treatment ($\text{lin } R^2 = .72$), pretested control ($\text{lin } R^2 = .75$), and unpretested control ($\text{lin } R^2 = .68$) groups. The same pattern emerged when looking at the proportion of children who were best fit by each function, with 46% of children’s judgments in the unpretested treatment group being best fit by the linear function versus 23% of children in the pretested treatment group, 21% of children in the unpretested control group, and 17% of children in the pretested control group.

![Figure 4.5 Breadth of change. Fit of linear model to children’s category judgments (transfer task) across four groups: Pretested treatment group (Group I), Unpretested treatment group (Group II), Pretested control group (Group III), Unpretested control group (Group IV).](image)

Finally, to test for the predicted interaction between categorization pretesting and treatment, I conducted a 2 (categorization pretesting: yes, no) X 2
(treatment: yes, no) factorial ANOVA on the fit of the linear function for each child’s judgments. Categorization pretesting and treatment produced no main effects, but there was a substantial interaction between the two variables, \( F(1, 56) = 5.32, p < .05 \). The unpretested treatment group provided significantly more linear judgments (mean \( R^2 = .76, SD = .02 \)) than did the pretested treatment group (mean \( R^2 = .54, SD = .07 \)), \( t(22) = 2.43, p < .05, d = 4.27 \), and they also provided slightly more linear judgments than did the control groups (pretested, mean \( R^2 = .64, SD = .03 \); unpretested, mean \( R^2 = .59, SD = .07 \); \( p’s < .07 \)), which did not differ from each other. To appreciate just how powerful transfer was when children received treatment but no categorization pretest, it is useful to compare the unpretested treatment group’s performance on the categorization task (shown in Figure 4.5) to their last trial blocks of performance on the estimation task (shown in Figure 4.4): the variance accounted for by the linear function in the two tasks is nearly identical (estimation: mean lin \( R^2 = .79 \); categorization, mean lin \( R^2 = .76 \)), which is consistent with an almost perfect transfer of a linear representation across the two contexts.

In summary, the results regarding transfer supported two major conclusions. First, changes in estimation again seemed to occur at the level of the whole representation rather than as a series of local (i.e., task- and treatment-specific) repairs, thereby leading children to generalize a little treatment both to new numerical ranges within the same task (the number line task) and to the whole numerical range on a new task (the number categorization task). Second, as predicted by the pretest interference hypothesis, the initial use of the logarithmic representation on the number categorization pretest made it more difficult to change performance on this task, thereby leading to the finding that
quite robust transfer was almost completely blocked by the administration of a categorization pretest.

4.4 Discussion

Spontaneous transfer of learning is notoriously difficult to elicit (Barnett & Ceci, 2002; Gick & Holyoak, 1983; Thorndike, 1922), even in microgenetic studies that allow one to control for the amount of learning that occurs prior to administration of the transfer task (Siegler, 2006). I hypothesized that one source of this difficulty comes from previous experience with the transfer task, such as that provided on a pretest. To test my hypothesis about this difficulty with transfer, I used Solomon and Lessac’s (1968) four-group design to examine how children acquired linear representations of numerical magnitude on an estimation task (see Process of Change in Numerical Estimation), whether they transferred these representations to a new context in which they were asked to categorize numbers, and whether pretesting interfered with this transfer (see Transfer of Learning to Numerical Categorization).

Regarding the process of change in numerical estimation, the present study lent further support to the representational change hypothesis about development of estimation (Opfer & DeVries, in press; Opfer & Siegler, 2007; Siegler & Booth, 2004; Siegler & Opfer, 2003), replicating results from the only previous microgenetic study of numerical estimation (Opfer & Siegler, 2007) in fine detail. As in Opfer and Siegler (2007), I found that feedback in a single session yielded a large increase in accuracy of estimates, and this change was associated with a shift from generating logarithmic to linear estimates (see Source of change). Further, this change occurred after feedback on the magnitude of just one numeral—150—and was remarkably stable thereafter (see Rate of change),
with the linearity of an individual child’s estimates typically increasing greatly in one trial rather than gradually over many trials (see Path of change). These three findings were important because they supported the idea that children’s estimates improved due to a shift from a logarithmic to a linear representation of numerical quantity.

Regarding the transfer of learning to numerical categorization, my study of transfer (see Transfer of Learning to Numerical Categorization) revealed surprisingly robust transfer of learning from the estimation context to the categorization context—at least when no categorization pretest was administered. That is, children who were given feedback and no pretest (i.e., the unpretested treatment group) categorized 1 as “very small,” the numbers 78 and 150 as “small,” and 1,000 as “very large,” which was consistent with the linear representation they learned to use in the estimation context. Indeed, the most striking evidence for transfer of representations across the two contexts came from performance of the unpretested treatment group shown in Figure 4.5 (which depicts linearity of number categorization) to the last trial blocks shown in Figure 4.4 (which depicts linearity of number-line estimates after learning to generate linear representations): the variance accounted for by the linear function in the two tasks is nearly identical (estimation, mean lin $R^2 = .79$; categorization, mean lin $R^2 = .76$). This performance is consistent with an almost perfect transfer of a linear representation across the two contexts. In contrast, when children were given neither treatment nor categorization pretest, they categorized the number 1 as “very small,” the number 78 as “medium,” and the numbers 150 and 1,000 as “very large,” which was consistent with use of a logarithmic representation. Indeed, as Figure 4.5 depicts, children who had received no
feedback (i.e., children in the control groups) and children who had been pretested on the transfer task (i.e., children in the pretested groups) generated category judgments that were just as non-linear before training as after training. Thus, without the administration of a categorization pretest, I observed nearly perfect transfer to the categorization task; with the administration of a categorization pretest, I observed virtually no transfer at all.

