

Spring 2009

Keeping visual-auditory associations in mind: The impact of detail and meaningfulness on crossmodal working memory load

Anne T. Gilman

University of New Hampshire, Durham

Follow this and additional works at: <https://scholars.unh.edu/dissertation>

Recommended Citation

Gilman, Anne T., "Keeping visual-auditory associations in mind: The impact of detail and meaningfulness on crossmodal working memory load" (2009). *Doctoral Dissertations*. 473.

<https://scholars.unh.edu/dissertation/473>

This Dissertation is brought to you for free and open access by the Student Scholarship at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact nicole.hentz@unh.edu.

**KEEPING VISUAL-AUDITORY ASSOCIATIONS IN
MIND: THE IMPACT OF DETAIL AND
MEANINGFULNESS ON CROSSMODAL WORKING
MEMORY LOAD**

BY

Anne T. Gilman

B.A., Vassar College, 1992

M.A., University of Texas at Austin, 1997

M.A., University of New Hampshire, 2006

DISSERTATION

Submitted to the University of New Hampshire
in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

in

Psychology

May 2009

UMI Number: 3363717

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI[®]

UMI Microform 3363717
Copyright 2009 by ProQuest LLC
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

This dissertation has been examined and approved.

Dissertation Director, John Limber,
Associate Professor of Psychology

Colin Ware,
Professor of Computer Science and of Ocean Engineering

Michelle Leichtman,
Associate Professor of Psychology

Andrew Leber,
Assistant Professor of Psychology

John Sparrow,
Associate Professor of Psychology, UNH Manchester

Date

DEDICATION

I would like to dedicate this work to my family and friends who have supported me every step of the way. Inquisitive minds like Fred, Peggy, Marion, Carl, and Carole did much to light my way, as did the passion and willingness to get to the heart of the matter of Ed (Papa) and Diane.

When I went about seeking support for this graduate-school journey, I wrote about my desire to use my research into thinking and learning to serve kids who want to learn, who think differently, and who do not have access to my mom's tireless advocacy and guidance. For all that my research topic has narrowed and grown hard-to-understand wings of jargon, that desire remains.

As I've grappled with that same jargon and with the excitement of first-hand research, having expert translation help from my dad—still tops when it comes to following me while my ideas are still forming—has time and again brought me closer to my goal. And again, in honor of any learners, young or old, that I may help with my ability to see more than one organizing structure in the same pattern of data, I thank him.

My sister has a pastel drawing of mine displayed in her house. There have been days when I've wondered if learning another field made even less sense, in view of the outer world, than my adolescent spate of drawing vaguely elfin faces. The currency has not yet been invented which could measure the worth of my sister's support in these and all my passions.

To add a sillier note, I'm not sure if my grandmother Billie had any idea, as she worked so hard to make dreams of greater educational opportunities come true for her family, that one of her grandkids would spend quite *this* much time engaged in higher learning. Happily for both of us, she's always open for good surprises. I've worn my new jumpers made for the new instructor with pride, I look forward to wearing some of her snazzier new creations

later this week, and I'm wearing a blouse she made as I type. Her support in my work and her reminders to stop and have some fun have made a huge difference.

Since so many of my aunts and uncles and cousins are environmentalists, including the honorary ones, I'm hoping they will understand my curtailing my family thanks at this point, and further understand how much I appreciate their words of good cheer. Finding a good psychology program in New England has been a particular boon as far as seeing Nancy, Jeff, and my amazing nieces Margaret and Laura. My delightful step-niece Elinor has brightened my family visits, and her parents' insights into higher learning have guided my steps.

My stalwart Schenectady friends—Rob, Sam, Aaron, Janis, Alexandra, Steve, Vrinda, Jin, Bowden—have been there for me in those scary moments when my research engine was coughing and stalling, and also there to remind me to celebrate! This work is dedicated to them, too—their encouragement has gotten me all the way to the finish line. The same applies to those dear hearts in Albany—Dorothy, Susan, Nicole, Anne—, in Montreal, and beyond who have kept me grounded.

Particularly for the dissertation push, the kind thoughts and emails of friends and colleagues like Gavin, Muthu, and Jocelyn have renewed my strength. My former colleagues at GE's Service Algorithms Lab and project mates in other labs gave me the best possible start in my new career direction. May this dissertation and the work I do in the future bring out all the more the potential all these dear connections have seen in me.

ACKNOWLEDGEMENTS

Through the grace of my Creator, the financial support of the University of New Hampshire and the Graduate School, and the help of so many others, I have had the chance to investigate research questions more thoroughly than was ever previously possible for me. Not all of this intellectual growth has been comfortable, and I credit my advisor, John Limber, for bringing his curiosity, his humane approach to education, and his general unflappability to bear when needed. Working with him has been one of the best parts of my coming to UNH.

So, too, has my collaboration with Colin Ware been a highlight of my graduate program; I cannot thank him enough. Feedback from other Visualization Lab scientists and my inclusion in the CCOM community—from the high-mental-octane lunchroom to the beautiful seacoast picnics—have also enriched the last two years. A few highlights include Kurt and Matt’s expert presentation feedback, Lynn and Nick’s “Seymour DuMoore” tours of mountains and waterways, and Val and Brian’s help with the L^AT_EX template for this document. I greatly value Larry Mayer’s welcome and the moral and logistical support from Maureen Claussen, Linda Prescott, and Abby Pagan-Allis.

My other committee members have been so helpful: Michelle Leichtman, always applying her keen analysis and professional standards with a kindness and wit I much admire; Andrew Leber, providing in-depth and up-to-date perspectives from even before he was fully hired; and John Sparrow, deftly pointing out ways to sharpen my experimental design with unflappability to rival John Limber’s! All the faculty and students keeping the “BBC” flame bright have helped me grow as a scientist, and Erika, Jan, Kie, Başak, Dan, Chris, Tyler, and others have taught me and kept me laughing. Ken Fuld and David Pillemer’s thoughtful guidance on my master’s research improved this later work. Bill Stine’s willingness to ask—and answer—questions both serious and silly and Brett Gibson’s support

and encouragement both made a huge difference. Space does not permit me to thank every member of the department, although each has taught me in some way; Anne, Robin, Donna, and above all Janice deserve mention for the many times their skills and/or adept use of humor got me through stressful moments in teaching and research. And if Stella Arambel's guidance and inspiration are at all reflected in this work, I have succeeded.

For teaching me new ways to approach the stresses and satisfactions of graduate-student life, Patricia Hanley deserves many thanks. Coaches Brian and Kathy and brown belts Tanya and Kurt taught me mental and physical agility in my brief but so-beneficial Aikido training. The Lindens, the Heilbronnens, the Hermans, the amazing Schwartz family, Al Spaien, and Rabbi David Mark made Temple Israel Portsmouth my Seacoast home.

Several tools have literally and figuratively transformed my work. Shane Mueller's Psychology Experiment Building Language (PEBL) rescued me from an unstable version of PsyScope and freed me to code and run experiments on every scrap of hardware I finagled access to, including the less-popular machines in the McConnell cluster. Athanassios Protopapas, source of the `apa.cls` format, merits much thanks, as does Stephen for setting me on the `LATEX` path. The R Core Development Team keeps me equipped with a Mac-friendly and enormously powerful statistical tool, at no cost to me. I want to thank Antonio Possolo for introducing me to R and for believing in my potential as a scientist before I knew how.

These acknowledgements would not be complete without further mention of the leadership of the graduate school—the deans and their assistants have done so much for our educational development, our health, and our quality of life. Stephanie Bramlett's amazing powers of social catalysis got me to perform in an open-mic night on campus. And I'm honored that Associate Dean Cari Moorhead could find the time—amid all of those battles for the future of graduate programs—to give me thoughtful and insightful mentoring for my own doctoral path.

Responsibility for the gaps and gaffes in this work remains squarely on my shoulders, leaving undiminished the contributions of those uncountable others on whose shoulders I now stand.

TABLE OF CONTENTS

DEDICATION	iii
ACKNOWLEDGEMENTS	v
LIST OF FIGURES	xiii
ABSTRACT	xiv
1 INTRODUCTION	1
1.1 Relevance of Crossmodal Working Memory	1
1.1.1 Conflicting Predictions about Combining Modalities	1
1.1.2 Problem Statement	3
2 MEASURING WORKING MEMORY CAPACITY	4
2.1 What is the function of working memory?	4
2.1.1 How is working memory different from long-term memory?	4
2.2 Are sensory modalities important for WM?	6
2.2.1 Modality and the psychology of learning	7
2.2.2 Neuroimaging of crossmodal processing	10
2.2.3 Current approach to modality	12
2.3 Measurement: What units does WM use?	12
2.3.1 Chunks, slots, and information load	12
2.3.2 Change detection and visual and crossmodal WM	13
2.3.3 Visual and auditory details play a conflicting role	15
2.3.4 Associations—is that the same as meaning?	17
2.3.5 Prior knowledge and working memory	19

2.3.6	Which features are we counting?	20
2.4	Goals of the present work	21
3	CROSSMODAL CHANGE DETECTION: DEVELOPING A GENERAL	
	METHOD	23
3.1	Crossmodal Adaptation of Change Detection Procedure	23
3.1.1	General Procedure	23
3.1.2	Participants	26
3.1.3	Apparatus	27
3.1.4	Stimuli	27
3.2	Experiment 1: Contrasting Unimodal and Crossmodal Performance	27
3.2.1	Method and stimulus specifics	27
3.2.2	Hypotheses: Experiment 1	28
3.2.3	Results: Experiment 1	29
3.2.4	Discussion: Experiment 1	29
3.3	Experiment 2: Contrasting Simple and Complex Stimuli	30
3.3.1	Hypotheses: Experiment 2	31
3.3.2	Results: Experiment 2	31
3.3.3	Discussion: Experiment 2	32
3.4	Experiment 3: Does Test-Probe Location Matter?	32
3.4.1	Hypotheses: Experiment 3	32
3.4.2	Results: Experiment 3	33
3.4.3	Discussion: Experiment 3	34
3.5	Experiment 4: Will Words Improve Performance?	34
3.5.1	Hypotheses: Experiment 4	35
3.5.2	Results: Experiment 4	35
3.5.3	Discussion: Experiment 4	35
3.6	Experiment 5: Contrasting Simple and Complex Stimuli—Replication	35
3.6.1	Results: Experiment 5 (replication)	36

3.6.2	Discussion: Experiment 5	36
3.7	Experiment 6: Do Image Differences Matter More, Absent Location?	37
3.7.1	Hypotheses: Experiment 6	37
3.7.2	Results: Experiment 6	37
3.8	Discussion: Method Development	38
3.8.1	Overall Results	38
3.8.2	Suggested Modifications	40
4	CROSSMODAL WORKING MEMORY LOAD, DETAIL, AND MEANINGFULNESS: REFINING METHODS	41
4.1	Image selection	42
4.1.1	Overall constraints	43
4.1.2	Representational Images	43
4.1.3	Abstract Shapes	44
4.1.4	Meaningfulness Analysis	45
4.2	Experiment 8: Revised Crossmodal Change Detection Method	45
4.2.1	Adding in Visual Search	45
4.2.2	Centered-Test Crossmodal Change Detection	46
4.2.3	Stimuli and Apparatus	46
4.2.4	Procedure	47
4.2.5	Results: Comparison to method development	48
4.2.6	Results: Search speed as WM capacity predictor	52
4.2.7	Comparison to Individual Association Variety	54
5	FURTHER DISCUSSION AND CONCLUSIONS	62
5.1	Working memory impact of adding detail crossmodally	62
5.1.1	Modality itself	62
5.1.2	We Do Not Just Count Features	63
5.1.3	Location, location, location	65

5.1.4	Modality itself: Visual and auditory differences	66
5.1.5	Intra-set distinctiveness and working memory	66
5.1.6	Color, WOW!	67
5.1.7	Are individual WM capacity differences at play?	68
5.2	What does meaningfulness mean crossmodally?	69
5.3	Future Work	70
5.3.1	Conceptual distance as category delineator	70
5.3.2	Taking another long-term look at detail	71
APPENDICES		74
A IMAGES USED		76
A.1	Images: Experiments 1 through 6	76
A.1.1	Images: Experiments 1 through 5	76
A.1.2	Images: Experiment 6	77
A.2	Images: Experiments 7 and 8	78
A.2.1	Images for Training	78
A.2.2	Images for Testing	78
A.3	Procedure Specifics: Instructions	81
A.3.1	Experiments 1–6 Instructions	81
A.3.2	Experiment 8 Instructions	81
A.4	Procedure Specifics: Timelines	84
A.4.1	Visual Search	84
A.4.2	Change Detection	84
A.4.3	Overall Structure	85
B FORMULAS AND CAPACITY ESTIMATES		87
B.1	Estimating working memory capacity	87
B.1.1	Application of Pashler’s capacity estimator	87
B.2	Accuracy Measures	89

B.2.1	Corrected Accuracy Formula	89
B.2.2	Signal Detection Theory Formulae	89
C	IRB APPROVALS	90
	References	95

LIST OF FIGURES

1-1	Visual stimuli used in to determine visual working capacity for multi-featured, complex objects.	2
1-2	Visual gratings used to test recognition memory.	3
1-3	Do all mushroom images elicit identical cognitive processing?	3
3-1	Change Detection: This would count as a <i>no-change</i> trial, as the test object has the same visual-auditory pairing as the first object presented. Depending on the experimental condition, the balls might be uniform gray or each a different color.	24
3-2	Change-detection accuracy results measured using $p(\text{Hit}) - p(\text{FA})$, with residual standard error indicated.	30
3-3	Change Detection: This would count as a <i>change</i> trial, because the glass was first paired with a frog sound, not a cat's. With the test image placed in the center, observers cannot rely on location cues to detect changes in sound-image pairings.	33
4-1	Average number of total and unique associations provided by each participant per image in a given category, crossing image type (representational or abstract) with detail level (color or grayscale). Error bars indicate residual standard error.	57
4-2	Crossmodal change-detection accuracy by several measures, using 3-item arrays for comparison to earlier experiments: Corrected accuracy ($p(\text{Hit}) - p(\text{FA})$), participant z scores by condition, and signal detection accuracy along with bias. Error bars indicate residual standard error.	58

4-3	Experiment 8: Search-time estimates (in ms) for each image type and an estimate of the crossmodal working memory load for associating each image type with an animal sound. Load estimates are based on the inverse of the number of items in memory arrays for which participants show 75% raw change-detection accuracy.	59
4-4	Search-time cost per image type for each participant plotted as a function of estimated working memory load of that image type. The latter estimate was calculated as the inverse of the number of objects affording the participant 75% accuracy.	60
4-5	Average search-time cost per image type plotted as a function of estimated working memory load of that image type.	61
5-1	Relation between working memory load and search times for crossmodal stimuli tested here (see key) and prior visual-only stimuli (range and domain shown by dotted rectangle). WM load is based on the inverse of participants' capacity at 75% accuracy: a load of .25 indicates that they could detect changes with that accuracy for 4 items, a load of .5 reflects a capacity of 2 items, and so on.	73

ABSTRACT

KEEPING VISUAL-AUDITORY ASSOCIATIONS IN MIND: THE IMPACT OF DETAIL AND MEANINGFULNESS ON CROSSMODAL WORKING MEMORY LOAD

by

Anne T. Gilman
University of New Hampshire, May, 2009

Complex objects have been found to take up more visual working memory—as measured by lowered change-detection accuracy with such stimuli—than simple colored shapes (Treisman, 2006; Xu, 2002). While verbal working memory studies have similarly shown reduced apparent capacity for longer words (Baddeley, 2007), other research has demonstrated that features contributing to object categorization and recognizability can help visual working memory capacity (Olsson & Poom, 2005; Alvarez & Cavanagh, 2004). Until very recently, no measures of **crossmodal** working memory capacity had been proposed, even though crossmodal associations are part of the fabric of learning, from classical conditioning to calculus. The working memory load of a range of complex crossmodal (visual–auditory) objects was measured here in a sequence of experiments adapting classic visual change detection procedures (Vogel et al., 2001). The adapted method involves rapid sequential presentation of objects, each comprising a sound and an image, with a test object appearing after a 1-second delay. Application of this method shed light on the working memory impact of two sources of complexity, featural detail and object meaningfulness. Displaying the test object in a previously unused location—in this case, the center of the screen—resulted in lower change-detection performance compared to placement in its original location. Test location interacted with the role of different image types (gray and colored shapes, drawings, and photos). Image type showed no consistent pattern of influence on working memory capacity

when test objects appeared in their original locations; when shown in an alternate location, crossmodal associations involving more-detailed images were more accurately recalled. Independent of test location, more-complex animal sounds provided better crossmodal change detection performance than abstract tones. An association measure showed consistently higher numbers of associations for representational images than abstract ones. Observers' response bias was lower for meaningful images, but their change-detection accuracy did not differ by image meaningfulness. The results obtained with this novel crossmodal working memory measure demonstrate that perceptual detail contributes to effective crossmodal working memory capacity for sounds and for abstract and realistic images.

CHAPTER 1

INTRODUCTION

1.1 Relevance of Crossmodal Working Memory

Rare is the university lecture which has no visual material—projected slides being perhaps the most common—accompanying the presenter’s spoken words. This incorporation of sights and sounds together is quite natural: a sighted, hearing child raised with a cat learns about its meows, its sensitive ears, and its method of locomotion at roughly the same time. Psychologists have long studied this type of joining together of heterogeneous sensory input—after all, how could Pavlov’s dog have become conditioned without some internal processing able to treat both sounds and sights and/or smells as relevant to each other and to tie them together in the first place? Returning to the world of formal learning by humans, the increasing presence of computers in homes and classrooms over recent decades has provided fertile ground for the growing use of interactive multimedia learning programs. Many people now use multimedia displays in their cars to help them reach the venue for their work or their evening’s entertainment.

1.1.1 Conflicting Predictions about Combining Modalities

While combining modalities for learning is natural, its effectiveness in conveying unfamiliar information is not guaranteed. Extensive evaluations of the effect of multimedia technologies on learning has provided many practical recommendations on how to make instructional multimedia more successful (Mayer, 2005a, 2001). Influential factors for slide-like presentations of scientific material include tone, redundancy, contiguity, and many others. Puzzles remain, however, as some of the conclusions in this literature about adding detail (Mayer,

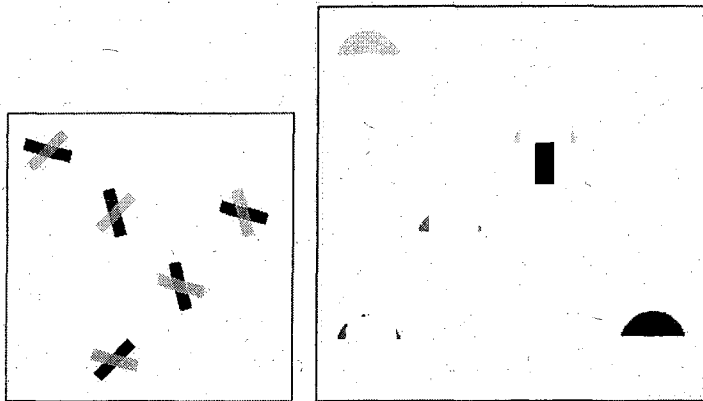


Figure 1-1: Visual stimuli used in to determine visual working capacity for multi-featured, complex objects.

Heiser, & Lonn, 2001) within and across modalities (in this case, visual and auditory) conflict with established studies of visual and verbal cognition, and some of the key authors differ on the conditions constraining when *prior knowledge*—comprising what people know and can talk about as well as learned responses to situations in the world which are harder to articulate—helps people learn new information. Several of the relevant findings have been found not to apply in a real-life schoolroom situation (Muller, Lee, & Sharma, 2008; Tabbers, Martens, & Merriënboer, 2007).

On the other hand, cutting-edge visual cognition research provides explanations of the impact of detail and recognizability on our ability to remember new information that do not always lend themselves readily to real-life application. For example, how meaningful do you find the objects in Figure 1-1, taken from Xu (2002)? Some scientists test short-term memory using even more abstract stimuli “that are little burdened by the complexities of extra-laboratory associations” (Sekuler & Kahana, 2007), such as the examples shown in Figure 1-2. Do you think the mushrooms shown in Figure 1-1 might be stored and retrieved any differently than mushroom images such as Figure 1-3? These particular puzzles rest on conflicting assumptions about the impact of adding detail and combining different modalities on *working memory (WM)* capacity. Student learning is a multifaceted process, and working memory capacity plays a key limiting role in that and any process where people are



Figure 1-2: Visual gratings used to test recognition memory.

Figure 1-3: Do all mushroom images elicit identical cognitive processing?

forming new associations which they may or may not recall later. Later recall is the domain of long-term memory, which covers durations from several minutes to many years. At the other extreme, sensory memory may last only for a few milliseconds, up to a second or two. Forming novel associations—between a meow and a texture, between a bell and a meal, or between a drawing and the typed word “synapse”—happens in the bridging processes of working memory, played out over a few seconds.

1.1.2 Problem Statement

The goal of the proposed work is to shed light on those particular conflicting predictions about the impact of added detail on working memory capacity for objects which combine visual and auditory features. This requires the development of a new measure of *crossmodal working memory* capacity, considering both perceptual detail and stimulus meaningfulness—an important outgrowth of prior knowledge—as influences on individual short-term recall of novel crossmodal associations.

CHAPTER 2

MEASURING WORKING MEMORY CAPACITY

2.1 What is the function of working memory?

Working memory (WM) makes fleeting sensory responses and stored representations available for cognitive processing, allowing us to keep varied kinds of information “in mind” at one time (Baddeley, 2007; Miyake & Shah, 1999b). A whole host of mental computations rely on working memory, from long division to sentence comprehension.