4.4.1 Why Did the Pretest Interfere with Transfer?

One mechanism that might account for the pretest interfering with transfer is simple fatigue. Within this hypothesis, my giving pretested children so many problems to solve reduced their performance relative to the unpretested children because the pretest caused children to become tired of the experiment. As a direct test of whether fatigue caused the pretested and unpretested groups to differ, I ran an additional condition consisting of 20 children, who were given an equally long, but non-numeric pretest (identifying the category labels of 10 animals) before estimation training. For these 20 children who received the animal pretest, pre-testing did nothing to block transfer of estimation training: their numeric categorizations were quite linear after estimation training ($R^2 = .73$), almost identical to those of the unpretested treatment group in Figure 4.5 ($R^2 = .76$), and more linear than both children pretested for numeric categorization and given the treatment ($R^2 = .54$) and children who received no estimation training ($R^2 = .47$). Thus, my follow up study indicated that the numeric content of the categorization pretest—and not its length—blocked transfer.

Another mechanism that might account for the categorization pretest interfering with transfer is children avoiding inconsistent answers. The strongest version of this hypothesis would maintain that children remembered their
categorization pretest answers and repeated them on categorization posttest to appear consistent; a weaker (though perhaps less plausible) version of the hypothesis would be that children remembered which representation they used on categorization pretest and continued to use that representation to appear consistent. Although I cannot directly rule out either hypothesis (because I did not include a condition in which children answered questions secretly, which would allow them to change their answers while saving face), at least their literal answers did differ between categorization pretest and posttest. For example, among the children in the two groups that received a categorization pretest, most changed their category judgments to some degree between pretest and posttest (e.g., on categorization posttest, children tended to provide judgments that were generally smaller than on pretest), though almost all continued to provide category judgments consistent with a logarithmic representation. Thus, unless children had some remarkable insight into the distribution of estimates they were providing (which I myself could not discern without performing statistics), it seems unlikely that the categorization pretest interferes with transfer because children avoid giving inconsistent answers.

In my view, categorization pretests interfered with transfer for an altogether different reason: practice on the number categorization pretest strengthened the association between the logarithmic representation and the categorization pretest, and thereby blocked transfer of the more optimal (but under-practiced) linear representation learned during training on the estimation task. This explanation is consistent not only with the results presented and my follow-up condition, but is also consistent with previous findings of multiple representations of numeric magnitudes (Siegler & Opfer, 2003; Holyoak, 1978),
with interference of the linear representation in adults’ performance on a fractional magnitude task (Opfer & DeVries, in press), with findings of set effects in analogy and transfer (Duncker, 1945; Gick & Holyoak, 1980; Luchins, 1942, Luchins & Luchins, 1950), and with the formation of memory traces through practice (Roediger & Payne, 1982; Shrager & Siegler, 1998). To my knowledge, however, the implications of this research for the possibly harmful effects of repeated testing on transfer has not been examined previously, despite its importance in interpreting the narrow transfer of learning sometimes observed in microgenetic studies.

If correct, my explanation for children’s trouble with transfer in this experiment suggests a quite widespread obstacle to children’s transfer of learning. Specifically, my depiction of pretests being associated with the representations used to solve them is clearly analogous to Duncker’s (1945) observation of functional fixedness. In both cases, previous practice leads to something (in my case, a numerical representation; in Duncker’s, memory for functions) becoming more resistant to change with increasing use. A broader conceptual fixedness is also evident in the mental set (Einstellung) phenomena studied by Luchins and Luchins (1950), who used many different strategies (all unsuccessful) to induce students to abandon the use of an equation that yielded correct solutions for one problem type but that was less than ideal for other types of problems. And a similar conceptual fixedness also seems evident in the “change resistance” phenomena described by McNeil and Alibali (2004, 2005), in which children’s concepts of what follows the equals sign (“= __”) becomes strongly associated with the instructions to complete arithmetic operations, leading to their difficulty in learning to solve problems (e.g., $7 + 4 + 5 = 7 + __$)
that deviate from the practiced form (e.g., $7 + 4 + 5 = \_\_\_\_$). In all of these cases, performance on transfer tasks that had been well-practiced were difficult to improve by training on different but related tasks. On transfer tasks that are not well-practiced, however, surprisingly robust transfer of learning can be obtained. For example, throwing darts at an underwater target is probably not something that students perform everyday, yet in Hendrickson & Schroeder’s (1941) classic study of transfer, a brief classroom lesson on light refraction was sufficient to improve students’ aim. Thus, this handful of examples suggests that the variable of previous practice can explain some of the differences in transfer that researchers have already observed.

4.4.2 Implications for Microgenetic Studies

The finding that pretesting inhibited transfer in my microgenetic study has a number of implications for future microgenetic studies. First, the primary implication is not to abandon pretests or trial-to-trial assessments. The rationale for including pretests—allowing each learner to serve as his or her own control—remains valid and serves the critical function of also allowing researchers to characterize the rate and path of change. Without trial-to-trial assessments, it would have been impossible for me to test the representational change hypothesis about estimation, which predicted rapid and abrupt changes in the linearity of estimates. Moreover, these features of the change process were not affected by categorization pretesting in my study, nor would the practice interference hypothesis predict them to be given that children received feedback (treatment) on that task.

Rather, I take the primary implication to be that research on the breadth of change would be greatly enhanced by the use of the Solomon and Lessac (1968)
four-group design. Pretesting on the categorization task strongly interacted with treatment in this study, with the result that transfer appeared almost perfect when no categorization pretest was given and almost non-existent when a categorization pretest was given. Thus, previous reports that transfer is difficult to achieve may depend more on the administration of pretest than other variables, such as the particular type of knowledge being transferred (e.g., whether transfer is more difficult for novel categories than problem-solving strategies).

In summary, this study suggests that children may be much more likely to transfer their learning than previously supposed. A number of researchers have made this argument on both theoretical and empirical grounds (e.g., Anderson et al., 1996). In addition to these arguments, I believe I have identified an important mechanism for the trouble that children do have with transfer (viz., conceptual fixedness), and I have provided a demonstration of a potentially systematic methodological bias that has contributed to the impression that trouble with transfer is a necessary feature of learning. Assessing the magnitude of this bias will presumably require many more studies that control for the effect of pretesting.
CHAPTER 5
GENERAL DISCUSSION

Whether transferring knowledge from one classroom to another, from early in the school year to later in the school year, or from one example to other similar examples, conceptual representations allow learners to generalize over situations that differ merely in place, time, and superficial details (Murphy, 2002). This is extremely important since it is rarely the case that a person encounters the exact same problem, in the exact same way, on multiple occasions (Bassok & Holyoak, 1993). Having the ability to transfer solutions across problems that differ merely in place, time, and superficial details is an example of transfer of learning that may occur automatically—that is, without learners consciously monitoring the breadth of their generalizations.

In this dissertation, I proposed the representational alignment hypothesis of transfer that posits transfer is merely generalization. The representational alignment hypothesis focuses on two types of generalization: 1) narrow generalization that occurs when a participant maps identical elements from the training context to the transfer context, and 2) broad generalization that occurs when a participant draws commonalities between the abstract, underlying structural similarities between the training and transfer context. Across three sets of experiments, I tested the merit of the representational alignment hypothesis.
5.1 Summary of Findings

5.1.1 Broad Generalization of Linear Representations

In Experiment 1a, I traced the development of participants’ linear representation across the lifespan and found that the characteristic logarithmic-linear switch is an all-or-none switch within one order of magnitude, but it is not an all-or-none switch across all orders of magnitude. Rather, participants gradually switch from use of a logarithmic to linear representation on one order of magnitude at a time. In Experiment 1b, I provided evidence for robust transfer of learning by simply highlighting the similar superficial perceptual details (e.g., number of zeros) across increasingly larger numeric contexts. By aligning the low numeric scales with the high numeric scales, second grade students learned to "scale up" their linear representation to large, previously unfamiliar numeric contexts. For instance, when 0-100 training problems were progressively aligned with 0-1,000, 0-10,000, and 0-100,000 generalization problems, participants used analogical reasoning to draw similarities between the placement of 15 on a 0-100 number line with 150 on a 0-1,000, 1,500 on a 0-10,000, and 15,000 on a 0-100,000 number line without receiving explicit instruction to do so. In this experiment, progressive alignment served as a successful mechanism of representational change and led to fewer errors and more adult-like, linear estimates far outside the original 0-100 training context.

5.1.2 Non-Strategic Generalization

Use of a learned, more adult-like, linear representation of number is advantageous in many scholastic situations, but if children automatically deploy a linear representation of number, regardless of numeric context, this might lead to costs as well as benefits. In Experiment 2a and 2b, participants estimated
salaries denoted in fractional notation (e.g., $1/60 minutes) on a “money line” ranging from $1/1 minute to $1/1440 minutes. Younger participants, who ignored the common numerators, made more “accurate” placement of salaries because they had a logarithmic representation of the fractions’ denominators (e.g., 60 is closer to 1,440 than to 1), whereas older participants, who similarly only paid attention to denominators and possessed a linear representation of numbers (e.g., 60 is closer to 1 than to 1,440), made less “accurate” placements on the money line since k/60 is actually closer in magnitude to k/1440 than to k/1.

When participants received training on a number line task that highlighted the maximally discrepant point between a logarithmic and linear function (e.g., 150 on a 0-1,000 number line), these participants transferred their learned linear representation automatically to the fractional magnitude context without being told to do so. Increases in accuracy on the number line task were correlated with decreases in accuracy on the fraction line task. Similarly, with increasing age and experience, children deployed an automatized linear representation of number which was equated with benefits in accuracy on the number line task with whole numbers, where this was the most appropriate representation to employ. However, costs ensued on the fractional estimation task, where the linear representation was an inappropriate representation to employ.