Decades of research have been devoted to quantifying WM capacity (Sweller, 2005, pp. 21–23), as it is a central limiting factor in our ability to make new associations. WM capacity has been shown to constrain learning in traditional classroom settings, for reasons ranging from teachers’ instructions exceeding some students’ ability to retain sequences of steps to intricate nuances of the reading process (Pickering, 2006).

2.1.1 How is working memory different from long-term memory?

Researchers disagree on *how* distinct WM is from long-term memory (LTM). Jonides and colleagues (2008) reviewed competing models which suggest WM’s information stores reside in the same or in different neural circuits as information that we can recall over the long term. While these debates continue, neuroimaging studies showing WM involvement of early perceptual areas in the visual system sustain core questions about the extent to which working memory is any sort of storage structure versus a set of processes (Harrison & Tong, 2009).

Taking another tack, scientists have differentiated WM capacity from attentional capacity by comparing observers’ success with two combinations of interleaved tasks. One

combination required detecting changes in an array of visual objects (a working memory task) and tracking the movement of multiple visual objects unrelated to those in the first task (this would measure attentional limits). The other combination involved two working memory tasks: looking out for changes (as above) in one type of display and attending to specific screen locations to count appearances of shape targets in a rapid stream of comparable shapes in another (Fougnie & Marois, 2006). If the capacity of WM were the same as the number of things we can attend to at once, the measured capacity for both task pairings should have come out equal, and it did not.

Broad agreement, though, surrounds the view that the amount of information that people can keep “in mind” to process simultaneously is vastly smaller than the amount of information they retain over the long term. This means that the theoretical distinction between WM and LTM is useful, and that prior knowledge plays a role. As an illustration, ask yourself which of the following nine-letter strings would be easier to keep in mind if you covered this page immediately after reading them, FBICIAGRE or EGFRCIIBA? Most Americans, seeing the familiar abbreviations (FBI, CIA, and GRE) in the first string, would have better luck repeating back all of those nine letters in order than they might with the second string, even though it comprises the same letters. This quick example highlights some of the difficulty in measuring how much an individual’s working memory will hold: is the first string of letters nine “things” to be held in mind, or three, or just one?

Some researchers grappling with this type of question have proposed elegant information-theoretic solutions to arguments about how working memory and its capacity limits may be implemented in the brain, suggesting that capacity itself may be fixed (Brady, Konkle, & Alvarez, 2008).

This simple alphabetic example above also illustrates the ongoing relevance of behavioral methods for working memory research. Whether or not information from the outside world is stored in different places or using different signal properties according to the sensory modality through which that information was acquired, if individuals’ success in forming audiovisual associations varies according to perceptual or conceptual characteristics of the

stimulus presentation, the latter variation adds to our overall understanding of working memory. Since this dissertation work examines which stimulus characteristics show consistent patterns as measured by individuals' short-term recall, any references below to WM capacity should be taken to refer to *demonstrated* capacity. Relative deficits or advantages between stimulus types for short-term recall indicate greater or lesser load on working memory resources, wherever and however those resources are implemented in the brain.

2.2 Are sensory modalities important for WM?

Working memory has been a central focus of cognition research due to its pivotal role both in maintaining new sensory impressions long enough for them to enter long-term memory and in making the content of those long-term stores available for manipulation (Miyake & Shah, 1999a). Early work relied on recall of verbal material—words and non-word letter combinations—as the defining measure for working-memory capacity, with heavy use of specialized words like numbers (Baddeley, 2007), which can be represented with only one or two visual symbols (graphemes or pictographs) in many languages.

Even in the formative days of working memory research, attempts were made to separate out the contributions of different sensory modalities. One standard approach used in WM research is to require participants to repeat words out loud or under their breath in order to block the strategy of mentally rehearsing words relating to stimuli to be remembered. This technique, called *articulatory suppression*, is often used as a litmus test to detect the use of verbal encoding strategies, as it has been shown to impair many memory tasks but not classical visual change detection (Baddeley, 2007; Vogel, Woodman, & Luck, 2001), described in the section on measuring WM capacity (see p. 13).

Lists of words, though portrayed here as a fairly limited initial approach to studying WM capacity, sparked what has become an enduring exploration of the role of modality in short-term memory resources. The memory research that inspired the applied studies mentioned in the introductory chapter stem from Paivio's groundbreaking work on short-term memory for words with different characteristics (Paivio, 1969). This makes use of

a classic measure of WM capacity: number of words (including number words) recalled from a list, in presented order or in any order (Bavelier, Newport, Hall, Supalla, & Boutla, 2008). Paivio discovered that people would remember more words from a list that referred to visible objects—e.g. “pig, house, square”—compared to less “imageable” words such as “democracy” or “joy” (Baddeley, 2007, see p. 86). This finding sparked the notion that different modalities access different short-term memory stores, often referred to as the dual-channel hypothesis.

How might the leap be made from recall differences among verbal stimuli (remember, Paivio’s result came from lists of words) to claims of capacity benefits from combining modalities such as sound and sight? Mayer and colleagues’ adaptation of Paivio’s channels for their cognitive model of multimedia learning construed those channels as pipelines running from sensation through working memory with very limited interaction before integration for longer-term encoding (Mayer et al., 2001, p. 190). This idea sheds light on the logic applied by Mayer and many other top researchers in cognition and learning, but some questions remain. After all, the visual and conceptual elements contrasted by Paivio were already bound together in participants’ long-term knowledge as semantic representations of words they could recognize and use. Learning *novel* associations joining image-based and auditory or other sensory inputs may or may not follow the same patterns. The work proposed below aims to answer a few of the questions stemming from Mayer’s and others’ extensions of Paivio’s lexical finding into crossmodal studies.

2.2.1 Modality and the psychology of learning

Combining modalities, such as when a GPS display talks to you while also displaying arrows on a screen, has been extensively explored for its potential to work around the limits of our working memory and boost how much new information we can learn at once (Mayer, 2005b; Sweller, 2005). Both of these strains of research are built on the idea that visual and verbal modalities use different working-memory resources; one of the most widely-used

memory models is that of Alan Baddeley (2007). We know that the model-makers' work involves many different trials of word lists; how do these more applied research lines look in practice?

Many of Mayer's studies used brief computer presentations about the formation of lightning or the operating phases of a piston (Mayer, 2001). Novice learners (primarily undergraduates) would see this presentation of simple diagrams, they would not control the pace of the presentation, and the accompanying narrative would either be a recorded voice or typed words above each diagram. A later test would assess not merely recall of facts, but participant ability to apply the scientific principles to a new related problem. These principles would not be required for later coursework for the participating students. Using this type of protocol, combining voice and diagrams had better learning outcomes than text and diagrams. While this method brought many beneficial changes to the study of the psychology of learning, particularly the emphasis on transfer of principles, it is worth noting that the topics tested were not related to participants' required schoolwork.

Another concern about some of the above literature is that the combined-modality advantage is only present for animations where the viewer has no control over its pacing (Low & Sweller, 2005; Moreno, 2006; Ginns, 2005). Slower readers may have been penalized in the text condition, and on a real-world level, many instructional software programs aim to put as much pacing control in the users' hands as possible!

Sweller's line of research began with younger learners in an even more ecologically valid setting than Mayer's: younger students learning math—a required subject. Many studies comparing ambiguous instruction with worked examples showed an advantage for the latter, and in the case of geometry proofs, keeping explanatory text next to the forms they related to led to more successful learning (Sweller, 2005). The core explanation in these cases rests on working-memory demands: novice learners do not yet have larger knowledge structures to help them retain the different pieces of a geometry or trigonometry problem as one "thing" in working memory, so they need several solutions laid out before them step-by-step, without risking losing their hold on one aspect of a given step while looking back and

forth between pages.

The common ground between these two research lines has been summarized for a book chapter, as follows.

The capacity limitations of working memory are a major impediment when students are required to learn new material. Furthermore, these limitations are relatively inflexible. Nevertheless, in this chapter we explore one technique that can effectively expand working memory capacity. Under certain, well-defined conditions, presenting some information in auditory mode can expand effective working memory capacity and so reduce the effects of an excessive cognitive load. This effect is called the *modality effect* or *modality principle*. It is an instructional principle that can substantially increase learning (Low & Sweller, 2005, p. 147).

As mentioned above, the necessary condition of having learning materials presented without learner control over the pacing of those materials suggests that the amount of time required to read text associated with graphics may be more explanatory than claims of changes to working memory capacity itself. In the research reviewed in the above chapter, the vast bulk of the cited results confirm that people show less interference in their performance of two completely unrelated tasks if those tasks involve different modalities. Oddly enough, the key result offered to counter explanations based simply on the time overlap between reading and visually scanning a diagram rests on a comparison between learner success with a geometry diagram and audio instructions versus successive presentations of the diagram and written instructions—so the latter group had no access to the diagram while reading the instructions, violating the researchers' own cognitive load principles for instructional design (Low & Sweller, 2005, p.153). Thus, many questions thus remain about the learning impact of combining modalities; at least in certain situations, researchers have found results opposite to Griffin and Robinson's (2005) finding that including a map did not aid student learning of Roman facts.

Looking at another modality, the use of gesture has been found to aid learning in

arenas as different as acting and geometry. Novice actors showed better recall of lines where they had been blocked to move than those where they stood still (Noice & Noice, 2001), and in math learning, gesturing by both teacher and students confers significant learning benefits for later problem solving (Goldin-Meadow & Wagner, 2005). Gesture itself may turn out to be an exception, since other neuroimaging work has demonstrated that different modality pairings are handled by different circuits, not one central modality-combining module (Calvert, 2001). Either way, the gestural results merit examination for other modality combinations, but this proposal will concentrate on auditory and visual modalities exclusively.

The predominant means of establishing and proving independence between modality-specific WM stores, particularly those beyond visual and verbal, has been to use unrelated tasks in differing modalities and show that they do not interfere with each other at the level that same-modality tasks do (see summary in Park et al (2007)). Those results, while very important and compelling, do not in themselves prove that coordinating tasks between modalities will increase WM capacity. This serves as another reason to keep some questions in mind about the learning studies cited above until they are bolstered by additional foundational experiments.

2.2.2 Neuroimaging of crossmodal processing

In the neuroimaging literature, simultaneous crossmodal input is seen as a strain or a luxury, used largely in case of unclear input from a single sense (Calvert, 2001); examples include people's use of a speaker's facial movements to identify unclear speech sounds. Many of these experiments include non-visual and non-verbal modalities such as touch in addition to the pairings discussed by educators. Unconnected, simultaneous tasks in different modalities cause processing logjams (Dux, Ivanoff, Asplund, & Marois, 2006). In other studies, however, crossmodal activation has been shown to happen even for irrelevant input (Fort, Delpuech, Pernier, & Giard, 2002). Once people know about a particular object, recognizing it will automatically activate crossmodal associations relevant to the

object (Amedi, Kriegstein, Atteveldt, Beauchamp, & Naumer, 2005; Postle, D'Esposito, & Corkin, 2005). These automatic activations may well complicate attempts to test only one modality to the complete exclusion of another, even with behavioral tools such as articulatory suppression available to the experimenter.

Another complication worth mentioning is that perceptual systems or processes that we commonly think of as one modality may actually be heterogeneous groupings of capabilities whose parallel findings can be composed into a unitary, consciously-perceived representation of the world. V.S. Ramachandran refers to vision as “a bag of tricks” (Ramachandran, 1990), and the analysis of face processing supports this view. By taking photographs of human faces and separating out the low- and high-spatial-frequency elements (loosely, the coarse and fine details) has revealed that people respond to emotional expressions in faces using more than one neural circuit, and that these circuits—not all of which involve the “vision” center in the occipital cortex—provide complementary information (Halit, Haan, Schyns, & Johnson, 2006; Holmes A, 2005; Pourtois, Dan, Grandjean, Sander, & Vuilleumier, 2005).

The final wrinkle to add in to the discussion comes from older work showing that modalities *change* even in adult humans, in ways that relate to prior knowledge. Within the auditory modality, trained musicians and novices show opposite patterns of hemispheric dominance when recognizing melodies (Bever & Chiarello, 1974). Three decades later, using functional imaging techniques, researchers have found that although all of their participants damped down visual processing to deal with auditory input, this resource clampdown leveled off for orchestra conductors, while it continued to intensify for non-conductors faced with more and more difficult auditory discriminations (Hairston et al., 2008). That result is consonant with demonstrated changes in utilization of peripheral vision for frequent video-game players (Green & Bavelier, 2006, 2003). And in an interesting reversal for long-held views about hard-wired gender differences, a mere ten hours of play on a specific action video game improved female players' spatial processing enough to eliminate gender differences among players (Feng, Spence, & Pratt, October 2007).

2.2.3 Current approach to modality

Relying on a behavioral measure such as the one used in the final stage of this dissertation research offers a chance to resolve some of the contradictory predictions around working memory for crossmodal associations from learning studies and a possible new addition to the neuroimaging toolbox if successful. The proposed work will also contrast crossmodal working memory capacity with past and concurrent assessments of unimodal working memory. For the latter, I will create image-image pairings and combined sounds to insert into the exact same structure of trials to assess unimodal change detection accuracy for associations of a comparable complexity to the crossmodal ones.

2.3 Measurement: What units does WM use?

2.3.1 Chunks, slots, and information load

Defining the units with which to measure WM capacity—and particularly whether those units should take whole things (words, visual objects, etc) or their component features (syllables, oriented lines, and so on) more into account—is a longstanding source of debate. Ever since George Miller’s influential treatise proposing the “magical number seven, plus or minus two” (Miller, 1956), many scholars have attempted to refine a solid answer to the question of how much working memory can hold. The more complex groups of information, such as the three three-letter abbreviations used in my simple example (see p. 5) would be called “chunks” in his terminology, which is still in use today.

While some might interpret human chunking ability to extend to all kinds of information groupings, for working memory, various aspects of the things to be remembered do make a difference. For example, people can recall longer lists of short words than of mixed short and long or all long words (Baddeley, 2007, p. 9). More modern research treatments replace the “chunk” with a “slot”, arguing that we have four slots with some ability to add in extra items; other scientists argue that truly the WM capacity limit is expressed in *information* content rather than any fixed number of objects (Jonides et al., 2008; Brady et al., 2008;

Alvarez & Cavanagh, 2004).

According to the latter view, working memory functions a little like compression software (e.g. StuffIt, PKZip, JPEG, tar, etc.): information from the outside world that is for some reason easy to compress will use up less of our total working memory capacity than information which is harder to compress. What might be *psychologically* easy to compress can be affected by prior experience, and thus in this view the WM load imposed by a given stimulus cannot be calculated without reference to the viewer's knowledge. Using the term very loosely, one could argue that JPEG or other compression algorithms make use of prior knowledge (in that case, encoded geometrical knowledge) to save and serve up an image covering many many pixels using less storage space than would be required to save a bitmap (a pixel-by-pixel format) of the same image. Referring back to our FBICIAGRE example, this aspect of the information-based approach fits well with observed memory behavior: people who already know something about the string "FBI" as a unit have a recall advantage with that string compared to, say, "IFB". Specific solutions used experimentally to gauge information load of visual stimuli will be discussed further with regard to detail, p. 16.

2.3.2 Change detection and visual and crossmodal WM

In recent decades, scientists such as Pashler and others (Pashler, 1988) have sought ways to measure capacity for *visual* working memory in a way that does not rely on symbols for spoken language: the core method that has been developed is called *change detection*. Viewers are presented with various visual stimuli, those stimuli disappear for a specific length of time, and then one or more of them reappears (Vogel et al., 2001). A comparable real-world example might involve a billiards table: say there are four different-colored balls left. If you commit their locations to memory, look away for a second, and then look again, how successful would you be at identifying any changes made by an interloper? While no method is free of limitations (see Alvarez and Thompson (2009)), this method as developed by Vogel and colleagues avoids both production issues (decreased apparent capacity due to the processing required to tell the experimenter what you saw or heard) and order

dependency. The latter—that is, basing WM capacity on the number of words a person can recall *in order* from a list—tends to indicate greater WM capacity for people using spoken language compared to signed languages like ASL, while unordered measures do not show the same discrepancy (Bavelier et al., 2008).

Factors important for visual WM

Visual WM research has further evaluated the mnemonic role of objects and features, with conflicting results. Support for greater memory capacity for objects was shown in Vogel, Woodman, & Luck's (Vogel et al., 2001) extended series of refinements on Pashler's (1988) visual WM methods. While their and other studies found that between three and four objects can be stored with varying amounts of featural detail recalled “for free” (Alvarez & Cavanagh, 2004; Vogel et al., 2001), comparable attempts by Treisman and colleagues supported separate memory stores for features and for bindings while failing to support any “free” featural complexity storage accruing to bound objects (Wheeler & Treisman, 2002). Xu's contrasts of multi-part objects showed that combining features from different dimensions (color and orientation) in the same object provided memory benefits, while combining different values on one dimension (with bicolored objects, for instance), did not (Xu, 2002). These unimodal results offer the possibility that crossmodal feature combinations may display greater working memory capacity than unimodal combinations.

A final complication in explaining featural complexity is the role of image location, long recognized for its importance in visual object binding and storage. Further discussion of location and its possible role in crossmodal feature binding follows the more detailed discussion of the change-detection methods used to test working memory. One of the aims of this research is to evaluate whether those prior (visual-only) findings have bearing on crossmodal (visual-auditory) associations.

Research in the less-thoroughly explored arena of crossmodal working memory must investigate the key factors of interest identified in verbal and visual WM research: featural complexity and location.

Crossmodal WM results

Recent work has sought to remedy the lack of attention to crossmodal feature binding in the cognitive literature (R. J. Allen, Hitch, & Baddeley, 2009), using crossmodal stimuli which divided up visual features that in unimodal trials were joined in a single image—color and shape—, pairing spoken words with grayscale or ambiguously-shaped images. My dissertation research seeks to clarify the outer bounds of crossmodal working memory capacity, while Allen and colleagues focused more on comparisons of crossmodal and unimodal performance under varied dual-task conditions. More specifically, their methods involve *splitting* visual objects' features and delivering some featural information verbally (R. J. Allen et al., 2009) rather than the present approach of *adding* auditory features to simple and complex visual objects. In addition, their method uses crossmodal presentation but tests participants' recall using only visual stimuli, assessing whether other-modal input can be incorporated into the visual working memory system. The present work evaluates participants' recall of crossmodal associations using crossmodal stimuli.

2.3.3 Visual and auditory details play a conflicting role

Evaluations of working memory capacity date have not yet developed fully-plumbed analyses of the role of perceptual detail. Assessments of visual working memory capacity compare freely between studies using using crisp and recognizable but relatively meaningless shapes which do have verbal labels (e.g. “square”, “triangle”) (Gajewski & Brockmole, 2006; Vogel et al., 2001), shapes carefully chosen to extremely difficult to describe verbally (Postle et al., 2005), cute cartoon animals (Horowitz et al., 2007), and the crossed bars and mushrooms (Xu, 2002) seen in the introduction. LTM is helped by visual detail, where photos in general are better recalled than typed words (Schmidt, 2006). Another recent study contrasting a more-realistic and a more-schematic depiction of the human heart sounded a different note, with more successful learning of ventricular blood flow from the latter (Butcher, 2006). Although the results were presented as an indictment of detail in instructional graphics, the two stimuli used differ in more ways than in their level of detail, including in the

clarity of grouping between atria and ventricles; such influences on Gestalt grouping have long been known to affect the utility of instructional diagrams (Vekiri, 2002). Spence (1990) showed that some of the same visual features maligned as “chartjunk” by Edward Tufte did help readers better understand the chart’s contents. Later research confirmed that although the addition of uninformative depth cues to graphs did reduce accuracy slightly, a larger influence was exerted by the perceptual context and the size and proximity of neighboring graph elements (Zacks, Levy, Tversky, & Schiano, 1998). Time-to-naming results *improved* for hand-drawn objects once color was added (Rossion & Pourtois, 2004). And complex sounds are apparently not harder to memorize than simple ones (Demany, Trost, Serman, & Semal, 2008). Finally, Jun-ichiro Kawahara’s work on contextual cueing (in press) also contrasted more and less complex sounds: recorded speech played backwards for two experiments, and telephone tones for a third. While configurations which had appeared regularly with the voice sounds showed faster search times in the last trial block, telephone tones failed to produce a cueing effect (Kawahara, in press).

Similar concerns about detail have been voiced with regard to learning materials for children, warning that overly attractive math and science manipulatives (physical learning tools) may take away from learning (Uttal, Liu, & DeLoache, 2006; Callanan, Jipson, & Soennichsen, 2002); both of these research teams, though, fail to distinguish between *novelty*—which has well-known cognitive effects—and perceptual detail itself. Also, more-distractible individuals—who otherwise perform more poorly than others in finding distinct objects in a visual display—are brought up to the performance level of their more-focused counterparts when the visual background is more crowded (Forster & Lavie, 2007)!

While background effects are beyond the scope of this dissertation, novelty effects are relatively easy to control for—or at least to exclude. For that purposes, all the experiments used to develop the crossmodal change-detection method involved participant training (see Appendix A.3) with the exact same limited range of images and sounds which were used in the change-detection experiments themselves.