5.1.3 Why Transfer is Rare in the Literature

Despite the abovementioned evidence that transfer of numeric representations are deployed in rather broad and automatic ways, in some learning situations, transfer of learning can be rather difficult to illicit. What then is this trouble with transfer? In Experiment 3, I offered a possible explanation:
proactive interference from previous practice on an unsupervised pretest. In the experiment, participants in the pretested treatment group first received a number categorization pretest (Is a number like 150 a really small, small, medium, big, or really big number?). Then they received training on a number line estimation task where they learned a linear representation of number. Finally, they were presented a number categorization posttest equivalent to the pretest. These participants did not transfer their learned linear representation of number from the number line training context to the number categorization posttest. I proposed that this was because the participants had previously made logarithmic category judgments on the pretest without receiving corrective feedback. In contrast, participants in the unpretested treatment group, who received no number categorization pretest, were able to successfully transfer their learned linear representation of number to the categorization posttest. These results imply that previous experiments on transfer may have actually underestimated the amount of transfer that occurred due to pretesting activating inappropriate representations.

5.2 Support for Representational Alignment Hypothesis

Across a series of three experiments that investigated transfer in the domain of numerical estimation, preliminary support was offered in favor of the representational alignment hypothesis. In this section, I summarize the representational alignment hypothesis criteria and detail the evidence in support of the hypothesis.

First, the main criteria for the representational alignment hypothesis posits that the representations that participants use are automatically and unconsciously deployed without participants monitoring the breadth of their
generalizations across training and transfer contexts. Empirical support for this hypothesis was found across all three experiments. Without hints from the researcher, children were able to transfer linear representations learned in a number line estimation training context to a transfer context containing larger orders of magnitude (Experiment 1), a fractional magnitude estimation transfer task (Experiment 2), and a category judgment transfer task (Experiment 3). Further, the participants clearly were not consciously deploying their practiced numerical representation because, in the fractional magnitude estimation task used in Experiment 2, use of a practiced linear representation actually led to costs in estimation accuracy. It seems unclear why a participant would consciously sabotage his or her performance on an estimation task by using a numerical representation that was clearly disadvantageous for the task at hand.

Second, the representational alignment hypothesis claims that representational change could occur through use of an analogy that aligned the underlying structural similarity of the training and transfer problems. Experiment 1b is the first experiment of its kind to directly test and find support for the hypothesis that analogy facilitates representational change in the domain of numerical estimation. By aligning low numerical scales with high numerical scales, participants were able to ignore the concrete differences at the surface level (e.g., number of zeros that differed between the 100, 1,000, 10,000, and 100,000 contexts) and focus on the underlying structural similarity that united the problems (e.g., decimal system).

Third, the representational alignment hypothesis states that highlighting irrelevant surface features or relations might actually detract from participants’ attention to relevant surface features or underlying structural relations.
Experiment 1b provided evidence in favor of this point. Consider the full alignment condition and the focused alignment conditions. In the full alignment condition, participants were able to look back on all of their previous answers because all training and generalization problems were presented on one page. It might have been the case that participants did not focus on how the units aligned with the magnitudes (as they were forced to in the focused alignment condition), but might have paid more attention to irrelevant surface features, like the way the units in the cherry training problems differed from the units in the carrot training problems.

Fourth, the representational alignment hypothesis posited that surface similarities between the base and target domains or a participants’ consistent internal state (e.g., activation of mental representation) might facilitate transfer between the training and transfer domains. The number line estimation task, the fractional magnitude estimation task, and the category judgment task contained 1) similar surface features (e.g., a line flanked by a beginning point/ending point of the numerical context, the buckets in the category judgment task were arranged in a left-right pattern, the to-be-estimated numeric stimuli), and 2) overlapping structural relations (e.g., decimal system) that facilitated transfer.

Experiment 3 tested claims about the mental activation of the entrenched logarithmic representation. Specifically, the pretested treatment group practiced their logarithmic representation without feedback on the category judgment pretest. This unsupervised practice made the logarithmic representation more strongly associated with the category judgment task, so that when these participants encountered the category judgment posttest task within the same experimental setting, they automatically deployed their logarithmic
representation despite training that should have prompted them to abandon usage of this inappropriate representation. These findings are consistent with existing evidence in the literature pertaining to mental sets and the availability heuristic as well as Chen & Klahr’s (2008) transfer taxonomy that indicated that transfer should be facilitated when there is a very minimal lapse in time between the training and transfer contexts.

Fifth, the representational alignment hypothesis suggested that transfer might be blocked in a situation where participants were forced to practice an inappropriate representation on a task without receiving feedback from the researcher, despite the fact that the participant had just learned a more appropriate representation on a related task. In Experiment 3, children completed a category judgment task without feedback; therefore, these second grade students produced a logarithmic series of estimates. Then the children were taught a linear representation on a number line estimation task during training. Instead of transferring their newly-learned linear representation to the categorization posttest, the participants again produced a logarithmic series of estimates on this task which was equivalent to the category judgment task administered at pretest. It was as if the participants approached the task in the exact way they had originally approached the task, and they could not inhibit the logarithmic representation that was originally associated with this task.

In total, I believe the representational alignment hypothesis’ criteria were well supported by the present experiments conducted in the domain of number. Future research will further test the claims of the representational alignment hypothesis and focus on some unanswered questions.
5.3 Broader Implications of Representational Alignment Hypothesis

More broadly, what theoretical and educational contributions does my representational alignment hypothesis make to the transfer literature? First, because I consider transfer to be the same phenomenon as generalization, findings from the literature on the generalization of learning can inform the understanding of transfer of learning.

Second, broad and narrow transfer can be accounted for by my representational alignment hypothesis. As mentioned previously, the representational alignment hypothesis claims that there are three dimensions that may impact a participant’s ability to transfer: 1) the concrete surface similarity between the training and transfer context, 2) the underlying relational structure common between the training and transfer context, and 3) the degree of activation of a common mental representation across the training and transfer context. When there are many surface similarities between the base and target domains or the underlying structural similarity between the two domains is highlighted, participants will be prompted to mentally represent the two tasks in similar ways, and broad transfer will result. When surface similarities are at a minimum, and there is no underlying surface similarity between the training and transfer context, participants will not be prompted to necessarily represent the two domains similarly, and transfer will be much more narrow.

Therefore, one situation that shares no surface commonalities across the training and transfer context as well as no underlying relational similarities, and another circumstance where there are no surface similarities, but a rich underlying structural similarity that is highlighted, may both lead to broad transfer. The difference is that transfer is much more likely to be facilitated in the
latter than the former example. These results pose an interesting implication for the transfer taxonomies (Ceci & Barnett, 2002; Chen & Klahr, 2008) because they fail to differentiate on the basis of whether or not structural similarity is present between the training and transfer domain, and thus, the taxonomies are much less predictive of the actual success rate of transfer that might otherwise be expected in a particular experiment.

Finally, my findings pertaining to the representational alignment hypothesis are very revealing for educational advancements as well. My results suggest that teachers should be advised to highlight the underlying structural similarity of related tasks if they expect their students to spontaneously transfer knowledge learned in one task to another instead of simply assuming that the students will be capable of mapping the similarities between the tasks themselves without such prompting. More specific to education in the domain of number, teachers should be aware that young students bring their logarithmic representation to bear on the numerical tasks they are presented in the classroom, and that this representation will pervade children’s thinking for some years. With increasing age, logarithmic discrepant information (e.g., feedback on 150), or alignment of lower scales (0-100) with higher ones (0-1,000 to 0-100,000), children can be prompted to think about numbers in a much more adult-like manner years before regular classroom instruction might lead them to do so.

5.4 Future Directions

My research on the transfer of numerical representations has indicated that under some circumstances, children can adopt a linear representation of number and abandon a logarithmic one after only one trial of corrective feedback or in situations that highlight the relational structure of lower and
higher numeric scales. Since children’s adoption of the linear representation is so abrupt and freely generalizable, it seems that the methodology used in my experiments might be a rather simple and effective classroom intervention to undertake.

Specifically, I would also like to focus on classroom interventions for special populations, such as first and second grade students possessing mathematical learning disabilities. Recent research by Geary and colleagues (2007; in press) has indicated that children with mathematical learning disabilities are less accurate in their number line estimates than are age-matched, typically-developing children. One reason for their lower performance may be that the children with the mathematical disabilities rely to a greater extent than typically developing students on their logarithmic as opposed to linear representation of number. Siegler and his colleagues (Ramani & Siegler, 2008; Siegler & Ramani, in press) have found that exposing low-SES preschoolers to an hour of numeric board games increased the children’s proficiency in counting, ability to identify printed numerals, ability to compare the relative sizes of numbers, and performance when estimating the positions of numbers on number lines. I would like to extend this methodology to children in the first and second grade who possess mathematical disabilities. For instance, half of participants in my proposed study would play Chutes and Ladders, a board game comprised of spaces numbered 1-100, whereas the control group would be asked to count from 1-100 and complete addition and subtraction problems for an equivalent amount of time. Participants in both the experimental and control groups would complete a pretest consisting of a 0-100 number line estimation task, numeral identification, counting, magnitude comparison, numerical category judgments,
and two-digit addition problems. The hypothesis is that children playing the numeric board game would see greater improvements in the battery of mathematics tasks mentioned above from pretest to posttest.

Further, I have indicated an experimental artifact—pretesting—that leads to the “blocking” of transfer. In future research endeavors, I would like to investigate ways to “unblock” transfer. For instance, might 1) changing the surface characteristics of the transfer pretest and posttest (e.g., shape, size, or spatial configuration of the “really small”-“really large” labeled boxes), 2) asking children to take the transfer pretest and posttest in separate rooms (cf. state-dependent memory; Godden & Baddeley, 1975), 3) having children complete the pretest and posttest under the direction of different experimenters, or 4) having children listen to different mood-inducing songs (cf. state-dependent memory/mood congruency effects; Bower, 1981), lead to successful transfer of the linear representation? Also, if the participants received a categorization pretest in a numerical context in which they already possessed an appropriate linear representation (0-100), and then received training on the number line estimation task in the 0-1,000 context, would they be able to transfer the linear representation to a 0-1,000 number categorization posttest since they had not previously practiced an inappropriate representation without corrective feedback?
LIST OF REFERENCES


Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology, 95*, 393-408.