Alvarez and Cavanagh (2004) used visual search time to quantify the visual information

load for each of their participants, an approach adopted in the present research. They found that a participant's speed in finding visual stimuli (which was faster for colored squares than for unfamiliar Chinese characters) was indeed predictive of the working memory load imposed by that type of stimulus (Alvarez & Cavanagh, 2004). By scaling their stimuli according to participant behavior, they were able to study the impact of visual information load without needing to resolve every difficulty involved in modeling it for all subjects. Their results are consonant with findings from an international and multilingual study of picture naming speed, where none of the measures of purely visual complexity (e.g. file size of the graphic according to multiple formats) was predictive of naming speed, whereas a behavioral measure was: pilot-tester ratings of how well each graphic depicted what it was supposed to represent did indeed predict naming speed (Székely et al., 2003). Each of the image sets tested by Alvarez and Cavanagh consisted exclusively of members of the same category: all Roman letters in one set, all shaded cubes in another, etc. Their training set of line drawings, though, comprised objects with similar outlines but very varied identities: a bunch of celery, a pen, and so forth. Could the very high WM capacity recorded for those drawings relate at all to the fact that they were different objects?

Some of the key findings in visual search which inform this discussion specifically rely on a count of low-level features and *not* an information-theoretic count to measure object informativeness or complexity such as that suggested by (Alvarez & Cavanagh, 2004). Many of those same findings, however, rely on abstract images which might not tie in to prior knowledge to the same extent that a picture of a bagel (or the mushroom pictured earlier) or a tricycle might. The latter images might also be considered more *meaningful* to observers, particularly in light of the findings discussed earlier (p. 10) showing that remote likenesses of known objects activate viewers' conceptual associations automatically.

2.3.4 Associations—is that the same as meaning?

Researchers of human memory would be hard pressed to make any progress without factoring meaning into their theories. Even memory for three-letter nonsense syllables is

influenced by their relative meaningfulness—closely tied to their pronounceability—, as exhaustively documented by Underwood and Schulz (1960). The roots of that research extend all the way back to Ebbinghaus' work in the 19th Century!

Describing the distinctions drawn above between Roman letters and Chinese pictographs in terms of *familiarity* may not generalize well to the images used in the method-development phase of this dissertation research: shaded balls and similar-sized color drawings (icons) of a tree, a chair, an apple, and a castle. For example, one University of New Hampshire student might have only rarely seen images of castles, but might have played so much croquet as a child that the colored balls would be highly familiar. Another might have been obsessed with English castles in junior high, seeing hundreds of images of them and perhaps even downloading the same drawing used in these experiments. Can we differentiate images in a manner less subject to individual variation? Although participants' responses will reliably not be identical—otherwise one would suffice—there are also clear cross-participant working memory patterns stemming from prior knowledge such as the distinction found between upright numerals (“2” and “5”) and their rotated equivalents, which were harder to find and to recall correctly (Alvarez & Cavanagh, 2004).

An artist or an art historian might characterize the castle and tree drawings as *representational* and the shapes as *abstract*—see for instance Merriam-Webster's fourth sense for “abstract” below.

having only intrinsic form with little or no attempt at pictorial representation
or narrative content <*abstract* painting>

This distinction has been used not only to distinguish art forms within one culture and era, but also to discuss the cognitive and cultural accomplishment levels of different groups of people over time (Halverson, 1992), and even as a personality measure (Furnham & Walker, 2001). A cognitive scientist might be more inclined to label the icons as more *meaningful* than the abstract shapes—how could such a claim be supported?

While decades of work in more than one subfield has been devoted to defining meaningfulness, established picture-naming (Székely et al., 2003) and semantic (Pexman, Harg-

reaves, Siakaluk, Bodner, & Pope, 2008; Pexman, Hargreaves, Edwards, Henry, & Goodyear, 2007; McRae, Cree, Seidenberg, & McNorgan, 2005) norms focus on concrete objects, omitting terms and images representing the abstract shapes (e.g. “oval”, “checkbox”) often used as item markers or icons in multimedia learning materials. Classic works measuring associations with abstract shapes, on the other hand, used such shapes to the exclusion of everyday concrete objects (Vanderplas & Garvin, 1959; Attneave & Arnoult, 1956). Apparently, neither the older nor the more recently established norms resolve whether such abstract shapes do empirically demonstrate less of the kinds of meaning assessed by those research groups. Also, meaningfulness (also termed *semantic richness*) explains variance in the newer studies only when considered alongside multiple other variables, requiring hundreds of words or images per study, rather than two dozen. For this dissertation, whose basic methods would not support hundreds of stimuli, a meaningfulness measure based on word associations will be loosely adapted (as described in Section 4.1.4, page 45) from McRae et al.’s (2005) number of features (NoF) measures and from many years of assessments of the diversity of word use among children (Hess, Sefton, & Landry, 1986; Johnson, 1944).

Whether the meaningfulness metrics or the behavioral (search-speed) measure is more predictive of crossmodal working memory capacity, both types of measures tap into observers past experience, or *prior knowledge*, of the world around them.

2.3.5 Prior knowledge and working memory

Prior knowledge has been shown to play a large role in learning in general, and Sweller and colleagues have found the same pattern when mixing levels of expertise among their participants, with more-knowledgeable learners requiring less structure and guidance than novices (Kalyuga, Chandler, & Sweller, 2000). Where the theoretical conflict arises between the Mayer and Sweller schools of research is in arenas where the prior knowledge does not relate directly to what is being learned—say in the case that you are learning to use an onboard GPS system while visiting your hometown versus while on your very first visit to Mount Rushmore. In Sweller’s (2005) model, a novice GPS user would not benefit

from prior location knowledge without explicit instruction, whereas researchers following Baddeley's (2007) model (Mayer among them) would predict such a benefit.

Some kinds of prior knowledge are easier to talk about than others—giving driving directions to my mother's house is substantially easier for me than explaining how to tell when to release the clutch after changing gears manually, yet both of those involve procedural knowledge. Our vast stores of long-term memory include procedural, narrative, and myriad other kinds of information.

Included in those stores are category distinctions: a toddler might not distinguish horses and cows, but adults can. Colors, dry or sweet wines, modern or early jazz—all these distinctions can become working parts of our sorting processes, giving definition to what we see, taste, or hear around us. Categorical distinctions are essential for the maintenance of multiple objects in visual working memory, as shown by Olsson and Poom (2005): Using change-detection methods very similar to those of Vogel, Woodman, and Luck (2001), they constructed stimuli which varied along non-categorical dimensions. For instance, some stimuli consisted of a white circle enclosed by a ring of blue—the ratio between blue area and white area varied continuously between all the objects to be remembered. In that case, participants' change-detection accuracy suggested that they could hold about one such object in WM at a given time, far less than the capacity found for stimuli distinguished by categories (e.g. blue versus red squares). This role of category distinctions may influence the role of the other factors under consideration for this study.

2.3.6 Which features are we counting?

Another approach that may explain conflicting effects of detail in different experiments suggests that for objects we can recognize, low-level features may not be the right ones to count anyway. Extensive work attempting to model object recognition suggests that when people recognize an object, the genuinely identifying may be mid-level features like a rough eyebrow pattern rather than low-level features like color or orientation of specific tiny regions (Ullman, Vidal-Naquet, & Sali, 2002; Borenstein & Ullman, 2002). The Ullman

work is an information-theoretic approach, where features are derived from calculations of which such mid-level features are most common within the category to be recognized (face, dog, car) and least common in non-category-members.

If the latter model closely describes the object-recognition process by which working memory distinguishes between items to store, then indeed, greater visual detail, *if such detail identifies an item as belonging to one or another category*, should boost change-detection capacity. On the other hand, perhaps greater distinguishing detail only helps up to a certain threshold, after which it becomes burdensome for working memory. Recent work in visual working memory suggests that we do indeed process identifiable objects differently than we do more abstract combinations of visual features (Hommel & Colzato, 2009; VanRullen, 2009). Recent modeling work has explored how visual objects could be encoded in such a way that familiar combinations take up less working memory capacity than rare or novel feature combinations (Brady et al., 2008).

Since an observer's prior knowledge holds such a major influence over WM capacity, how can we design experiments which take that knowledge into account—without first solving years' worth of modeling challenges—and still vary detail effectively? Image-based accounts alone, counting features or bits or other image representations, have not been predictive of visual working memory—after all, the luminance of an upright “5” is not different than a lying-down “5”—or of picture-naming speed (Alvarez & Cavanagh, 2004; Székely et al., 2003). Adopting a behavioral measure of visual information load will extricate the proposed work from some thorny debates about how to define detail while potentially clarifying the memory and modality dynamics inspiring those debates.

2.4 Goals of the present work

To arrive at the full comparison of crossmodal working memory capacity to behavioral measures used in visual cognition research, first and foremost a measure is needed for crossmodal working memory capacity. Thus this dissertation is divided into two main phases: developing such a method and applying that method to evaluate claims about

detail and meaningfulness of images associated with sounds.

The measure is an adaptation of change-detection procedures used to assess visual working memory, basing its capacity estimate on observers' rate of accurately detecting a change in one object out of several initially presented. Both abstract and representational sounds paired with images with more and less featural detail were evaluated, as was the impact of different test-object locations.

In applying the method, a meaningfulness measure was also developed which supported claims of greater meaningfulness (in terms of associations) for representational images compared to abstract shapes. The meaningfulness measure was used to prepare stimulus sets which contrasted image meaningfulness with image detail (color versus grayscale), characteristics which were tested in separate experiments in the development phase. A visual search task was incorporated alongside the core change-detection task to evaluate whether search speed would correlate strongly with working memory load of visual-auditory associations.

CHAPTER 3

CROSSMODAL CHANGE DETECTION: DEVELOPING A GENERAL METHOD

3.1 Crossmodal Adaptation of Change Detection Procedure

To measure crossmodal working memory capacity independent of language, adaptations to visual change detection methods were made as described below to assess the contributions to crossmodal binding of different sound and image types as well as test-object location.

3.1.1 General Procedure

Rapid Serial Presentation

This sequence of experiments adapted classic visual change detection procedures (Vogel et al., 2001) to measure crossmodal (auditory and visual) working memory capacity. This new method involves rapid sequential presentation of sound-image pairs, 500ms for each, with a test pair appearing after a delay of 1000ms. Unlike the all-centered serial presentation method devised by Allen and colleagues (2009), all memory-array images were placed at a fixed distance (approximately 3.5°, well within the parafovea) from the center of the screen and equidistant from each other. (See Figure 3-1 for a sample sequence.) Observers' accuracy in detecting whether or not the pairing was indeed present in the training array provided a quantitative measure of their WM capacity for such visual-auditory associations. Further comparisons to Allen et al.'s methods are discussed with Experiment 4. Instructions for the experiments in this chapter involved brief, individual training with the experimenter (see Appendix A.3).

The 1000ms delay between the presentation of the initial array (the first three objects)

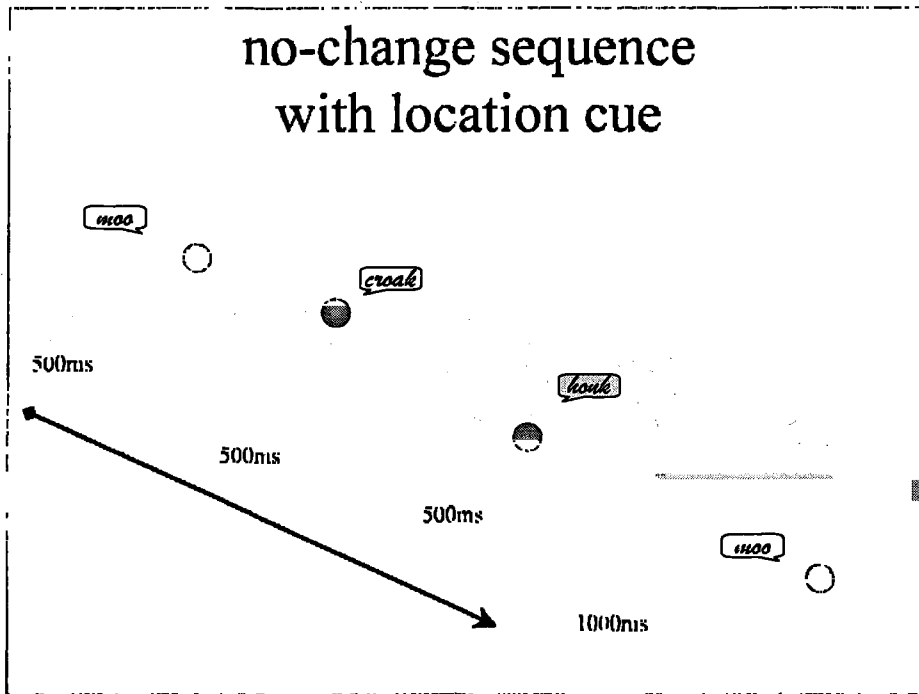


Figure 3-1: Change Detection: This would count as a *no-change* trial, as the test object has the same visual-auditory pairing as the first object presented. Depending on the experimental condition, the balls might be uniform gray or each a different color.

and the test object is identical to that used in Vogel, Woodman, and Luck's (2001) experiments, and it is well into the working-memory (rather than iconic-memory, closer to 100ms) range (Treisman, 2006). Delays of two seconds or more lead to loss of memory performance (Eng, Chen, & Jiang, 2005).

While the presentation duration—half a second for each object, with a total study time of one second—is longer than many of the prior visual-only WM capacity studies (Vogel et al., 2001), those authors argue that more-complex stimuli may require more time to process, as do subsequent studies (Jonides et al., 2008, see p. 201). Each object receives the minimum exposure time accorded to complex stimuli in one of the key studies Jonides is summarizing (Eng et al., 2005), and the total study time of 1,500ms is slightly longer

than their medium-length exposure time (1 s).

Image placement for test objects

Location has been identified as more crucial than other features used to perceive objects in many studies of feature binding and visual working memory capacity (Treisman, 2006). Therefore, two different approaches to placement of test objects were compared: displaying them in the location where they had originally appeared in the memory array (the original location) or in the center of the screen, a location never used for presenting stimuli to remember (the centered-probe condition).

Feature swapping

The other major difference between the crossmodal change detection method used here and some of the visual working memory precedents from Luck, Vogel, and Treisman and colleagues is its exclusive reliance on feature swapping. Wheeler and Treisman (2002) contrasted memory for binding of features with memory for the (presumably unbound) features themselves, and in the latter case, a changed test display would include features that had not been present in the memory display. Related work testing out the impact (described earlier) of varying features along the same or different dimensions also would introduce features (colors in this case) at test which had not been present in the original memory display (Xu, 2002).

The present method is specifically geared towards testing binding between features conveyed through different modalities, so—more like the binding-test trials from Wheeler and Treisman—any feature presented in the test phase of a trial would have been seen or heard immediately before. In other words, the only changes were the associations between features. If our tests included features that had not been presented immediately before, participants could successfully identify changes without making any connections between auditory and visual features at all: if no dog sound was present in the memory array, and the test pair is a barking apple, voilà!—a change. Each memory trial used three visual-auditory pairs,

with images and sounds drawn randomly from a set of four possibilities. In change trials, only images and sounds used in that trial were used in the test object.

Focusing in this manner on memory for crossmodal binding does permit a possibly confounding strategy, identified by (R. J. Allen, Baddeley, & Hitch, 2006), where observers can identify a variety of possible binding changes based on one accurately-remembered pairing. Using the trial displayed in Figure 3-1, an example of this strategy would be for the observer to remember only that the lower-left ball (appearing last in the memory array, the third box in the figure) honked like a goose, without any retention of the other two stimuli. This participant could correctly detect a change in the image-sound pairings for this trial if that last pairing comes back, if the honk sound comes with any other image, or if that ball from the lower-left corner makes any non-honking sound. Since test probes in this swapping protocol only select from the three sounds and three images presented in the memory array, there are nine possible visual-auditory pairings for a given trial, and a person using this single-pairing strategy will be able to respond correctly to five out of those nine.

This strategy option and our sequential presentation complicate the interpretation of prior estimates of absolute WM capacity, and these experiments are agnostic about the correct units with which to measure it. Instead of focusing on absolute capacity, my analyses concentrate on response accuracy, as follows. Subtracting participants' false alarm rate from their score for correct change detection, allowing us to penalize guessing and remove individual response bias (R. J. Allen et al., 2006). Chance responding would score 0% accuracy, versus 100% for error-free change detection.

3.1.2 Participants

Twelve students from the University of New Hampshire—each taking part in a single experiment—participated in each experiment, with the exception of the last experiment, which had 13. Most received course credit or \$12 in compensation; the time involved was approximately 50 minutes.

3.1.3 Apparatus

Experiments 1 through 5 were run on a Dell desktop computer with an LCD screen, using custom software created by Colin Ware. The sixth experiment was presented on two Macintosh iBook G4s, using code written by the author in PEBL (Mueller, 2006). Sounds for all experiments were presented using adjustable headphones (Philips SHP2500). Participants pressed keys labeled "SAME" or "DIFF" to record their answers.

3.1.4 Stimuli

To keep our experimental crossmodal associations as simple, novel, and consistent as possible, the images to associate were chosen to either represent abstract shapes (shaded balls) or non-noise-making objects (e.g. trees, castles, household items). Please refer to Appendix A.2 to examine the images used.

Associated sounds consisted of 400ms recordings of animal sounds, monosyllabic words, or pure tones; specific stimulus choices for each experiment are listed in Table 3.1 and described in more detail for each experiment.

3.2 Experiment 1: Contrasting Unimodal and Crossmodal Performance

3.2.1 Method and stimulus specifics

Crossmodal accuracy was first compared to unimodal performance with serially-presented colored balls which could change color or not, with test objects always appearing in the same location as where they were first presented. (Image stimuli for Experiments 1-6 are shown in Appendix A.1.) Associated sounds consisted of 200, 350, 500, or 650 Hz constant tones (Fabiani, Kazmerski, Cykowicz, & Friedman, 1996). Articulatory suppression was used to assess involvement of the phonological loop: observers spent half of their trials repeating randomly-chosen pairs of numbers presented at the start of each trial.

<i>Experiment</i>	<i>Images</i>	<i>Sounds</i>	<i>Test Location</i>
Expt. 1	Colored Balls	Pure Tones	Original Location
Expt. 2	Gray & Colored Balls & Color Drawings	Pure Tones & Animal Sounds	Original Location
Expt. 3	Gray & Colored Balls & Color Drawings	Animal Sounds	Original or New Location
Expt. 4	Gray & Colored Balls & Color Drawings	Animal Sounds or Spoken Words	Original Location
Expt. 5	Gray & Colored Balls & Color Drawings	Pure Tones & Animal Sounds	Original Location
Expt. 6	Color Photos or Grayscale Drawings	Animal Sounds	New Location

Table 3.1: Experimental Design: Stimulus Contrasts

3.2.2 Hypotheses: Experiment 1

Perhaps binding visual and auditory features as crossmodal objects is comparable to binding visual features from very different dimensions—in other words, a sound feature is simply *more* different than an orientation feature associated to a color. If so, extrapolating from Xu's (2002) differentiation rather broadly, observers' accuracy for crossmodal stimuli should equal or exceed their change detection results for unimodal stimuli. If instead crossmodal binding is less successful than binding different-dimension visual features, unimodal accuracy should be higher. If rapid serial presentation of colored shapes parallels simultaneous presentation exactly (Vogel et al., 2001), articulatory suppression should not affect visual-only trials. Finally, if the phonological loop is used to process abstract tones as crossmodal features, articulatory suppression should impair change detection on crossmodal trials.

3.2.3 Results: Experiment 1

Excluded Data

Technical issues and data analysis dictated the removal of data sets, as follows. One participant in the first experiment was unable to successfully operate the experiment program, and two others had more than three error trials in specific blocks; those data were excluded. Also, zero to two participants in each experiment responded with accuracy rates which were not distinguishable from chance—these records were also excluded. Reported *Ns* reflect the adjusted total number of participants from whom valid and complete data sets were obtained.

Effects

Corrected accuracy was significantly higher for unimodal change detection, $M_{no-suppr} = 77.1\%$, $M_{suppr} = 71.5\%$, than for crossmodal, $M_{no-suppr} = 55.9\%$, $M_{suppr} = 32.3\%$. The unimodal advantage was significant, $F(1, 88) = 45.4$, $MSE = 77.1$, $p < .001$, and articulatory suppression depressed accuracy across the board, $F(1, 88) = 10.6$, $MSE = 77.1$, $p < .01$, with a marginally significant ($p < .06$) interaction reflecting a greater impact on crossmodal trials. Crossmodal detection accuracy for this and all subsequent experiments is shown with a standard error term based on the residual standard error of the analysis of variance used to determine significant effects. See Figure 3-2 for a graph of the results of Experiment 1 and other method-development experiments.

3.2.4 Discussion: Experiment 1

These results suggest that, compared to associating color information with a located shape, associating sounds with uniquely-colored shaded balls consumed considerably more working memory resources. Similar to Allen and colleagues' recent results (R. J. Allen et al., 2009), articulatory suppression impaired recall for both kinds (visual-visual and visual-auditory) of feature binding.