APPENDIX A

SEX DIFFERENCES IN MAGNITUDE ESTIMATION
Sex Differences in Magnitude Estimation

Sex differences are an intriguing test of the representational alignment hypothesis. For instance, it might be the case that boys produce a more linear series of estimates on numerical estimation tasks than do girls, and this may be an explanation for the widespread differences in boys’ and girls’ mathematics performance (see Halpern et al., 2007, for a recent review). If this is indeed the case, this leads to an interesting prediction of the representational alignment hypothesis. Previous research (Opfer & DeVries, in press) found that when adults used their automatized, linear representation in a fractional magnitude estimation task, their accuracy suffered due to the inappropriateness of this representation for the task at hand as compared to children who utilized their automatized logarithmic representation for the task (see Experiment 2a and 2b for a similar finding regarding younger vs. older children). Extending these findings to sex differences would lead to the hypothesis that girls would actually out-perform boys in the context of a fractional units task since they are likely to produce a series of estimates more consistent with a logarithmic than a linear representation. On the contrary, boys should produce more accurate number line estimates than girls because their automatized linear representation is more appropriate for this task than is the logarithmic representation. Given the centrality of numerical representations to mathematical thinking and given widespread public speculation about sex differences in quantitative abilities, I examined whether boys and girls differed in their numerical representations that lead to estimation performance.

What sex differences have already been reported for mathematical reasoning? Typically, differences between boys’ and girls’ quantitative
performance varies greatly depending on the variable measured and children’s ages (Halpern et al., 2007; Hyde, Fennema, & Lamon, 1990). For instance, females tend to outperform males during early elementary school years on quantitative tasks related to verbal abilities and curriculum content, whereas males tend to outperform females from grades 4 through 12 when quantitative tasks involve visuo-spatial concepts, reasoning about real-world problems, and estimating answers to arithmetic problems like 76 x 89 (Dowker, Flood, Griffiths, Harriss, & Hook, 1996; Doolittle & Cleary, 1987; Geary, 1996; Hyde, Fennema, & Lamon, 1990; Levine, Huttenlocher, Taylor, & Langrock, 1999; Willingham & Cole, 1997). Given the correlation between these latter measures and numerical estimation performance that have also been found to occur during the early elementary school years (Booth & Siegler, 2006; Siegler & Booth, 2004), I examined whether sex differences might also exist in numerical estimation, possibly due to boys possessing more linear representations of numerical magnitude than girls.

Theoretically, sex differences in numerical representation are interesting for at least three reasons. First, a sex difference has been predicted on the basis of neurophysiology (Halpern et al., 2007). There is now considerable evidence that the brain represents numerical quantity in at least two regions over the course of development: the prefrontal and inferior parietal cortex (Ansari et al., 2005; Dehaene, Piazza, Pinel, & Cohen, 2005; Dehaene et al., 1999; Nieder & Miller, 2002; Pinel et al., 2004; Rivera et al., 2005), and sex differences in the function and architecture of these regions have been reported in studies of human and non-human animals (Goldstein et al., 2001; Kavaliers, Osenkopp, Galea, & Kolb, 1998; Knops et al., 2006). For example, in human adults, Goldstein and colleagues (2001) found that, adjusting for overall brain volume, the inferior parietal lobe
was 20% larger in males than in females. What has remained unclear is whether these sex differences in the architecture and function of the parietal cortex makes a difference in *early* numerical representations, such as that tapped by the number-line task. Interestingly, Halpern et al. (2007) noted that to the extent that these regions are larger in males than females, a male advantage is predicted for the mental number line. I will examine this prediction by comparing the accuracy and linearity of boys’ and girls’ estimates on a number line estimation task.

Second, if sex differences exist at the level of automatic numerical representations, as the representational alignment hypothesis suggests, then such sex differences should have *costs* as well as *benefits*. Specifically, sex differences in the accuracy of symbolic magnitude estimates should be complementary, with the sex that performs more accurately in the numerical magnitude context performing less accurately in the fractional magnitude context. The costs of representational change are well understood in the field of perceptual learning (e.g., Petrov & Anderson, 2005), but to my knowledge, this consideration has not figured at all in the discussion of how basic sex differences might contribute to mathematical proficiency. In the extreme, an implication of my theoretical analysis is that robust sex differences in numerical representations might not confer any *overall* advantage in accuracy but could nevertheless confer some large *task-specific* advantages to both sexes (e.g., male advantage on whole number estimation; female advantage on fractional magnitude estimation).

Finally, as Newcombe, Mathason, and Terlecki (2002) have observed, while it is scientifically interesting to document sex differences, the more interesting question is how experience affects these sex differences. For example, Opfer and Siegler (2007) have shown that even a small amount of focused
training can virtually eliminate age differences in numerical estimation. Is the same true of sex differences as well? To examine this issue, I determined whether there were any sex differences in the performance of participants that completed Experiments 2a and 2b. I examined the accuracy and linearity of males and females estimates prior to, during, and after the course of training.

**Sex Differences Results**

First, I examined whether the boys and girls from Experiment 2a differed in the accuracy of their numerical and fractional magnitude estimates before they received any corrective feedback (see Figure A.1). Boys’ *numerical* magnitude estimates were more accurate on average ($M = 88\%, SD = 10\%$) than girls’ estimates ($M = 78\%, SD = 10\%$), $F(1, 63) = 14.78$, $p < .0005$, $d = 1$. In contrast, girls’ *fractional* magnitude estimates tended to be more accurate on average ($M = 50\%, SD = 16\%$) than boys’ estimates ($M = 40\%, SD = 13\%$), $F(1, 34) = 3.66$, $p = .06$, $d = .69$. When comparing boys and girls median estimates, differences in accuracy reflected a logarithmic-to-linear shift in the $0 – 1,000$ numerical magnitude context. Boys’ numerical estimates (average lin R$^2 = .81$, $SD = .24$) were more linear than girls’ estimates (average lin R$^2 = .60$, $SD = .22$), $F(1, 63) = 13.8$, $p < .0005$, $d = .91$, and boys were less likely than girls to provide whole number estimates best fit by the logarithmic function (30% versus 65%, $\chi^2 = 7.75$, $p < .01$).
Sex differences were chiefly evident for younger children (7- to 8.49-year-olds) as compared to older children (8.5- to 9.5-year-olds). Among younger children, a greater proportion of boys generated linear estimates than did girls (44.4% versus 4.8%, Fisher’s $p = .02$), whereas among older children, boys and girls generated linear estimates at similar rates (83.3% versus 75%, ns) on the number line estimation task. Whether differences between boys and girls in this sample indicated a developmental delay for girls’ numerical and fractional magnitude estimation or a stable sex difference is an issue that I explored further in by analyzing sex differences from participants in Experiment 2b.

Effect of Learning on Sex Differences

I was also interested in whether sex differences in numerical and fractional magnitude estimation persisted after children received feedback in the
context of Experiment 2b. I was particularly interested in whether conditions that substantially reduced age-differences in estimation (as in Experiment 2a) would also reduce sex-differences. Specifically, I wanted to know whether boys continued to generate more accurate numerical magnitude estimates than girls and whether girls continued to generate more accurate fractional magnitude estimates than boys. To find out, I examined three measures of estimation performance for each task: accuracy (0-100%), fit of the function ($R^2$, .00 – 1.00) associated with greater accuracy (i.e., linear for numerical estimation, logarithmic for fractional magnitude estimation), and percent of children best fit by the function providing better accuracy (see Table A.1).

| Sex | Number Line | | Fraction Line | |
| --- | --- | --- | --- | |
| | Accuracy | Fit of Linear Model | % Best Fit by Lin | Accuracy | Fit of Log Model | % Best Fit by Log |
| Boys | M = 90% | Boys: Lin $R^2 = .99$ | Boys: 81% Lin Best | Boys: M = 39% | Boys: Log $R^2 = .78$ | Boys: 35% Log Best |
| Girls | M = 83% | Girls: Lin $R^2 = .97$ | Girls: 59% Lin Best | Girls: M = 46% | Girls: Log $R^2 = .91$ | Girls: 59% Log Best |
| $F_{(1, 62)} = 9.34, p = .003$ | $\chi(1) = 3.04, p = .08$ |

Table A.1 Analyses of sex differences at posttest.

On each of the six dependent variables, boys and girls continued to differ in their estimation performance, and I observed no interaction between sex (0 = boys, 1 = girls) and condition (0 = control, 1 = treatment) on posttest scores ($F$’s < 1.25, ns). As at pretest in Experiment 2a, boys’ accuracy on the number line task ($M = 90\%, SD = 8\%$) continued to be greater than girls’ accuracy ($M = 83\%, SD = 10\%$), $F (1, 62) = 9.34, p = .003, d = .77$, whereas girls’ accuracy on the fraction line task ($M = 46\%, SD = 16\%$) continued to be greater than boys’ accuracy ($M = 38\%$, $p = .003, d = .77$).
$SD = 12\%$, $F(1, 62) = 5.69, p = .003, d = .57$. On the number line task (which was favored by a linear representation), the fit of the linear model to boys’ median estimates ($R^2 = .99$) was slightly greater than the fit of linear model to girls’ median estimates ($R^2 = .97$). On the fraction line task (which was favored by a logarithmic representation), the fit of the logarithmic model to girls’ median estimates ($R^2 = .91$) was greater than the fit of the logarithmic model to boys’ median estimates ($R^2 = .78$). Finally, on the number-line task, a higher proportion of boys generated estimates best fit by the linear model (81%) than did girls (59%), whereas on the fraction-line task, a higher proportion of girls generated estimates best fit by the logarithmic function (59%) than did boys (33%). Thus, sex differences in estimation performance appeared to be stable over short-term training, whereas age differences in estimation performance were not stable over short-term training (see Experiment 2b).

Effect of Context on Sex Differences

Although the number line task has been used extensively in prior research on numerical representations (e.g., Opfer & DeVries, in press; Opfer & Siegler, 2007; Laski & Siegler, 2007; Booth & Siegler, 2006; Siegler & Booth, 2004; Siegler & Opfer, 2003), sex differences on the task have not been reported previously, leading some researchers (e.g., Spelke, 2005) to infer quite reasonably that sex differences on the number line estimation task do not exist.