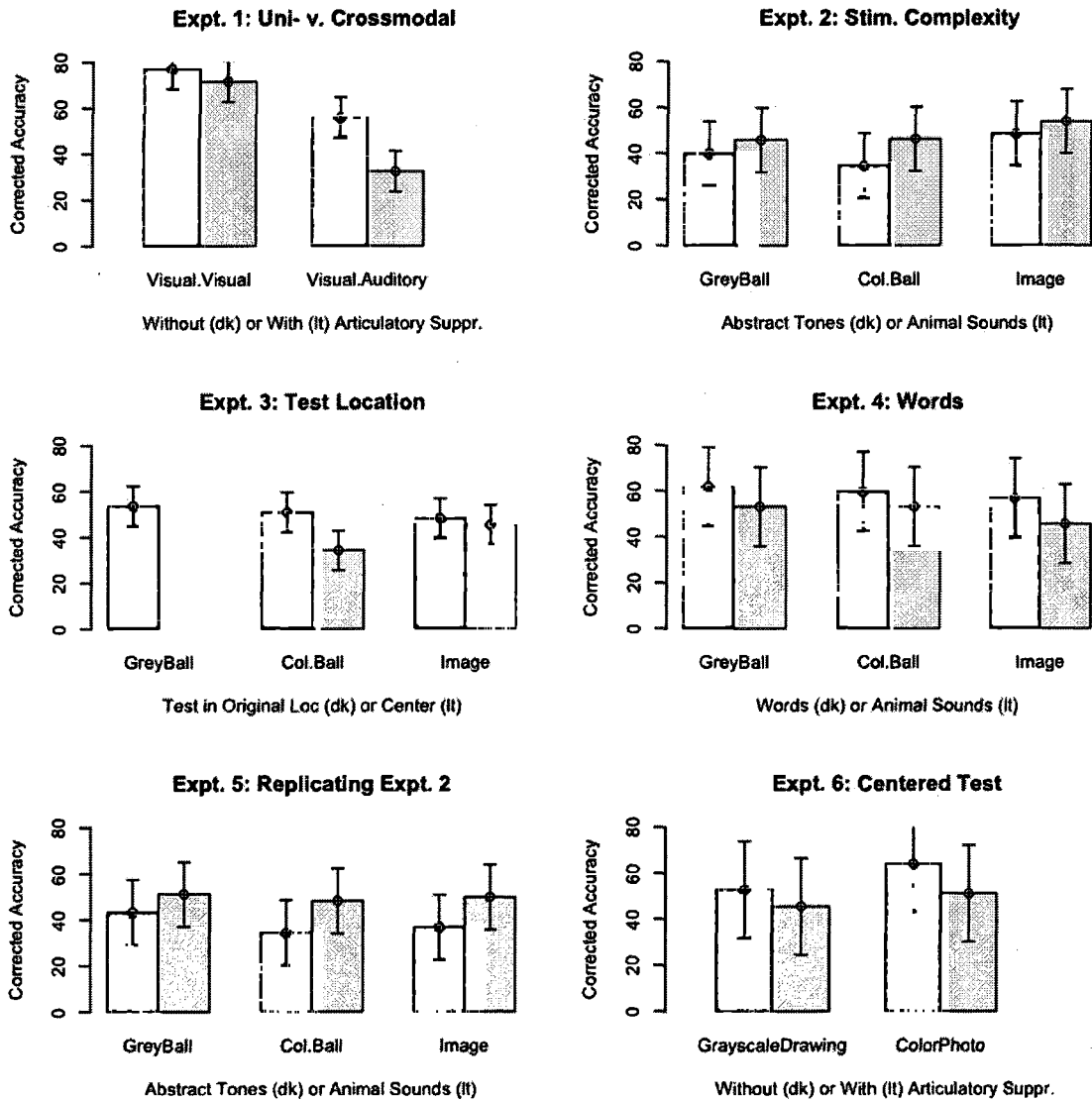


Figure 3-2: Change-detection accuracy results measured using $p(\text{Hit}) - p(\text{FA})$, with residual standard error indicated.

3.3 Experiment 2: Contrasting Simple and Complex Stimuli

To compare the impact of more complex (more highly-featured) stimuli on crossmodal WM capacity to related visual research, the second experiment diversified both the auditory and visual stimuli used. A second set of more meaningful and feature-rich sounds was selected

to contrast with the pure tones: animal sounds, each lasting within 50ms of the 336ms duration of the tones. The sounds chosen—a cow’s moo, a goose’s honk, a frog’s croak, and a loon’s laughing call—differed from each other on many auditory dimensions, including pitch, rhythm, and timbre. All trials involved crossmodal associations.

To complete the comparison of meaningful stimuli with abstract ones, an additional set of images was included, showing a tree, a stone castle, an apple, and a wooden chair. None of these images portrayed animals, to minimize any chance of semantic connections between the images and the animal sounds. Also, a set of identical grayscale balls were included to provide a less-featured abstract comparison for the colored ones. (See images in Appendix A.1.)

3.3.1 Hypotheses: Experiment 2

If number of low-level features exclusively determines the working memory load of visual and auditory stimuli, participants’ change detection accuracy should be highest for gray balls and for abstract tones, with drawn images and animal sounds showing the worst performance. If the meaningfulness of those images and sounds—tying in with long-term knowledge of some kind—boosts working memory capacity, the reverse should hold.

3.3.2 Results: Experiment 2

The more complex sounds, $M_{animal} = 48.6\%$, $M_{tone} = 41.1\%$, $F(1, 41) = 12.78$, $MSE = 2.82$, $p < .001$, and the drawings of real-life objects, $M_{image} = 51.2\%$, $M_{coloredball} = 40.5\%$, $M_{grayball} = 42.9\%$, $F(2, 41) = 3.93$, $MSE = 2.82$, $p < .05$, both showed main effects of improving change-detection accuracy. (N.B.: The effect of image type failed to replicate in later experiments, as detailed below.) Accuracy results for this experiment are shown in Figure 3-2.

3.3.3 Discussion: Experiment 2

These results do not support a straightforward featural complexity cost to crossmodal working memory capacity. More-complex animal sounds and images—both plausibly more meaningful than their abstract counterparts—were recalled more successfully than abstract tones or shaded balls.

3.4 Experiment 3: Does Test-Probe Location Matter?

To assess the contribution of location cues to recall of crossmodal pairings, the conditions of Experiment 2 were repeated with the following changes. Abstract tones—greatly disliked by participants—were removed. New trial blocks were added without location cues at test, with test image-sound pairs presented at the center of the screen. Center-test trials could only be implemented for the unique image types (colored balls and images), as sound associations with the the identical gray balls could only be differentiated by locations, resulting in an asymmetrical experiment design. Control trial blocks, with test objects reappearing in their original location, were (other than randomization) identical to the animal-sound blocks of Experiment 2. (See Figure 3-3 to compare a center-test sequence to the original method.)

3.4.1 Hypotheses: Experiment 3

If location is as important for crossmodal associations using non-localized auditory features as it is for visual-visual feature binding, change-detection accuracy should be poorer for centered probes. At first blush, this hypothesis seems likely, as more retrieval cues are at play for test probes in their original locations. However, for single-probe tests of feature binding after simultaneous presentation of multiple shapes, location of the test probe made no difference in a key precedent (Treisman & Zhang, 2006). If single-probe test protocols for rapid serial presentation are identical to this precedent, no difference in crossmodal change-detection accuracy should be observed between test probes in their original locations and centered-test trials. Following Experiment 2, complex and recognizable images should show the best performance compared to the gray and colored balls.

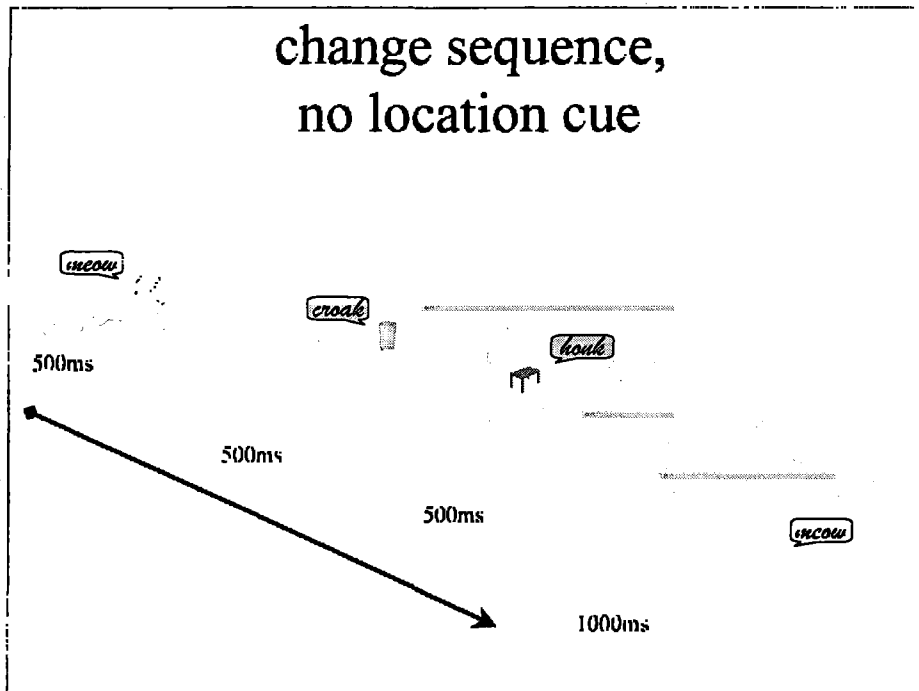


Figure 3-3: Change Detection: This would count as a *change* trial, because the glass was first paired with a frog sound, not a cat's. With the test image placed in the center, observers cannot rely on location cues to detect changes in sound-image pairings.

3.4.2 Results: Experiment 3

Participants recalled crossmodal associations with location cues $M = 50.9\%$ better than those tested at the screen's center, $M = 39.7\%$. The location-cue advantage was significant, $F(1, 121) = 15.66, MSE = 65.6, p < .001$, while image types did not exert a consistent main effect, failing to replicate the image difference found in Experiment 2. (See Figure 3-2 for full results.) Crossmodal associations using gray balls were most successful—and representational images least successful—in the location-cue condition, while representational images were easier to recall than colored balls in absence of location cues. This interaction was significant, $F(1, 121) = 4.79, MSE = 65.6, p < .05$.

3.4.3 Discussion: Experiment 3

Image-sound pairings were easier to recall accurately when test probes were presented in their earlier locations than when they were placed in a novel, centered position. The interaction between test-probe location and image type prompted both the replication of Experiment 2 (as described above) and the always-centered test design of Experiment 6.

3.5 Experiment 4: Will Words Improve Performance?

Past research on verbal working memory has shown that lists of five short words can be recalled and repeated with near-perfect accuracy (Baddeley, 2007, p. 9). To find out how much words might improve recall of novel visual-auditory associations, animal sounds were contrasted with monosyllabic English words in visual-auditory pairings. Recorded words—all monosyllabic and imageable, such as “sphere”, “sock”, “moon”, and “cup”—had similar durations to the animal sounds, and they were pilot-tested for intelligibility in a word-sequence identification task mimicing the auditory presentation protocol for the main experiment. Pilot testers did not make order-recall errors with lists of three words (chosen randomly each time from the set of four). This protocol was the most similar to Allen et al.’s crossmodal approach (2009), with two key differences. Their crossmodal stimuli divided up visual features that in unimodal trials were joined in a single image—color and shape—, pairing spoken words with grayscale or ambiguously-shaped images. The crossmodal stimuli used here combine complete images (compared to past unimodal controls) with additional auditory features which would be difficult to depict—for the animal sounds—or would require an additional image. The second key difference lies in the semantic plausibility of the verbal material used in their experiment (R. J. Allen et al., 2009). The words chosen for use in Experiment 4 did not obviously relate to the images used (e.g. a castle saying “sock”). Judging by what stimulus combinations prompted laughter in training, to many participants, small abstract images having a nameable shape is a more expected characteristic than for those images to bark, honk, or moo.

3.5.1 Hypotheses: Experiment 4

If image-based and sound-based WM resources constitute entirely separate stores, highly-imageable words should overload visual WM capacity when paired with irrelevant images, resulting in better performance for the animal sounds. On the other hand, if associating novel pairings incurs very little processing cost, the associated words should be so easy to recall that change-detection performance should approach the high rates found in unimodal trials in Experiment 1.

3.5.2 Results: Experiment 4

The visual-auditory pairings featuring words, $M = 59.5\%$, were better recalled than the ones using animal sounds, $M = 50.6\%$. The word advantage was significant, $F(1, 106) = 10.6, MSE = 268.8, p < .01$; no significant differences were found between image types. Accuracy results for this experiment are shown in Figure 3-2.

3.5.3 Discussion: Experiment 4

While images paired with spoken words were easier to recall than those paired with animal sounds, top performance on Experiment 4 still does not approach the unimodal accuracy rates observed in Experiment 1—this suggests that maintenance of the pairing itself requires cognitive resources. The absence of significant differences according to image type also provides an additional failure to replicate Experiment 2's image results.

3.6 Experiment 5: Contrasting Simple and Complex Stimuli—Replication

In light of the lack of consistent cost or benefit associated with different image types in Experiments 3 and 4, we replicated Experiment 2 in its entirety, crossing image-type differences (gray balls, colored balls, and drawn images) with sound-type differences (pure tones or animal sounds). The hypotheses for Experiment 5 were identical to those of Experiment 2.

3.6.1 Results: Experiment 5 (replication)

In this replication of Experiment 2, the complex-image benefit observed earlier failed to replicate, $F(2, 63) = 2.11, MSE = 198.99$, while the advantage for animal sounds, $M = 49.6\%$, over pure tones, $M = 38.2\%$ was supported, $F(1, 63) = 21.26, MSE = 198.99, p < .001$. See Figure 3-2 on p. 30 for full results. A significant interaction was observed between image type and individual participant identity in the replication $F(20, 63) = 2.10, MSE = 198.99, p < .05$; no such interaction was observed in Experiment 2, $F(12, 41) = 1.30, MSE = 2.82$. With only twelve people participating in each experiment, there could be individual perceptual or strategy differences whose patterns are not captured by these data.

3.6.2 Discussion: Experiment 5

This replication did not support a clear featural-complexity cost or benefit for associated images in crossmodal binding. While the software and hardware employed were identical, the two experiments (2 and 5) were run in different buildings, with more physical distance and less of a sight line between experimenter and participants in the latter. Although the experimenter used the same training process in both to deliver instructions (see Appendix A.3), reliance on verbal instructions—corrected in the final experiment—may have introduced errors, and changes in department policy meant that Experiment 5's participants had read a brief description of the experiment before signing up, while Experiment 2's had not. The Experiment 5 participants were from later in the semester's pool, in addition, and had more information about the experiment beforehand. All of these slight changes, combined with the relatively small sample size, make interpretation of the inter-replication differences difficult, other than clearly pointing to the need for further study as in Experiment 6.

More-complex animal sounds, on the other hand, were once again better recalled than abstract tones, while image types produced no significant pattern of cost or improvement. Comparing these results with those of Experiment 3 suggests a possible interaction where image complexity plays a different role when participants cannot use location cues.

3.7 Experiment 6: Do Image Differences Matter More, Absent Location?

To better evaluate contributions of image differences to crossmodal WM capacity, the Experiment 6 presented each test probe at the center of the screen, removing the option of correct change detection based on prior *location* rather than other image characteristics. In addition, more highly-featured meaningful images (color photos of a book, a butterfly, a tree, and a Granny Smith apple) were contrasted with grayscale drawings proven to be highly identifiable in multiple picture-naming experiments (Rossion & Pourtois, 2004). (See specific images used in Appendix A.1.2.) Responding to past participants' concerns about having two birds represented in the set of animal sounds, the loon recording was replaced with a cat's meow. This final experiment also relied on articulatory suppression, with participants asked to repeat random number pairs under their breath for half of the trials.

3.7.1 Hypotheses: Experiment 6

If the image-complexity benefit shown for centered-probe trials in Experiment 3 was solely due to the fact that the representational images depict real-world objects while the abstract shapes do not, then highly recognizable grayscale drawings and color photos should show comparable change-detection performance. If perceptual characteristics such as the greater number of differently-oriented lines in the images also contributed to that benefit, the color photos should demonstrate higher accuracy. If the more detailed images prompt observers to rely more on verbal encoding, subvocal articulatory suppression should impair performance on those trials more than on the trials using simpler grayscale drawings.

3.7.2 Results: Experiment 6

Full-color photographs, $M_{no-suppr} = 64.1\%$, $M_{suppr} = 51.2\%$ provided greater crossmodal change detection accuracy than grayscale drawings $M_{no-suppr} = 52.9\%$, $M_{suppr} = 45.5\%$. In addition to this significant effect of image detail, $F(1, 96) = 4.25$, $MSE = 441$, $p < .05$, subvocal articulatory suppression reduced accuracy across the board, $F(1, 96) = 6.01$, $MSE =$

441, $p < .05$, without any observed interaction with image differences, $F(1, 96) = 0.44$, $MSE = 441$. Accuracy results for this experiment are shown in Figure 3-2 on p. 30.

3.8 Discussion: Method Development

3.8.1 Overall Results

This crossmodal adaptation of classic change-detection methods has shown that sensory modalities and factors critical to object perception—featural complexity and location—themselves shape observers' crossmodal working memory capacity, and that those critical factors interact. Specifically, the results from these method-development studies show that participants' recall of image-sound pairings in certain cases does differ for different image types, and the image differences are not well explained by counting low-level features. Detailed images representing real-world objects sometimes provide differing WM results than more abstract depictions.

Application of this new crossmodal change-detection method suggests that WM capacity for visual-auditory associations is lower than capacity for visual-only associations. This series of experiments emphasized relative rather than absolute claims about WM capacity; further discussion of absolute capacity estimation can be found in Appendix B.1. Using their feature-splitting approach, Allen et al. found no change-detection differences across unimodal and crossmodal presentations, but their participants performed articulatory suppression more slowly under crossmodal presentation conditions (R. J. Allen et al., 2009, p. 100). For the present sequence of experiments, comparing crossmodal to unimodal performance—in our own first experiment and in others' (Vogel et al., 2001)—shows a clear crossmodal disadvantage.

Change detection itself may underestimate true working memory capacity (Alvarez & Thompson, 2009), as shown by comparing feature-swapping change detection techniques (Saiki & Miyatsuji, 2005) with others which have been used to show that binding persists in absence of attention (Gajewski & Brockmole, 2006). The method used here relied exclusively on feature swapping—that is, features presented at test were always drawn from those

which had been presented in the preceding memory array, so only bindings were altered for change trials. However, in the changed-location trials, the test location was not ever used in memory arrays, so those trials should exhibit less overwriting than the same-location trials. Having the mnemonic boost of a specific location contributed more benefit than any detriment from overwriting of features associated with that location.

Retaining image location at test showed a clear advantage for recall of visual-auditory pairings. This is similar to the results of (Prabhakaran, Narayanan, Zhao, & Gabrieli, 2000), which did involve changing test-probe locations to previously-used areas on the screen; their results supported a clear benefit for location. Our crossmodal results support the central role of location in binding, as the auditory features themselves were not localized acoustically and were bound to images only in the perception of the participants. These results contrast directly with recent findings that location did not boost visual working memory for single-probe test protocols (Treisman & Zhang, 2006), suggesting that even 500ms' exposure to a solo visual-auditory pair fostered its encoding with relation to the borders of the experiment presentation screen.

While location clearly did contribute to these crossmodal associations, the impact of visual and auditory featural complexity was not identical across experiments. In two replications of the same protocol, more-complex animal sounds provided consistent change-detection benefits compared to abstract tones. Those replications failed to support any consistent pattern of accuracy differences according to image complexity, however. Only in the cases when image location cues were not available did greater visual detail improve recall for crossmodal associations. Given that visual working memory itself depends on categorical distinctions (Olsson & Poom, 2005), one possible explanation would be that the abstract tones were only marginally distinctive enough to support greater working memory capacity. Alternatively, the location-cue trials (where the test probe was presented in its original location) may have permitted a variety of mnemonic strategies to participants, who could successfully detect changes by recalling associations between sounds and specific images *or* specific locations. If true, the availability of these qualitatively different strategies

may have obscured image-related differences in recall.

Finally, the pattern of results observed supports a nuanced treatment of features (Olsson & Poom, 2005; Alvarez & Cavanagh, 2004; Xu, 2002) more than a simple count of low-level features as predictive of working memory load. More-complex images were not associated with a significant decrease in change detection accuracy, and were in fact easier to recall than less-complex ones in cases where the test probe did not cue the prior location of the test pair.

3.8.2 Suggested Modifications

To clarify the results obtained thus far, several changes are in order. Firstly, centered placement of test probes (that is, testing for change detection in a location unused in each memory array) is essential to discover the working memory impact of image-type difference. Thus this approach, as used in Experiment 6, should be used for subsequent applications of this change detection method. Secondly, those image-type differences already attested in one case (Experiment 3) involved meaningful color images showing better accuracy than abstract colored shapes, and in the other case (Experiment 6) both sets of images depicted recognizable, meaningful objects, but the advantages accrued to the full-color photos compared to the grayscale drawings. Further analysis of the impact of these image characteristics on crossmodal working memory capacity requires an experiment design which contrasts meaningfulness with perceptual detail.

CHAPTER 4

CROSSMODAL WORKING MEMORY LOAD, DETAIL, AND MEANINGFULNESS: REFINING METHODS

As stated in the previous chapter, thorough examination of the effects of image type on crossmodal working memory capacity requires juxtaposing higher- and lower-featured instances of both representational images and abstract shapes within one experiment. Two further extensions are in order: First, image *meaningfulness* may influence the working memory load of image-sound associations; determining this requires an operational definition of meaningfulness which can be applied to all the image types used in this dissertation. Second, the image and sound complexity advantages found in the earlier experiments call for a comparison to a behavioral alternative to feature-based complexity measures for visual objects, developed by Alvarez and Cavanagh (2004).

The four image types required to cross the effects of featural complexity with the distinction between representational images and abstract shapes are: full-color representational images, grayscale representational images, full-color abstract shapes, and grayscale abstract shapes. These image classes, if six exemplars are chosen for each type, lend themselves to the same sort of comparison as Alvarez and Cavanagh's (2004) five image classes: colored rectangles, capital letters, Chinese characters, random polygons, and shaded cubes. Their behavioral measure, which was found to correlate very strongly with the working memory load of each image class, was an estimate of the per-item visual-search cost (in milliseconds) for search arrays with more and more members of that class of images displayed.

Further consideration of meaningfulness

As discussed in Section 2.3.4, another way to characterize the difference between representational images and abstract shapes would be to say that the former are more *meaningful*. Assessing the working-memory impact of meaning could possibly clarify whether perceptual expertise alone (e.g. having seen and recognized trees before possibly more often than a particular type of abstract shape) is helping people recall more centered probes with representational images (compared to the shapes), or if something more directly related to prior knowledge stored in long-term memory is in play. But how to define *meaning* in this case? Paired associate learning results showed again and again that pairs of more-meaningful pseudowords are easier to remember, with meaning there closely tied to pronounceability (Underwood & Schulz, 1960). Later twists on that research—using paired associates based on real words—attempted to tease apart the contributions of meaningfulness and imageability, with the latter helping long-term learning in ways that the former did not (Paivio, 1969).

Since none of the models of word or image meaningfulness reviewed in Chapter 2 fit the scope of the present work, a new way of operationalizing meaning had to be developed and tested. Pilot testers and experiment participants were shown each image one by one, and they wrote or typed several associations about each image used. Participants had no specific constraints on these associations, which could be descriptive, functional, or personal. The classes of images were then characterized as more or less meaningful according to the number and variety of word associations provided by each person for that group of images.

4.1 Image selection

Multiple images for each of the four classes (see Table 4.1) were selected or created to satisfy the needs of this new examination. For color photos, royalty-free thumbnail images were downloaded from a now-defunct stock photo distributor. Similar to the preparations for Experiment 6, grayscale images were adapted from color drawings (Rossion & Pourtois, 2004) by minimizing color saturation. Colored shapes were created by Colin Ware and the

	<i>Representational Image</i>	<i>Abstract Shape</i>
<i>Higher Detail</i>	Color Photo	Colored Shape
<i>Lower Detail</i>	Grayscale Drawing	Grayscale Shape

Table 4.1: Schematic showing the image categories to be tested.

author using several graphics tools from Adobe. Grayscale equivalents of those shapes were developed by setting color saturation to -100%, as for the drawings. To further strengthen any conclusions drawn from differences in observer performance, two separate sets of images were chosen that conform to the constraints laid out below.

4.1.1 Overall constraints

Each image measured between 1.5° and 3° on a side. Height- and width-based measurements were balanced across conditions and participants: Half had “wide” (or matched for horizontal visual angle of 3°) color photos and grayscale shapes but “tall” grayscale drawings and colored shapes; the other half saw the opposite. Please see Appendix A.2 to review the images chosen after the completion of this review process.

4.1.2 Representational Images

Since real-life object size can have an impact on perception and memory (Konkle & Oliva, 2007), the drawings and photos selected all depict objects that could easily be easily picked up using only one hand. To maintain similar levels of image-sound irrelevance across all trials, no images of talking, calling, or singing animals were used.

The featural contrast between the selected grayscale drawings and their color-photo counterparts involves not only color but differences in shading, and each representational image had far more variety of line orientation in the object’s edges than the abstract shapes had. This disparity allows for an additional experimental contrast illuminating the role of perceptual saliency in working memory research. Color and line orientation are both highly

salient cues for directing visual attention, and visual search is faster for arrays of more homogeneous than heterogeneous distractors (Wolfe & Horowitz, 2004). In visual change detection, though, any item in the original array can be a target or a distractor. Only using abstract shapes, Olsson and Poom (2005) found that greater *categorical* distinctiveness boosted change detection accuracy, and in grating and sound recognition, others have found lowered detection bias but unchanged accuracy for more heterogeneous displays (Sekuler & Kahana, 2007). This final experiment compares the role of color on the abstract-shape side, where that one additional feature difference between shapes makes a very large difference in the number of low-level features differentiating images in the set, to its role on the representational side, where it is merely one more among many differences, albeit a highly salient difference. This additional comparison also affords a new perspective on later studies of image complexity and its effect on change detection (Awh, Barton, & Vogel, 2007).

4.1.3 Abstract Shapes

Past work suggests that association-free shapes are extremely difficult to come by. Specifically, based on a spoken association protocol similar to the one used here (see below), out of fifty participants' responses, not a single randomly-generated polygon out of 180 tested elicited zero associations, with the least-meaningful shape prompting associations in 20% of respondents (Vanderplas & Garvin, 1959, p. 148). In more recent work, relatively meaningless shapes were chosen from the Vanderplas and Garvin list which had elicited associations from 30% or less of their participants (Postle et al., 2005, p. 206). All of the shapes tested in this dissertation were ones that can easily be created on common software tools such as Power Point that are used in creating slides or other materials for instruction.

To ensure comparability of results between the two image sets, the colored shapes for one group of participants was used (grayed out by setting color saturation to -100% again) as the grayscale shapes for the other. In order for these shapes to serve viable visual cues in centered-probe conditions, they needed to be *different* shapes rather than different-colored instances of the same shape.

4.1.4 Meaningfulness Analysis

Association Counts

One simple approach to quantifying image meaningfulness is to count the number of associations, whether words or phrases, provided by observers for each image.

Variety of Associations

The other approach used here to compare image meaningfulness using written associations considered whether participants tended to put the same associations down for images within a class (e.g. grayscale abstract shapes) or if all the image associations provided were unique. To calculate *unique* word associations for each image class, the associations provided by each participant were grouped by image class and repeated items were removed. For example, if a participant wrote “shape” down as an association for four of the grayscale shapes, it would be counted as only one unique association for that class. This measure is an adaptation of the type/token ratios used to assess children’s vocabulary development mentioned in Chapter 2.

4.2 Experiment 8: Revised Crossmodal Change Detection Method

4.2.1 Adding in Visual Search

Will per-item search cost for the images involved in novel crossmodal associations predict the working memory load of the visual-auditory pairings? Answering this question requires the addition of visual search trials into the previous method; exact procedural details are included below.

To keep a focus on participants’ overall change-detection capacity across the four image classes, leaving questions of mnemonic strategy for later investigation, articulatory suppression was not used in the final experiment.

The other change to the basic crossmodal change-detection method was a greater use of participant training in both search and change-detection procedures. Images used for

training consisted of the six Rossion & Pourtois drawings (2004) matching the (now-licensed) training images used in (Alvarez & Cavanagh, 2004).

4.2.2 Centered-Test Crossmodal Change Detection

Following Experiments 3 and 6, all test objects presented to participants were placed in the center of the screen, rather than in their original location.

4.2.3 Stimuli and Apparatus

Images

Appendix A.2 presents the training images and the two parallel sets of images used for the final experiment. These images were tested in a pilot study (initially called Experiment 7) involving 26 UNH undergraduates who tested the proposed meaningfulness measures. Their responses were also used to screen out any images with very strong or anomalous associations.

Sounds

Sounds consisted of 380 to 400ms excerpts of recorded animal sounds (dog, cat, cow, and bird), all tested in the development of the change-detection method.

Apparatus

Stimuli and survey questions were presented on Macintosh iBook G4s with 14-inch liquid-crystal screens, using custom software written in PEBL (Mueller, 2006). For participants seated with their eyes 57cm from the screen, a degree of visual angle was determined to be equivalent to 35.8 pixels on these screens. A 57cm string was attached to each monitor so that participants could easily check their (completely unrestrained) head position between testing blocks. The experiment program provides reminders to move around and relax briefly in those times.

Cubicles and small rooms were used to allow each participant to do the procedures individually, with an experimenter within earshot but not looking at the screen. Sounds were presented using adjustable headphones (Philips SHP2500).

4.2.4 Procedure

Participants performed two tasks in each condition, the visual search task used to estimate image complexity and the core crossmodal change-detection task, then repeated their condition sequence in reverse (except for the training condition). Following (Alvarez & Cavanagh, 2004), condition orders were fixed, with half of the participants using this sequence: training, training, grayscale shapes, color photos, colored shapes, grayscale drawings, grayscale drawings, colored shapes, color photos, grayscale shapes. The block sequence for the other half was: training, training, grayscale drawings, colored shapes, color photos, grayscale shapes, grayscale shapes, color photos, colored shapes, grayscale drawings.

Procedure for Visual Search

Participants searched for a target object in an array of objects taken from the same group of six images. The target was first presented for 500ms, then after a 900ms blank pause, an array of 4, 8, or 12 objects was presented; participants indicated whether or not the target was present by key press. (See Appendix A.4.1 for a listing of the timing sequence.) Distractors were randomly chosen to ensure broad coverage and varied groupings among the images, while presenting as many different objects in each search array as possible. These images were presented in pseudorandom positions on a 5×4 grid of $5^\circ \times 5^\circ$ squares, randomly jittered by up to 1° horizontally and vertically, directly following the Alvarez and Cavanagh (2004) procedure. Similarly, each stimulus set was tested separately; the training images were used for two sets of 12 trials each, while testing blocks had 78 search trials. (See Appendix A.4.3 for an overview of the experiment block sequence.)

Procedure for Crossmodal Change Detection

Based on the original crossmodal change detection procedure, this task presented two or three image-sound pairs in rapid sequence, followed by a one-second blank pause and a test image-sound pair. (See Appendix A.4.1 for exact timing sequence.) Participants indicated by keypress whether the pair was present in the original array—with a key labeled “no change”—or whether the image-sound pairing has changed (key labeled “change”).

For their training, participants went through two sets of eight trials of crossmodal change detection. Each condition-specific testing block consisted of 36 trials, half with two-item sequences and half with three-item, balanced between changed and unchanged trials. Images and their associated sounds were chosen randomly for each trial. These training and test blocks included far fewer trials than in the precedent (Alvarez & Cavanagh, 2004), which tested not two but eight change-detection array sizes: 1, 3, 5, 7, 9, 11, and 13 items. Not only does the audiovisual task, thanks to its serial presentation, require more time per trial than the visual-only precedent, but past results with the method also suggest that array sizes above 3 would lead to a major drop in performance. Blocks of 36 trials here provided similar representation for each stimulus in the condition’s stimulus set as in the precedent and a similar number of trials per array size. Participants typically spent 90 minutes performing these alternating trial blocks, and many expressed fatigue following their debriefing, supporting the choice to limit the duration of the experiment. Using fewer array sizes may reduce the statistical strength of the comparison between interpolated estimate of a 75%-capacity level per stimulus class to image search times, following Alvarez and Cavanagh’s (2004) precedent, but in their case much of the data from the far ends of their range of array sizes had to be discarded.

4.2.5 Results: Comparison to method development

Corrected accuracy for crossmodal change detection was greater for the color than the grayscale images, $F(1, 130) = 6.86, MSE = .029, p < .01$ and unsurprisingly greater when participants encountered two rather than three visual-auditory pairings to remem-

	<i>Representational Image</i>	<i>Abstract Shape</i>
<i>Higher Detail—Color</i>	70.0%	72.5%
<i>Lower Detail—Gray</i>	64.9%	64.2%

Table 4.2: Corrected accuracy scores for crossmodal working memory by image type. These scores are obtained using $p(Hit) - p(FA)$.

	<i>Representational Image</i>	<i>Abstract Shape</i>
<i>Higher Detail—Color</i>	55.5%–56.3%	57.1%–57.8%
<i>Lower Detail—Gray</i>	43.1%–52.2%	42.2%–53.2%

Table 4.3: Corrected accuracy scores for crossmodal working memory when encountering three-item memory arrays. Accuracy rate pairs above show the differences between the two image sets, listed with the East score followed by the West score.

ber, $F(1, 130) = 150.51, MSE = .029, p < .001$. Mean accuracy by condition for three-item arrays (for better comparison to Experiments 1–6) is presented in Table 4.2

For clearer comparison to the results of the method-development phase of this research, Figure 4-2 displays corrected accuracy rates only for memory arrays with three items; group means split by image set are presented in Table 4.3.

Unlike the first six experiments of this dissertation, this experiment had two (counter-balanced) between-subject differences: which of the two parallel image sets the participant viewed, and which condition order he or she experienced. Contrasting these two factors was added to the experimental design to strengthen any possible conclusions by increasing the generalizability of the results. Indeed, in spite of efforts to keep the parallel image sets as equivalent as possible, change-detection accuracy was higher for one (referred to in participant IDs as “W” or “West”) set than the other (see means displayed in Table 4.3). At the same time, the change-detection task involved showed high individual variation,

$F(21, 130) = 4.15, MSE = .029, p < .001$ Since this variation could obscure any patterns arising from the image manipulations discussed above, removing individual performance variability from the analysis of variation associated with experimental condition was employed as an additional analysis.

Concerns about muddying analyses of within-subject variation with high inter-subject variation have been raised over many years, with particular concern that graphs which include the full inter-subject variability might obscure detection of the within-subject patterns (Masson & Loftus, 2003; Loftus & Masson, 1994). Noting the drawbacks—including masking of between-condition variability differences—of Loftus and Masson’s particular approach, Cousineau proposed a linear norming of participant group means with relation to their individual mean (Cousineau, 2005). For the present analysis, an approach similar to Cousineau’s was taken by simply norming each participant’s change-detection scores with a z transform. In other words, one person’s accuracy on grayscale drawings was recorded as the number of standard deviations that raw accuracy lay from the person’s mean accuracy over all conditions.

Using Cousineau’s approach would allow an examination of differences in *magnitude* of any image-type-based effects between the parallel sets; in this case, the z transform—which will hide such differences—was chosen to first establish whether the *nature* of such effects were consistent or variable using different images. In this case, the transformed data also only showed differences according to detail, with crossmodal color stimuli better remembered than grayscale ones.

Splitting Accuracy: Signal Detection Theory

Up to this point, accuracy has been treated as a unitary measure; many years of past research have been devoted, however, to statistics which can divide response accuracy into two components, a standardized accuracy measure and the other a measure of *bias*, or how frequently observers report detecting whatever they are looking for. Signal Detection Theory (Pollack, 1964) has been an extremely influential approach to the study of perception,

and one measure from that literature which has been applied to change detection is A' (Xu, 2002; Grier, 1971). A' has been shown to be more accurate than the classic d' measure in some circumstances (Donaldson, 1993), and like the data Xu obtained, these data contain error-free trial blocks, which are better handled by A' (Xu, 2002). That precedent omits the accompanying bias measure, however, and computational error and misinterpretation of the A' statistic have since been rectified with the new measure A (Zhang & Mueller, 2005). Results from the first four experiments were evaluated according to A' with almost no change in significance of results—the only change was that part of the difference between performance on abstract tones versus animal sounds lay in a bias and not entirely in an overall accuracy difference. The formula for each measure (A and b) is shown in Appendix B.2.

Looking only at variability of A in Experiment 8, the same predictors are associated with significant effects observed in the analysis of corrected accuracy show up as important. Specifically, A was higher for the color images compared to the grayscale images, $F(1, 168) = 5.54, MSE = .0056, p < .05$ and having two rather than three visual-auditory pairings to remember was also significant, $F(1, 168) = 110.01, MSE = .0056, p < .001$. Which image set the participant encountered also played a role, $F(1, 168) = 5.53, MSE = .0056, p < .05$. The overall distribution is shown in Figure 4-2 on p. 58. None of these factors, on the other hand, showed significant ties to bias (or b), which responded only to the difference between representational images and abstract shapes, $F(1, 168) = 8.93, MSE = .285, p < .01$ as shown in Figure 4-2. Response bias is the only factor studied which revealed significant differences in crossmodal change detection according to the meaningfulness of the associated images. The lower bias scores for more meaningful shapes reflect participants' greater tendency to detect a change (both when a change had occurred and when none had) compared to their tendency to report detecting a change on abstract-shape trials. Further implications of this bias difference will be discussed in the concluding chapter.

<i>Detail Level</i>	<i>Repres. Image</i>	<i>Abstract Shape</i>
Color	8.02	7.36
Grayscale	16.42	18.97

Table 4.4: Mean search-time estimates in milliseconds according to image type.

4.2.6 Results: Search speed as WM capacity predictor

The alternate model introduced in this final experiment—which necessitated the addition of visual search tasks—was taken from Alvarez and Cavanagh (2004). This model predicts that the search speed cost (per added item in a visual array) will correlate closely with an estimate of the working memory load imposed by that type of visual image. Taking this model into the crossmodal realm requires applying the same visual-search metric to the images used for visual-auditory crossmodal associations, holding auditory featural complexity and other characteristics constant across conditions.

Visual search results

Using data from target-present trials, a search-time estimate was obtained by performing linear regressions for each participant on search reaction times as a function of the number of items in the search array. This estimate of the number of additional milliseconds required to find a target given an added image in the search array did vary according to condition, as shown in the following means table. These condition-based differences were significant, and analyzing the detail and meaningfulness factors separately showed that for search speed, only detail made a difference, $F(3, 83) = 14.63$, $MSE = 149.3$, $p < .001$. ($\eta^2 = .15$) See Figure 4-3 for a graph of these search-time estimates.

Comparing WM to visual search results

What is interesting to note from these results before comparing them to working memory capacity estimates is the lack of a significant difference between color images and colored shapes, and the parallel lack of difference between grayscale images. At the same time, within representational and abstract images, the more highly-featured items were found *faster* than the comparatively less detailed ones, even though the grayscale images have by far more individual low-level features than the colored shapes, if all such features are counted equally. This result is particularly surprising when compared to multiple object tracking research, where distinctive and meaningful images posed considerably more cognitive processing load than simpler ones (Horowitz et al., 2007).

Furthermore, the participants' average search times were all under 20 ms/item, comparable only to the two fastest image classes—colored squares and letters—tested by Alvarez and Cavanagh (2004).

Estimating the 75%-correct threshold

In order to complete this comparison of crossmodal to visual working memory load models, an estimate needs to be made of the change-detection array size (that is, the number of objects on the screen when starting a new change-detection trial) where observers' raw accuracy scores (average of hits and correct rejections) would be 75%. Getting from this estimate of WM *capacity* (as number of objects) to a measurement of working memory *load* requires finding the inverse of the capacity estimate, following Alvarez and Cavanagh (2004). Table 4.5 and Figure 4-3) show the working-memory load estimates arising from applying this procedure to Experiment 8's results.

No consistent main effects were found in an analysis of variance for this working memory capacity measure. Comparing participants' individual capacity estimates to the search times for each image type (see Figure 4-4 on p. 60) did not even trend towards a linear relationship, $p > .5$, although those results also suggest that screening out negative WM-load values holds some appeal. Taking group averages by condition for the search-time cost and the working

	<i>Representational Image</i>	<i>Abstract Shape</i>
<i>Higher Detail—Color</i>	0.212	0.202
<i>Lower Detail—Gray</i>	0.285	0.339

Table 4.5: Mean working memory load of novel visual-auditory pairings according to image type tested in Experiment 8. As explained in the text, this load figure is estimated by taking the inverse of the number of memory-array objects where the participant would achieve 75% accuracy.

memory load estimates, however, as done in the visual precedent (Alvarez & Cavanagh, 2004, p. 108), did reveal a linear relationship, $r^2 = .971$, $F(1, 2) = 66.99$, $p = 0.015$. The r^2 term, suggesting 97.1% of variance is explained, is comparable to that obtained in the prior visual research (99.2%, $p < .01$) (Alvarez & Cavanagh, 2004, p. 108). Figure 4-5 presents the points representing these averages for the four conditions. However, while in that precedent four out of their six participants individually showed a significant linear relationship based on these calculations, in this case only two did out of 22, with a third participant's model on the margin of significance ($p = .056$). This looser fit to the model tying together image-type search cost and working memory load could stem from multiple sources, including greater individual variability in performance of *crossmodal* than purely visual change detection, poorer estimation of participants' working memory capacity due to the reduced range of change-detection array sizes, and qualitative differences in crossmodal recall.

4.2.7 Comparison to Individual Association Variety

What if the inclusion of such different kinds of images requires a change in the model to reflect those qualitative differences? Also, how can we be sure that the word association variety that showed clear differences in the pilot participants' associations with these image types holds true for those people going through Experiment 8? Since all participants in the latter experiment were also asked, subsequent to the blocks of visual search and change

	<i>Representational Image</i>	<i>Abstract Shape</i>
<i>Higher Detail—Color</i>	4.50	3.77
<i>Lower Detail—Gray</i>	4.21	3.57

Table 4.6: Mean number of associations written down per image by type of image.

detection, to provide associations related to the images used in their image set, Experiment 8 provided an opportunity to examine the role of association-based meaningfulness in crossmodal change detection.

Participants in Experiment 8 did not provide as many associations per image (see Table 4.6) as were elicited in the pilot. In the earlier study, participants were writing by hand on a printed sheet with ten lines available per trial; in Experiment 8, participants were typing words into a text box on screen after nearly 90 minutes' effort on the main tasks. Once again, the number of associations for representational images was significantly higher, $F(1, 513) = 38.13, MSE = 1.69, p < .001$, and in this case, once individual differences were accounted for, a significant advantage for the full-color images in each category was observed, $F(1, 513) = 5.16, MSE = 1.69, p < .05$. This difference between grayscale and full-color images, as well as the main effect of representational images versus abstract shapes, held true also in the measure of variety of associations (see Figure 4-1 on p. 57). Participants provided one more unique association per representational image than they did for an abstract shape, on average—a significant difference, $F(1, 66) = 38.32, MSE = .622, p < .001$. And, unlike the results from Experiment 7, they provided an average of a half word (.55) more for a color than for a grayscale image of the same type—also significant, $F(1, 66) = 11.08, MSE = .622, p < .01$.

Adding these unique word-association counts into the model of the relation between search did not change the overall lack of a significant linear relationship between those measures; three participants (two of those mentioned before) demonstrated a significant linear relationship between visual search and working memory. However, adding the averaged

word-association variety scores into the model relating averaged search-speed and working memory load reduced the explanatory power of that model ($p > .15$), suggesting that search-speed itself does capture the informational load of both more- and less-meaningful objects comparably. Further considerations of this application of the search-time prediction model to visual-auditory working memory and exploration of the impact of stimulus meaningfulness follow in the final chapter.