It is important to note that I was unable to analyze sex differences in Experiment 1a and 1b. In Experiment 1a, there were only eight participants in each experimental condition (e.g., 0-1,000, 0-10,000, and 0-100,000 number lines) across the three age groups that could not allow for a reliable test of sex differences. Likewise, in Experiment 1b, participants received training and were
assigned to different experimental conditions that dictated the amount of progressive alignment the participants would receive after only one problem. Since these participants did not receive a pretest, I was unable to assess boys’ versus girls’ linearity prior to receiving corrective feedback.

On the contrary, I was able to analyze for potential sex differences in Experiment 3. When I examined numerical estimation in the first and second graders (mean age = 7.85 years) from Experiment 3, I found that, prior to training, the 27 boys tended to generate more linear estimates on average than did the 29 girls (boys, $M_{lin} R^2 = .69$; girls, $M_{lin} R^2 = .58$), $F(1, 55) = 3.13, p = .08$. After training, boys also generated more linear estimates on average than did girls (boys, $M_{lin} R^2 = .77$; girls, $M_{lin} R^2 = .58$), $F(1, 55) = 6.19, p = .02$.

Do sex differences on number line tasks generalize to other tests of children’s numerical representations? One reasonable concern is that my magnitude estimation tasks might not be representative of more widespread sex differences in numerical representations because the number-line task imposes a spatial performance demand (i.e., marking a position on a line), and other tests of spatial visualization (e.g., Levine, Huttenlocher, Taylor, & Langrock, 1999) have found sex differences in this ability, despite the inherently non-numerical nature of the spatial visualization tasks. To check this idea, I re-examined sex differences on a test of numerical concepts—the numerical categorization task from Experiment 3—that simply imposed a verbal performance demand. Recall that on this task, children were told that 0 is a “really small number,” that 1,000 is a “really big number,” and then were asked to say whether a novel number (e.g., 150) was “really small,” “small,” “medium,” “big,” or “really big.” With each category assigned an arbitrary ordinal number (i.e., 0 for “really small,” 4 for
“really big”), one can again regress the number given against the categorization judgment to assess the linearity of boys’ and girls’ verbal category judgments. Although the number of children who participated in this task in Experiment 3 was fairly small (14 boys, 8 girls) and the sex differences were not always statistically significant, absolute performance and magnitude of sex differences was nevertheless quite similar on the verbal number-categorization task (boys, mean lin R2 = .62; girls, mean lin R2 = .54) and on the spatial number-line task (boys, mean lin R2 = .69; girls, mean lin R2 = .58). I think this similarity suggests that early sex differences in numerical representation do not simply reflect a performance demand imposed by the number line, although clearly more tests are needed to test the claim (see Future Directions section).

Even if early sex differences in numerical representation do not reflect a performance demand, sex differences in accuracy clearly reflect a task demand. As Simon (1996) noted, accuracy is simply a measure of the fit between the approach that subjects use on a task and what the task demands for accurate performance. In the case of my number line estimation task, a logarithmic representation was less adequate for accuracy than a linear representation, and thus I observed boys to outperform girls on the number-line task. In the case of my fraction line task, the reverse was true, and thus I observed girls to outperform boys on the fraction-line task. Combining the two tasks led to there being no overall advantage for boys or girls (boys’ accuracy, M = 64%; girls’ accuracy, M = 65%; ns). Thus, the sex difference in accuracy was not absolute, but depended on the numeric context.
Broader Implications of Context Effects

My findings suggest that there are real age and sex differences in the representations used to solve magnitude estimation problems, but that age and sex differences in accuracy simply reflects the match of a particular numerical context to those representations. This point is an important one: if a difference between boys and girls (or younger and older children) is at the level of the representation, then the advantage of either sex (or age) reflects a kind of ‘dumb accuracy.’ That is, representations that are broadly generalizable across contexts (leading to positive transfer of accuracy), may also be inflexible to changing task demands (thereby leading to negative transfer of accuracy).

This conclusion is also important in that it provides guidance for making broader predictions about age and sex differences in mathematical cognition. Specifically, the analysis suggests that there is no necessary positive relation in accuracy among various mathematical tasks. Rather, positive relations in accuracy among specific mathematical tasks depend on the relative similarity of the abstract, underlying representations used to solve the specific tasks and the ability of those representations to accurately encode relevant mathematical properties. For example, in fractions expressed in decimal format or with common denominators, a logarithmic representation would be inappropriate because it fails to encode both the linear nature of the decimal system and the linear relation between the value of a fraction and its numerator. In contrast, for fractions expressed in simpler familiar units, such as expressing salaries in terms of “$1 per hour” versus “$1 per day,” the linear representation would be inappropriate due to the power relations between the implicit denominator in “hours” and “days” (60 minutes and 1,440 minutes, respectively) and the value
of the salary. A testable prediction of the current analysis is that girls and younger children should outperform boys and older children when estimating the magnitude of salaries expressed in familiar units, whereas boys and older children should outperform girls and younger children when estimating the magnitude of salaries expressed in decimals and in fractions with common denominators.