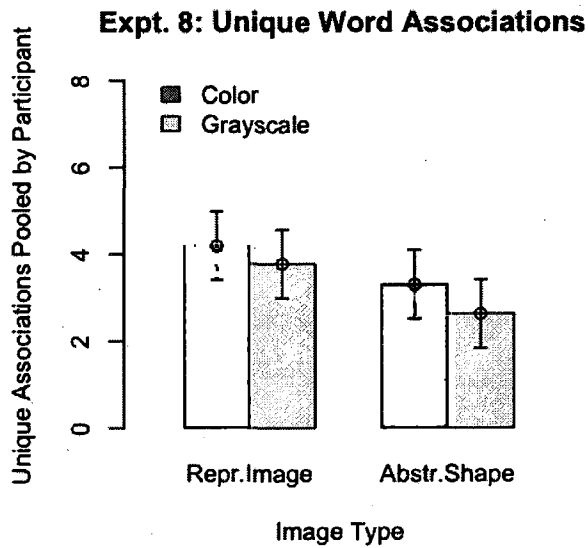
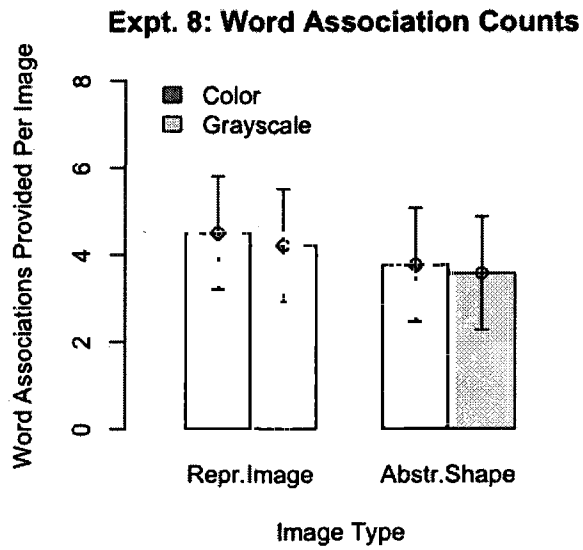


Figure 4-1: Average number of total and unique associations provided by each participant per image in a given category, crossing image type (representational or abstract) with detail level (color or grayscale). Error bars indicate residual standard error.

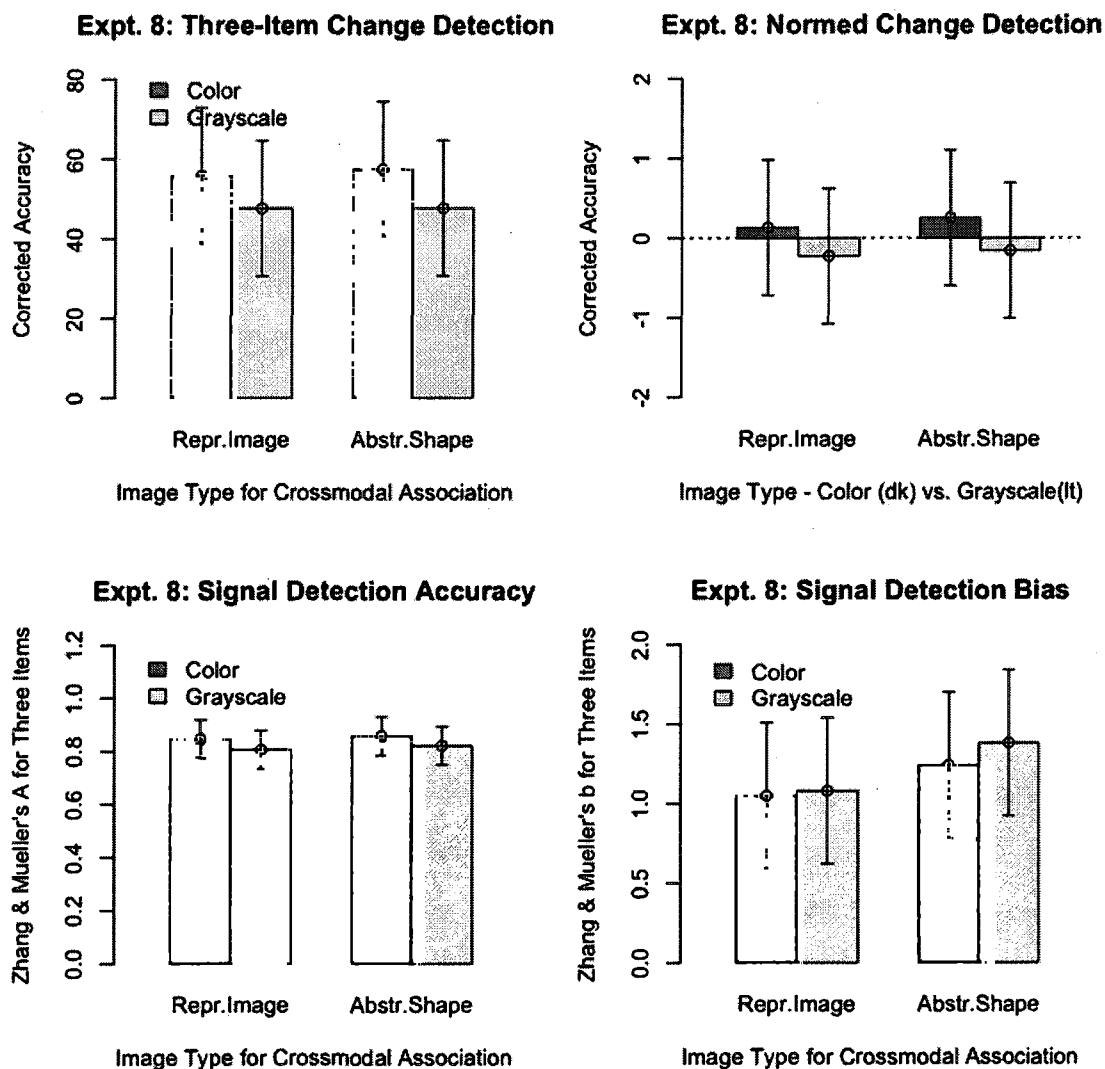
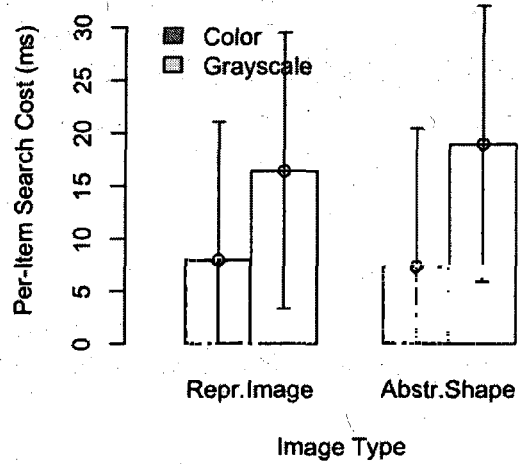


Figure 4-2: Crossmodal change-detection accuracy by several measures, using 3-item arrays for comparison to earlier experiments: Corrected accuracy ($p(Hit) - p(FA)$), participant z scores by condition, and signal detection accuracy along with bias. Error bars indicate residual standard error.

Expt. 8: Search Time Cost



Expt. 8: Estimated WM Load

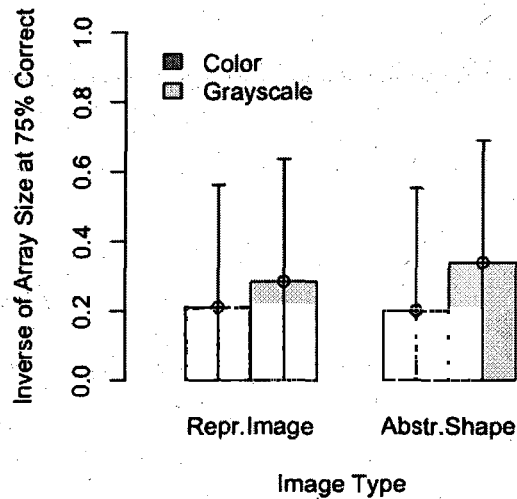


Figure 4-3: Experiment 8: Search-time estimates (in ms) for each image type and an estimate of the crossmodal working memory load for associating each image type with an animal sound. Load estimates are based on the inverse of the number of items in memory arrays for which participants show 75% raw change-detection accuracy.

Overall [Lack of] Relation Btw WM Load and Search Time

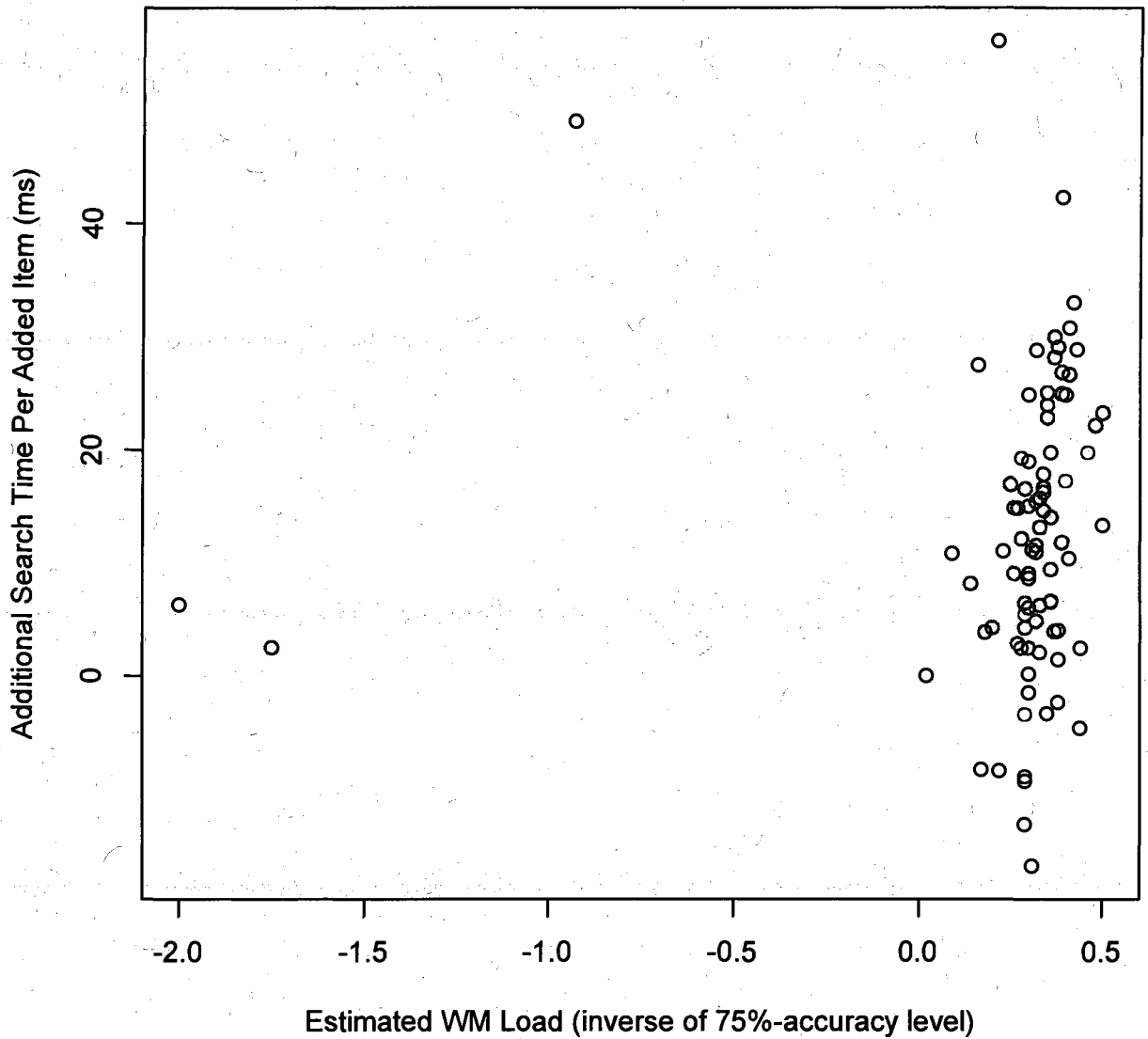


Figure 4-4: Search-time cost per image type for each participant plotted as a function of estimated working memory load of that image type. The latter estimate was calculated as the inverse of the number of objects affording the participant 75% accuracy.

Averaged Comparison of WM Load and Search Cost

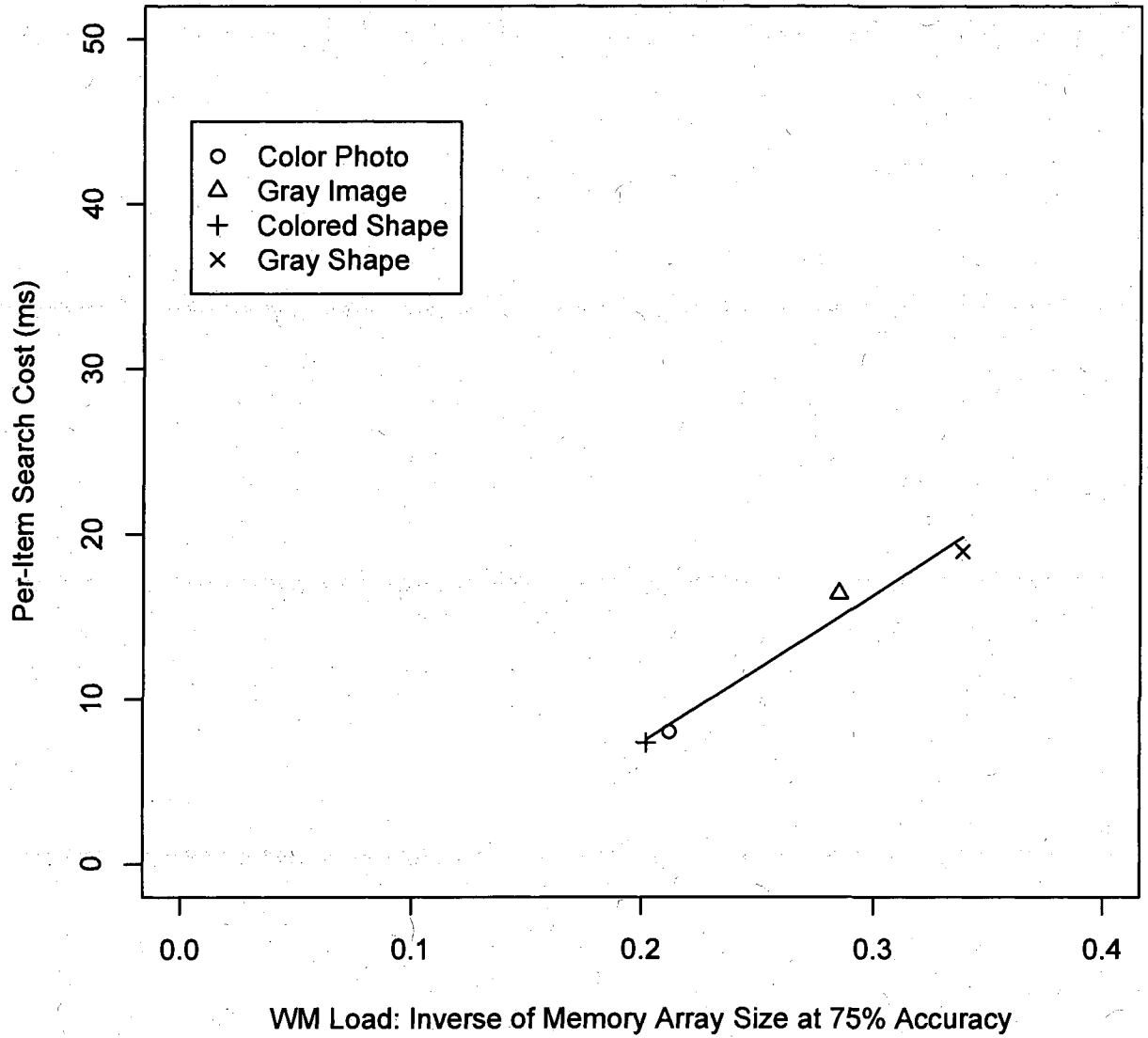


Figure 4-5: Average search-time cost per image type plotted as a function of estimated working memory load of that image type.

CHAPTER 5

FURTHER DISCUSSION AND CONCLUSIONS

The work described here pioneered a novel measure of crossmodal working memory capacity to shed light on the role of greater featural detail and stimulus meaningfulness in observers' capacity to correctly identify a changed feature binding taken from a crossmodal memory array. The experiments which served to develop this method showed that crossmodal objects' working memory load could be greater than that of similarly multifeatured visual objects. More-complex sounds afforded significantly better change detection accuracy, as did more-complex images, although the latter effect was only demonstrated for trials where the test object was placed in a novel location. In the final comparison, stimulus meaningfulness as measured by participant-provided verbal associations had only a subtle effect on crossmodal change detections—lowering decision bias—, while more-detailed full-color images very clearly improved accuracy compared to grayscale ones. These results support a nuanced view of the role of perceptual detail in working memory, as under most of the circumstances tested it provided a strong benefit. This dissertation research also raises new questions, as discussed below.

5.1 Working memory impact of adding detail crossmodally

5.1.1 Modality itself

Initial tests of this new crossmodal working memory measure suggested that adding featural detail in an alternate modality does not necessarily reduce, and in some cases may add to, the working memory load of a crossmodal object, as compared to a unimodal object with comparable amounts of featural detail. This result, based directly on the results of

Experiment 1 and indirectly on comparisons between the crossmodal results here and prior visual-only change detection rates, calls into question the claims that changes in effective working memory capacity are driving Mayer, Sweller, and colleagues' Modality Principle (Low & Sweller, 2005). This principle arises from the demonstrated benefits to learners of taking in words and images from an instructional system which specifically does not give them control over the pace of information delivery (Ginns, 2005). The limited consideration given to the effort, time, and processing involved in reading itself, combined with the present results, suggest caution in interpreting their findings as an across-the-board benefit for presenting added information in an alternate modality. Similar qualifications have been needed for instructional initiatives such as interactive learning, which was found to have a range of effects, not always optimal, in an extensive evaluation at the San Francisco Exploratorium (S. Allen, 2004). Beyond the introduction of a novel method for assessing crossmodal working memory load, what stands out about the present research is that highly meaningful images and sounds—in other words, stimuli greatly “burdened by the complexities of extra-laboratory associations” (Sekuler & Kahana, 2007)—, which might be used in a classroom or an interactive museum exhibit, displayed working memory dynamics similar to the more controlled stimuli used by those authors or by Olsson and Poom(2005).

5.1.2 We Do Not Just Count Features

Like the simple shapes in that earlier research (Olsson & Poom, 2005), short sound segments with much greater featural complexity and much greater categorical distinctiveness provided significantly better short-term recall than did abstract tones. Image differences proved hard to categorize, though they never supported a significant featural-detail penalty such as that found in multiple object tracking (Horowitz et al., 2007) or in some visual working memory studies (Treisman, 2006). These image differences only showed a consistent and consistently significant pattern of effects when test probes were presented in the center instead of in their original location.

To summarize, with location cues, differences of kind (representational image versus

abstract shape) or featural detail (color versus grayscale) did not display consistent patterns of influence on participants' change-detection success. When, on the other hand, crossmodal test objects were presented in an unused location—the center of the screen, never chosen as a location in the memory array—more-detailed images gave better performance than simpler ones. In the one direct comparison of same-location and new-locations test conditions, participants had significantly better change detection in the first (same-location) condition.

As mentioned above, all of these cases differ from what might be predicted from some earlier visual working memory research, where each additional feature to store (e.g. color, stripes) reduces available working memory capacity. Since the advantage accompanying location cues fits strongly with our general understanding of context and memory, while the effects of image detail and crossmodal working memory fall in more-disputed and less-charted territory, the second phase of the dissertation concentrated on two factors—meaningfulness and detail—which might be driving the less-predictable change detection results.

The experimental findings presented here are all compatible with information-theoretic approaches to visual object identification, despite some differences within that group of approaches. If meaningful images' working memory load is indeed based on feature counts, but the features being counted are fewer, higher-level features than line orientation, luminance and so on (Ullman et al., 2002; Borenstein & Ullman, 2002), their load would indeed be lower than abstract and unfamiliar objects which cannot be recognized with such higher-level features. If our perception and cognition makes use of association-based working memory encoding, where more-familiar conjunctions of features could be stored more efficiently than less-familiar ones (Brady et al., 2008), then the feature combinations in highly recognizable objects might take up fewer encoding resources than the less-familiar feature combinations. Since the complex-sound benefit in Experiments 2 and 5 operated independently of the image-type dynamics, we can contribute at least one new observation to those model debates: The entire crossmodal object does not need to be familiar and meaningful as a whole—neither a barking dot nor a barking apple is likely to be familiar—to benefit from the working memory advantages coming from one of its component features.

5.1.3 Location, location, location

Within visual working memory research, some features have demonstrated greater importance than others, with location ranking among the more important ones. The interaction observed in these experiments between test-probe location and the importance of associated image differences highlights an interesting difference between these results, where location cues led to better change detection (contrasting trials such as Figure 3-1 on p. 24 with ones like Figure 3-3 on p. 33, results from Experiment 3 shown in Figure 3-2) and earlier visual WM work, which found that location information did not help observers recall bindings correctly given a single test probe (Treisman & Zhang, 2006). While further work would be necessary to identify the reasons for this difference, a promising starting point would be to examine the ramifications of using rapid serial presentation instead of a fully simultaneous memory array. If the presentation method does explain the difference in impact of location cues, that would further inform questions of how much of a particular scene—even a scene on a computer display—is stored in visual working memory. This suggests that object features, possibly primarily their edge or boundary features (Alvarez & Cavanagh, 2008), are stored in relation to other feature boundaries caught in the same exact glimpse.

Other work from Alvarez and colleagues (Alvarez & Oliva, 2007) has supported the importance of relative geometry—termed global layout—in visual short-term memory. The advantage shown in these results for same-location test probes suggests that the impact of global layout is sensitive to the difference between multi-object displays where each object is presented separately for 500ms in a series with the rest and displays which present all the objects at once. Earlier work on visual repetition priming has suggested that rather than relying on basic feature priming, this visual-search effect utilizes episodic memory on a very short timescale (Huang, Holcombe, & Pashler, 2004). The present data supporting a location advantage for recall of visual-auditory associations are not geared towards making a precise distinction along the same lines, but the Huang et al. results support the possibility that participants can utilize the rapid separate views offered by the the serially presented memory arrays to help them remember crossmodal associations. At a more general level,

this location advantage adds credence to the claim that the images and (unlocalized) sounds are indeed bound together as objects in the observers' perception.

5.1.4 Modality itself: Visual and auditory differences

This interaction with location was not observed when comparing simpler abstract tones with more-complex animal sounds and spoken words: whatever the location cues available, participants had better change-detection recall with animal sounds compared to abstract tones, and spoken-word auditory features were better still. (See results from Experiments 2, 4, and 5 in Figure 3-2)

These findings support Olsson and Poom's (2005) contention that visual working memory relies on categorical distinctions, and again fail to support models where counting low-level features provides a linear prediction of working memory load. Using representational images complicates any attempts to make image or sound sets which differ by only one feature yet still represent easily-differentiated objects.

5.1.5 Intra-set distinctiveness and working memory

The visual WM precedent which handled a variety of images comparable to those used here and which paved the way for alternatives to feature-counting memory-load models is (Alvarez & Cavanagh, 2004). Looking at the averaged performance on both visual search and change detection in Experiment 8, support for their behaviorally-scaled model of working memory load can be observed.

Considering the marked difference in the percent of participants whose search times and WM loads were or were not strongly related between Alvarez and Cavanagh's (2004) study and Experiment 8, though, it is worth looking at the range of variation to be explained. Compared to that precedent, the images used in Experiment 8 were very close together and very fast in terms of search-time cost per item, and moderately close together and quite low in terms of working memory load. Figure 5-1 allows a visual review of the range of results obtained in the two experiments.

Comparing the precedent's image sets to those used in Experiment 8 raises the possibility that inter-item differences within the same memory array could be a determining factor in change detection performance. Awh and colleagues were able to demonstrate that cross-set visual object changes (e.g. putting a cube in as a test probe where a Chinese character had been presented in the memory array) were detected more accurately than changes involving more similar items (Awh et al., 2007). Their method was not an exact parallel to the basic method used here, and they did not measure the impact of gradations in intra-set heterogeneity. This hypothesized role of heterogeneity needs to be evaluated alongside hypotheses about the WM impact of the characteristics of individual images. Change detection does not demand exact recall from the observer, it only demands differentiation: did you see *this* one before or *that* one? One approach might be to use computational saliency measures to assess the extent of inter-item differences going along with different combinations of images. Unfortunately, those measures do not currently characterize meaningfulness or otherwise help to distinguish, say, between a display of six upright "2"s and a display of six "2"s rotated to a horizontal position.

5.1.6 Color, WOW!

The strongest WM effects demonstrated here, after individual differences and the role of array size, stemmed from the difference between full-color and grayscale images, with the former better recalled than the latter in all centered-probe contrasts of those two image detail levels. That color should serve as a highly salient feature is not in itself surprising, given the long record of visual-search studies establishing its importance. What is most surprising about the stand-out effects of color in accurate recall of crossmodal associations shows up in comparing color's impact between the representational and abstract image sets used in Experiments 3, 6, and 8.

Simply concluding that color is much more salient than the other varied features does not fully explain the crossmodal change-detection accuracy results obtained, as color photos fared better than colored balls in Experiment 3, and because location does play such a strong

role as well. The results of these experiments do not rule out the possibility that added detail does become a working memory burden after the minimum threshold of distinguishable category membership has been reached. However, if that were the full explanation, more-detailed images should do significantly worse in the experiments with test probes presented in their original location. Instead, they showed no consistent pattern of influence on change-detection accuracy.

Looking at the results from the final experiment (see Fig 4-2), we can see that the high salience of color holds greater influence over successful crossmodal WM performance than the difference between shape cues—in spite of the demonstrated importance of boundary features (Alvarez & Cavanagh, 2008)—and the more broadly-distributed and more-numerous small differences found between each representational image. If those many tiny differences could add up to a salience difference comparable to that conveyed by color differences, grayscale drawings would have easily outperformed grayscale shapes. And although those same grayscale shapes are nowhere near the visual complexity of the cubes used by Alvarez and Cavanagh (2004) and further examined by Awh and colleagues (2007), they are significantly harder to keep in mind, at least when associated with animal sounds, than their full-color counterparts. Specifically, Experiment 8 participants' threshold for 75%-accurate change detection was five crossmodal objects for the colored shapes and three for the gray ones. Further experiments contrasting change detection with same-color and varied-color images may help clarify the role of stimulus-specific and intra-set differences on working memory performance with crossmodal associations.

5.1.7 Are individual WM capacity differences at play?

At the level of observational anecdote, some participants in Experiments 3 and 4 said they preferred the least-detailed images, and at least a few of them did show worse working memory performance for the more-detailed ones. The opposite case (reports of participants preferring color photos who indeed performed better with those images) was also true. Individual differences in working memory capacity—which on the verbal side have inspired

much applied research regarding impacts on school performance—have been used to clarify mechanisms operating within visual working memory (Vogel & Awh, 2008). These participants with opposite preferences did not show overall capacity differences, though, in that their accuracy on their respective better and worse stimulus types was comparable between photo- and location-only preferrers. Further experimentation would be needed to see whether these individual differences are reproducible; the fact that both cases can exist once again rules out a simple measure of cognitive load for crossmodal associations based on low-level featural complexity.

5.2 What does meaningfulness mean crossmodally?

The association-based measure of meaningfulness very strongly supported initial claims that the representational images used would elicit a greater number of meaningful associations than the abstract shapes. While Experiment 7's participants did not show any associational differences according to image detail (color versus grayscale), the participants in Experiment 8 did, according more associations to color than grayscale images. The significant (if slight) advantage for the color photos compared to the grayscale drawings on its own would need to be followed by further experimentation—for instance comparing all drawings or all photos—for clarification, but the fact that it was accompanied by the same pattern in the abstract shapes suggests that comparatively vivid images may bring more associations to mind than duller ones.

Image meaningfulness was associated with differences in response bias but not with overall accuracy of crossmodal change detection, supporting recent claims that well-known objects are processed in different ways within visual cognition than novel feature combinations (Hommel & Colzato, 2009; VanRullen, 2009). This bias difference is similar to what Visscher and colleagues found in their study of recognition memory for synthesized, relatively meaningless sounds—specifically, detection bias was lower (participants more often said “yes”) for more-different stimuli and higher (participants were less likely to say “yes”) for stimuli only one just noticeable difference apart (Visscher, Kaplan, Kahana, & Sekuler,

2007). This heterogeneity-based difference in not accuracy but response bias has been found by those researchers in a range of studies of visual recognition memory, as well (Sekuler & Kahana, 2007).

5.3 Future Work

5.3.1 Conceptual distance as category delineator

Several immediately feasible projects which would clarify some of the issues raised above about category distinctions and featural detail suggest themselves as extensions to this work. One would involve crossmodal change detection using image sets which are either members of the same object category or of different categories; a starting point would be to use the images which demonstrated conceptually-based differences in long-term recall (Konkle, Brady, Alvarez, & Oliva, submitted). Such an experiment could possibly determine whether conceptual differences are sufficient to make distinctions between stimuli whose perceptual characteristics only vary by degree—suggesting WM capacity might go down to one for those objects (Olsson & Poom, 2005).

Another possibility would involve contrasting working memory performance using the colored shapes from Experiment 8 and contrasting them with similarly bright and colorful shapes whose outlines are the same but whose interior characteristics differ, such as having overlaid black stripe or dot patterns. These images should be contrasted with the colored balls from the method development phase of this research and the grayscale shapes. This study would go further to align crossmodal working memory research with related work about the impact of different kinds of features on visual cognition (Alvarez & Cavanagh, 2008).

A third avenue to explore involves contrasting rapid serial presentation with simultaneous presentation, to see whether that difference is the main source of the different role played here by location for feature binding.

5.3.2 Taking another long-term look at detail

The research presented in this dissertation applied experimental techniques from the field of visual cognition to crossmodal stimuli in a manner geared towards comparisons to related findings in multimedia instruction research, summarized in Section 2.2.1 on p. 7. The following claim about working memory was of particular interest.

It may be possible to increase effective working memory capacity by presenting information in a mixed visual and auditory mode rather than a single mode (Low & Sweller, 2005, p. 148).

Experiment 1 showed a substantial advantage for unimodal change detection (see Figure 3-2). Throughout the dissertation, higher detail within either the visual or the auditory modality on its own—unlike Butcher’s (2006) result for long-term learning—showed a consistent mnemonic advantage over simpler stimuli. Taken together, these results suggest that explorations of the working memory dynamics in formal learning settings consider image and sound heterogeneity in instruction as well as sensory modality.

Since use of multimedia learning materials is likely to keep growing in the foreseeable future, another research direction suggested by the present findings is to use considerably more-meaningful and more-detailed images in research on long-term learning. Incremental steps towards that goal might involve contrasting meaningful detail—such as scaling the size of a dot representing a city according to its population size—and irrelevant detail such as what was used in these tests of novel association formation. One challenge to such extensions is that for students who are learning about, say, the structure of DNA, an image of a helix may *change* in meaningfulness over the course of a semester.

Another possibility would be to manipulate the characteristics of bullet points or other, usually-uniform shapes used in instructional materials. Would using more unique, Miro-style points with different orientations next to three concepts which have been called out in a text box help readers to keep track of those concepts with repeated use of the unique shapes? This research would require careful coordination with instructors and in-depth

subject knowledge for appropriate preparation of stimuli.

In spite of the varied challenges facing research which integrates fundamental explorations of human cognitive processing with the pressures and constraints of formal learning, this dissertation provides evidence that laboratory research can indeed change our understanding of the dynamics of real-life learning. Designers of instructional media should understand and not eschew detail, and hopefully working memory researchers will extend our understanding of how short-term cognitive processes shape long-term retention.

Comparison of Crossmodal to Visual WM and Search Speed

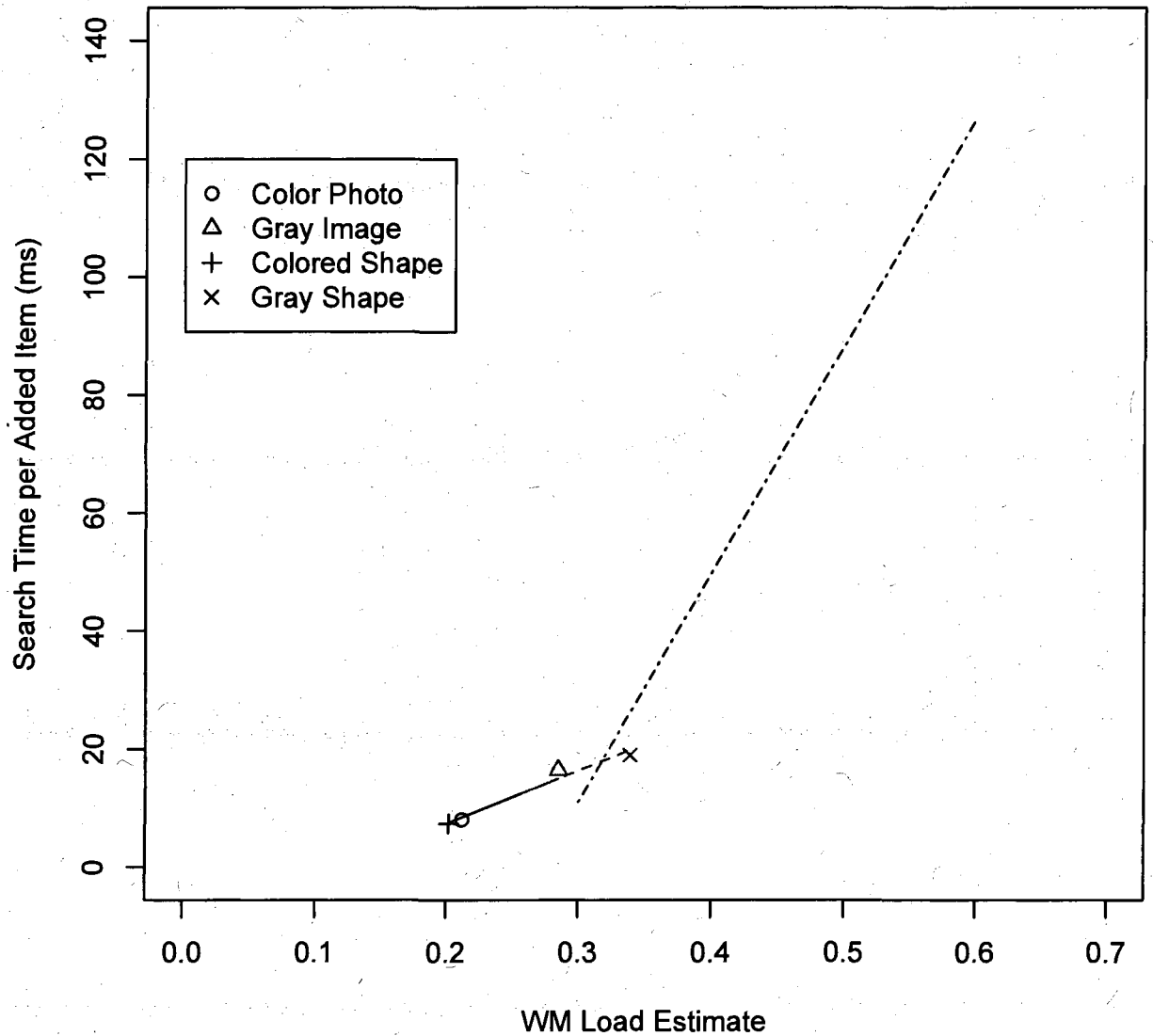


Figure 5-1: Relation between working memory load and search times for crossmodal stimuli tested here (see key) and prior visual-only stimuli (range and domain shown by dotted rectangle). WM load is based on the inverse of participants' capacity at 75% accuracy: a load of .25 indicates that they could detect changes with that accuracy for 4 items, a load of .5 reflects a capacity of 2 items, and so on.

APPENDICES

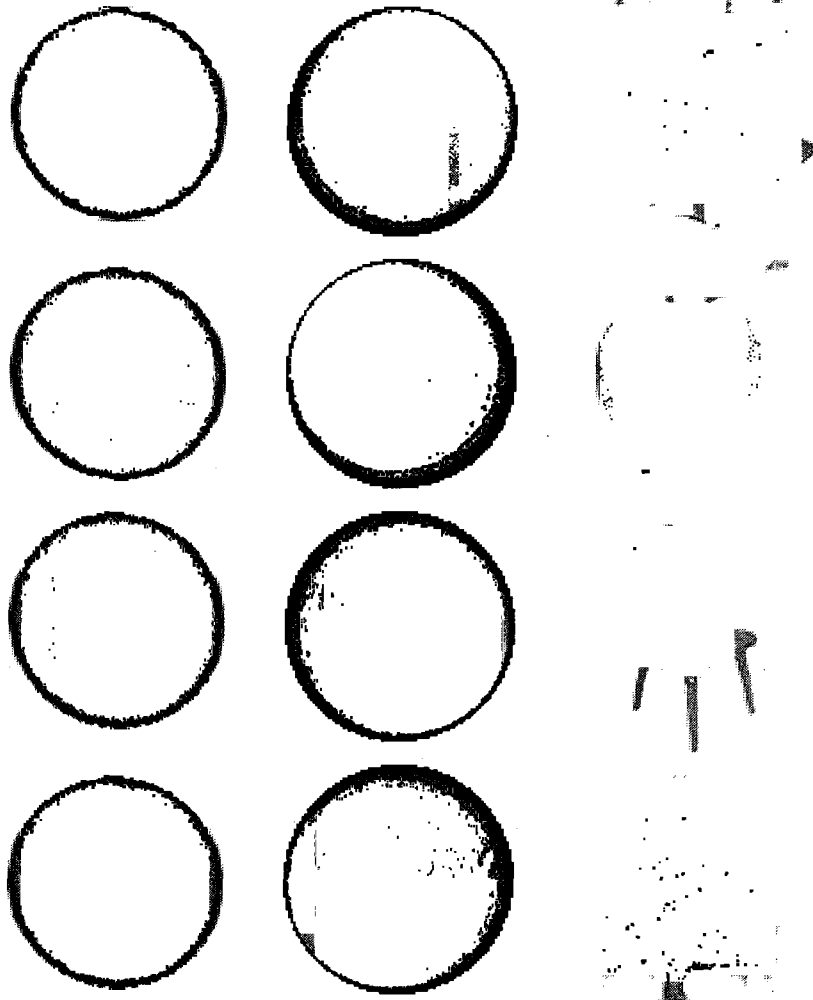
APPENDIX A

IMAGES USED

A.1 Images: Experiments 1 through 6

A.1.1 Images: Experiments 1 through 5

Experiment 1 used only colored balls, while Experiments 2-5 used all three of the image sets



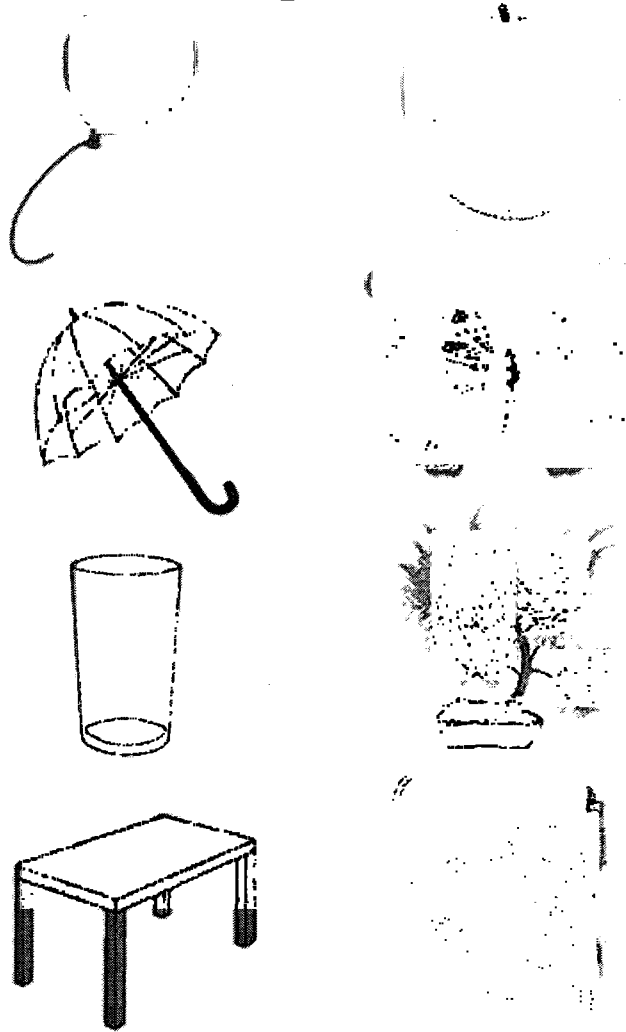
shown here.

The balls

have been re-created, due to a lack of compatibility between the experimental software and screen capture; the remaining images are those used in the experiments.

A.1.2 Images: Experiment 6

This final experiment in the method-development series contrasted the grayscale and full-color meaningful images shown below.

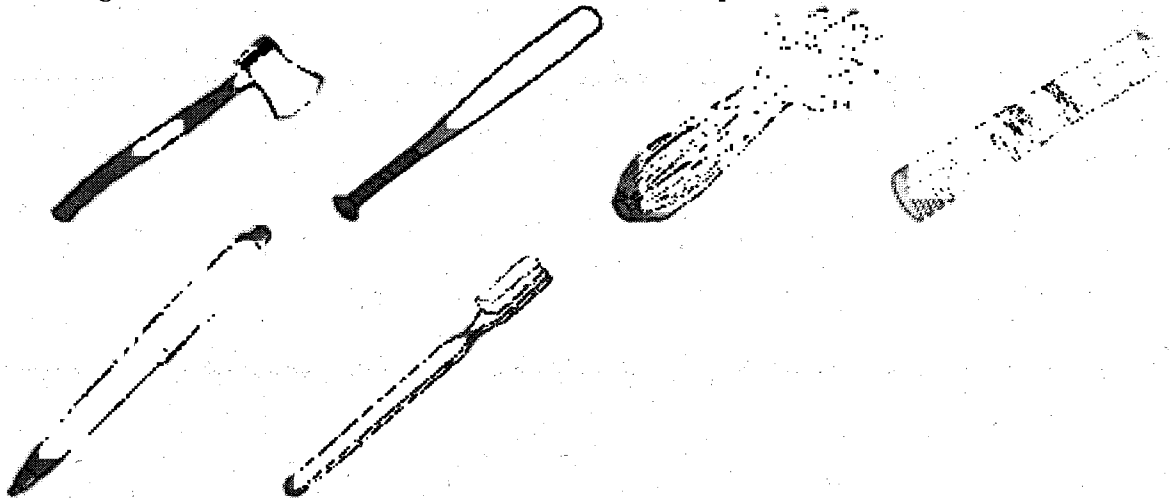


A.2 Images: Experiments 7 and 8

Experiment 7, the meaningfulness pilot, made use of all of the following images plus several more, as described in the main text. Results from Experiment 7 were used to finalize image selection for Experiment 8 and to characterize the images used.

A.2.1 Images for Training

The training images are the same as those used by Alvarez and Cavanagh (2004), except that they are shaded. As the original images used are now only available under license, the freely-available, full-color equivalents (Rossion & Pourtois, 2004) have been used, transformed by reducing color saturation to -100% with Adobe Photoshop.

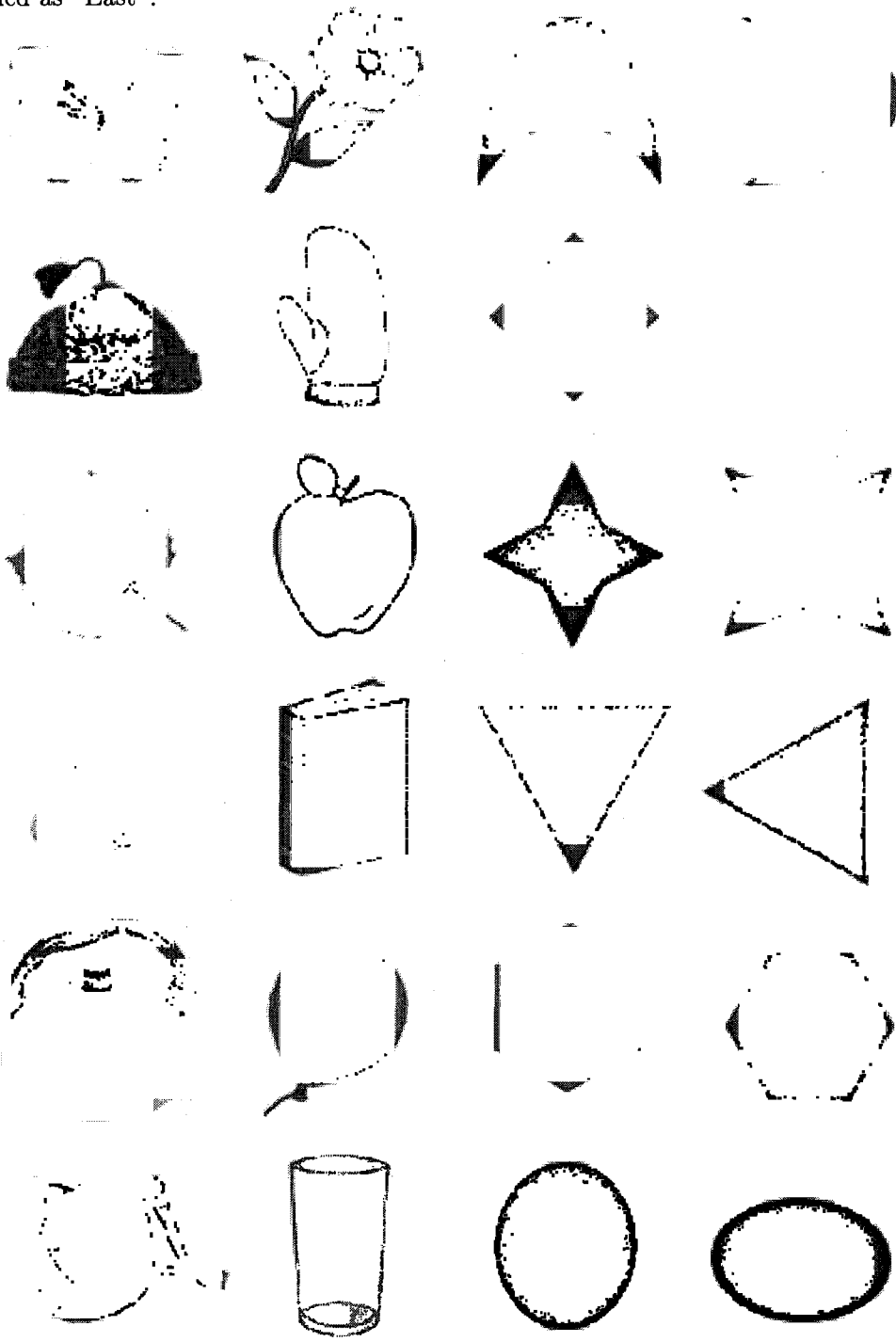


A.2.2 Images for Testing

For the two stimulus sets shown below, six images were chosen to represent each image class, all conforming to the constraints laid out in Section 4.1.

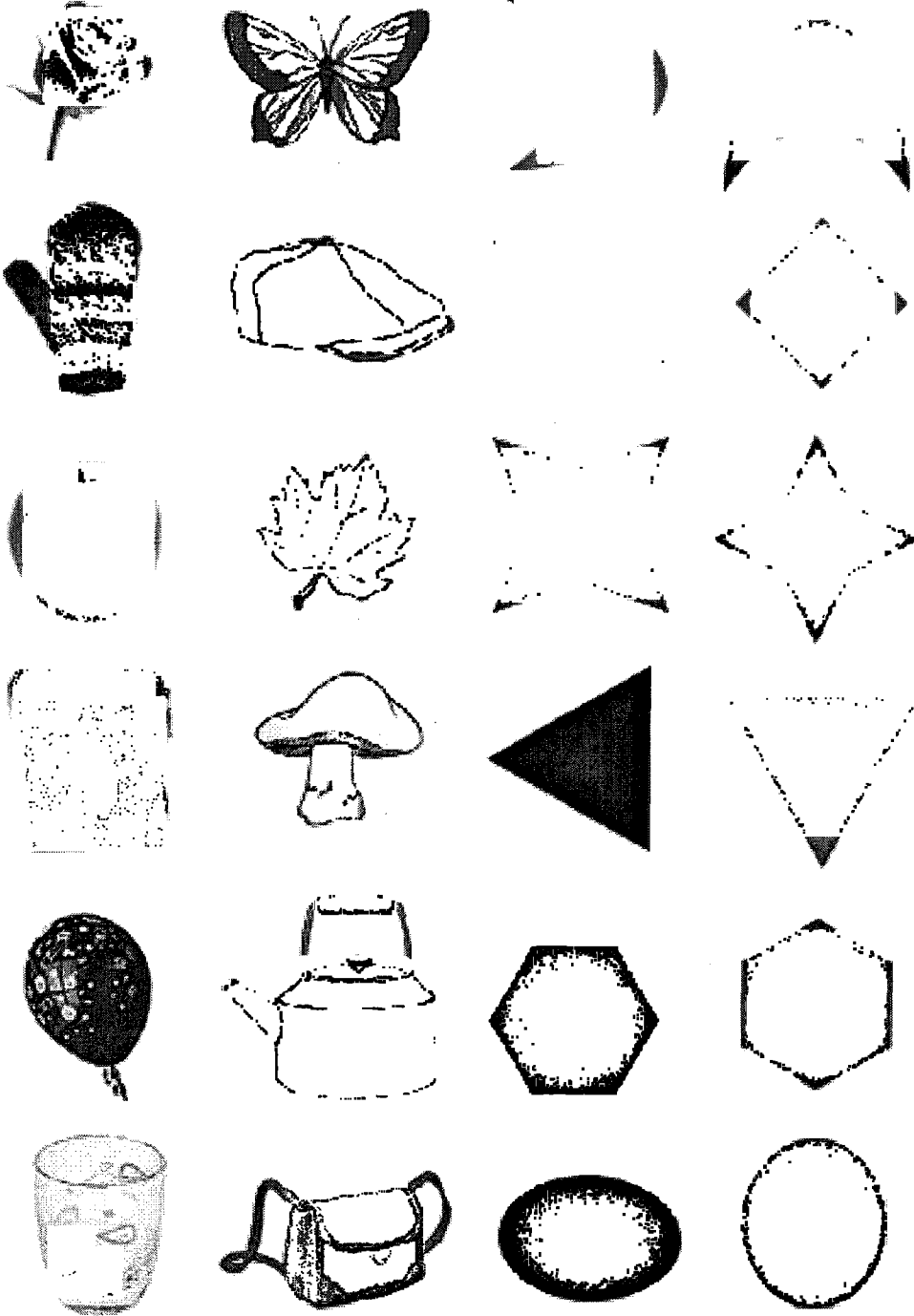
Set A

Half the participants had an image set with taller color photos and wider grayscale drawings, coded as "East".



Set B

The other participants, coded as "West", encountered the taller grayscale drawings and the wider color photos.



A.3 Procedure Specifics: Instructions

A.3.1 Experiments 1–6 Instructions

For these method-development experiments, participants received individual verbal instructions. All participants went through practice trials covering each possible experimental factor, though possibly not each combination of factors. For instance, in preparation for Experiment 2, each participant practiced with both sound types and all three image types to the point of comfort with the protocol before testing. Training did not necessarily involve all six combinations of sound and image types.

A.3.2 Experiment 8 Instructions

For the final experiment, instructions were made more uniform by including them in the experiment program. After collecting a signed consent form, the experimenter seated each participant, pointing out the labeled keys (see procedure instructions below) and the adjustment features of the chair, headphones, and laptop screen. Instructions were delivered on the screen, separated into small paragraphs which were displayed one at a time.

General instructions

The first specific instructions were as follows.

WELCOME TO THE EXPERIMENT! In Part I, you'll be searching for images and noticing changes in noise-making objects on the screen. In Part II, you'll play a word game and answer some surveys. Press any key to see more detailed instructions for Part I.

Your first main task will involve VISUAL SEARCH; the second main task is CHANGE DETECTION. On the next screen you'll get more instructions on the search task, then you'll get to practice it a bit. After that you'll get instructions and practice rounds for change detection. (Press any key to move on to the next page.)

After the practice trials, you'll alternate between those two tasks using different kinds of images. This program will remind you to stop and stretch now and again after you've completed a set of images. Ready? (Press any key to start.)

Experimental Procedure Instructions

Here are the instructions used for visual search. The designated keys were also labeled with small stickers saying "FOUND" and "NOT found".

VISUAL SEARCH: First you'll see a single object – keep it in mind, because next you're going to look for that same object. So you'll then see a bunch of similar objects appear all together. Press "P" if that single object you saw first is present in the group; press "W" if it's not. Then you'll get a new object to look for and start again. Press any key to begin.

Similarly, change-detection response keys were labeled "DIFF" and "SAME", with the following on-screen instructions.

CHANGE DETECTION: Note which objects make which sounds... there will be two or three objects in each trial, making different noises. When one of them comes back in the center of the screen, decide if it has CHANGED what it's doing compared to when you last observed it. Press the "." key if it changed its tune; if not, press "Z" for no change. (Press any key to start practicing.)

Progress Indication

Halfway through their training blocks, participants were encouraged to ask questions, as follows.

So now you've practiced both of the tasks that make up the main experiment. This would be a good time to get clarifications from the experimenter, or to tell her you've got the hang of it. Then press a key to continue.

Before starting the testing sequence, this message was displayed.

Okay, now you're getting to the main portions of the experiment – this is where we start changing up the image files. (Press any key to start.)

Participants also received reminders to stop and stretch after each experimental block, using randomly-selected and never repeating statements suggesting they—for example—touch their toes or shrug their shoulders.

Instructions for Wrap-Up

After completing all experimental trial blocks, participants viewed this message.

Whoa, all the searching and sound-tracking is done! You've been doing this for a while, so do take a minute or two off as you wish. For Part II of the experiment, you'll first provide some word associations, then answer two surveys. And Part II is quicker than Part I! When you're ready to continue, type a key.

Instructions for the word association task were as follows.

You will see a series of images. As each new one appears, write down whatever words come to mind easily. Type in a word and then hit the return key after each one; if you run out, type the word 'done' (no quotes) and hit return.

Let's practice once.

They then received instructions for the Cognitive Failures Questionnaire (Broadbent, Cooper, FitzGerald, & Parkes, 1982), with the second half taken directly from the original questionnaire.

Next you'll go through a survey that has been used in some related research on working memory and attention. Please answer as best you can without pondering the questions much. The post-experiment debriefing explains further why this may be relevant to working memory. (Press any key to go on.)

The following questions are about minor mistakes which everyone makes from time to time, but some of which happen more often than others. We want to know how often these things have happened to you in the past 6 months.

(Press any key to go on.)

The following bit of encouragement was displayed before the program asked participants to identify their dominant hand (right, left, or ambidextrous) and their gender.

Just a few more questions to help us make sure we have a good balance of participants between conditions... (Press any key to go on.)

After the entire experiment was completed, participants received these final instructions to go collect their debriefing.

Thank you for all your work!!! YOU FINISHED THIS WHOLE EXPERIMENT!! The experimenter has a debriefing form for you which says more about how your data will contribute to our research project.

A.4 Procedure Specifics: Timelines

A.4.1 Visual Search

1. Target presented at the center of the display for 500ms.
2. Blank interval of 900ms.
3. Array of 4, 8, or 12 objects (all "from the same stimulus class as the target" (Alvarez & Cavanagh, 2004, p. 107)) presented.
4. Participant indicates by keypress whether the target is present or absent.
5. Dependent variable: Search time as a function of number of items present.

A.4.2 Change Detection

1. Fixation mark appears for 650ms

2. First object and accompanying sound presented for 500ms
3. Second object and accompanying sound presented for 500ms
4. Third object and accompanying sound presented for 500ms, if used
5. Blank pause for 1,000ms
6. Test object and accompanying sound presented
7. Participant indicates whether a change was detected by key press
8. Dependent variable: Participant accuracy in detecting changes

A.4.3 Overall Structure

Participants were reminded to take a break after every block.

TRAINING		
	<i>Training Image Set</i>	
	Visual Search	12 trials
	Crossmodal Change Detection	8 trials
	<i>prompted to ask questions</i>	
	<i>Training Image Set</i>	
	Visual Search	12 trials
	Crossmodal Change Detection	8 trials
	<i>(continues on next page)</i>	

TESTING		
	<i>First Image Set</i>	
	Visual Search	78 trials
	Crossmodal Change Detection	36 trials
	<i>Second Image Set</i>	
	Visual Search	78 trials
	Crossmodal Change Detection	36 trials
	<i>Third Image Set</i>	
	Visual Search	78 trials
	Crossmodal Change Detection	36 trials
	<i>Fourth Image Set</i>	
	Visual Search	78 trials
	Crossmodal Change Detection	36 trials
	<i>Fourth Image Set</i>	
	Visual Search	78 trials
	Crossmodal Change Detection	36 trials
	<i>Third Image Set</i>	
	Visual Search	78 trials
	Crossmodal Change Detection	36 trials
	<i>Second Image Set</i>	
	Visual Search	78 trials
	Crossmodal Change Detection	36 trials
	<i>First Image Set</i>	
	Visual Search	78 trials
	Crossmodal Change Detection	36 trials
SURVEYS		

APPENDIX B

FORMULAS AND CAPACITY ESTIMATES

B.1 Estimating working memory capacity

This dissertation focuses on *relative* comparisons within a given participant's performance of accurate change detections for associations of visual and auditory features. However, researchers examining working memory have over many years looked for *absolute* measures of working memory capacity, such as the famous formulation of the “magical number seven plus or minus two” from Miller (1956) suggesting that people can keep about seven things of various sorts in mind at once.

Change detection accuracy itself has been used to provide a newer estimate of how many visual objects observers can maintain in working memory (Awh et al., 2007; Treisman, 2006; Vogel et al., 2001; Pashler, 1988). This approach was extended by Alvarez and Cavanagh (2004), who proposed using the *inverse* of a measure of object-based capacity as an indication of the working memory *load* of that type of object.

B.1.1 Application of Pashler's capacity estimator

Pashler suggested that the number of items that an observer can keep in mind (k) can be estimated from that observer's change-detection hit rate—how many times the person says the display changed when the display really had changed—and their false-alarm rate—the rate at which the person says the display changed when it had not (Pashler, 1988). Including the false alarm rate allows the measurer to account for differences in guessing rates between conditions and observers, since someone who catches every true change by simply pressing the “Changed” key for every trial is not as accurate as someone who catches most of the

true changes and almost never registers a false alarm.

Formula

Again using H for hit rate and F for false alarm rate, along with S for the size or number of items in the memory array, Pashler's (1988) formula is as follows.

$$H = \frac{k}{S} + \frac{S - k}{S} F$$

In words, this proposes that a person's hit rate will be the ratio between their working memory capacity and the number of items in each memory array, plus the product of their guessing rate and the proportion of items over their capacity included in the memory array.

That original formula can be solved for capacity (k), as shown below (Vogel et al., 2001, p. 95).

$$k = \frac{S \times (H - F)}{1 - F}$$

Results from method development

While rapid serial presentation of crossmodal objects may involve different assumptions than those underlying the Pashler (1988) estimator, applying it to the data obtained in the first six experiments suggests a capacity for somewhere between .75 and 1.75 crossmodal objects (such as a barking apple, beeping chair, or mooing balloon) being maintained at once. This is lower than the rough four-object estimate from other scientists (Awh et al., 2007; Alvarez & Cavanagh, 2004; Vogel et al., 2001; Pashler, 1988). The higher capacity demonstrated in the final experiment for color photos and varied colored shapes (almost five crossmodal objects), on the other hand, suggests that the earlier low capacity is not a direct outcome of combining an image with a sound.

Since this method *added* auditory features to simple and complex visual objects rather than *splitting* visual objects' features and delivering some featural information verbally (R. J. Allen et al., 2009), another perspective might be to treat each crossmodal object as two

objects within working memory. Even that adjustment only puts the most successful participants in the range of the three to four object capacity found for visual objects (Vogel et al., 2001). This difference suggests that binding features across different modalities does not behave in exactly the same fashion as binding features from different perceptual dimensions within one modality (such as vision).

B.2 Accuracy Measures

B.2.1 Corrected Accuracy Formula

All crossmodal change detection results have been reported using the **corrected accuracy** measure reported by Allen and colleagues (2006), whose formula is as follows.

$$\text{CorrectedAccuracy} = p(\text{Hit}) - p(\text{FalseAlarm})$$

B.2.2 Signal Detection Theory Formulae

The culminating experiment (Experiment 8) was also evaluated using Zhang and Mueller's (Zhang & Mueller, 2005) A and b , as discussed in the text (see p. 50). Using H for $p(\text{Hit})$ and F for $p(\text{FalseAlarm})$, the formulae for this measure of accuracy and bias are as follows.

$$A = \begin{cases} \frac{3}{4} + \frac{H-F}{4} - F(1-H) & \text{if } F \leq 0.5 \leq H \\ \frac{3}{4} + \frac{H-F}{4} - \frac{F}{4H} & \text{if } F \leq H < 0.5 \\ \frac{3}{4} + \frac{H-F}{4} - \frac{1-H}{4(1-F)} & \text{if } 0.5 < F \leq H \end{cases}$$

$$b = \begin{cases} \frac{5-4H}{1+4F} & \text{if } F \leq 0.5 \leq H \\ \frac{H^2+H}{H^2+F} & \text{if } F < H < 0.5 \\ \frac{(1-F)^2+(1-H)}{(1-F)^2+(1-F)} & \text{if } 0.5 < F < H \end{cases}$$

APPENDIX C

IRB APPROVALS

This dissertation research was approved by the University of New Hampshire's Institutional Review Board.

University of New Hampshire

Research Conduct and Compliance Services, Office of Sponsored Research
Service Building, 51 College Road, Durham, NH 03824-3585
Fax: 603-862-3564

01-Nov-2007

Gilman, Anne
Psychology, Conant Hall
327 Forest Park
Dover, NH 03824

IRB #: 4109

Study: Multimodal Binding Capacity in Working Memory

Approval Date: 30-Oct-2007


The Institutional Review Board for the Protection of Human Subjects in Research (IRB) has reviewed and approved the protocol for your study as Expedited as described in Title 45, Code of Federal Regulations (CFR), Part 46, Subsection 110.

Approval is granted to conduct your study as described in your protocol for one year from the approval date above. At the end of the approval period, you will be asked to submit a report with regard to the involvement of human subjects in this study. If your study is still active, you may request an extension of IRB approval.

Researchers who conduct studies involving human subjects have responsibilities as outlined in the attached document, *Responsibilities of Directors of Research Studies Involving Human Subjects*. (This document is also available at <http://www.unh.edu/osr/compliance/irb.html>.) Please read this document carefully before commencing your work involving human subjects.

If you have questions or concerns about your study or this approval, please feel free to contact me at 603-862-2003 or Julie.simpson@unh.edu. Please refer to the IRB # above in all correspondence related to this study. The IRB wishes you success with your research.

For the IRB,


Julie F. Simpson
Manager

cc: File
Ware, Colin
Ferguson, Ariel

University of New Hampshire

Research Conduct and Compliance Services, Office of Sponsored Research
Service Building, 51 College Road, Durham, NH 03824-3585
Fax: 603-862-3564

06-May-2008

Gilman, Anne
Psychology, Conant Hall
327 Forest Park
Dover, NH 03824

IRB #: 4109

Study: Multimodal Binding Capacity in Working Memory

Approval Expiration Date: 30-Oct-2008

Modification Approval Date: 04-May-2008

Modification: Addition of Cognitive Failures Questionnaire

The Institutional Review Board for the Protection of Human Subjects in Research (IRB) has reviewed and approved your modification to this study, as indicated above. Further changes in your study must be submitted to the IRB for review and approval prior to implementation.

Approval for this protocol expires on the date indicated above. At the end of the approval period you will be asked to submit a report with regard to the involvement of human subjects in this study. If your study is still active, you may request an extension of IRB approval.

Researchers who conduct studies involving human subjects have responsibilities as outlined in the document, *Responsibilities of Directors of Research Studies Involving Human Subjects*. This document is available at <http://www.unh.edu/osr/compliance/irb.html> or from me.

If you have questions or concerns about your study or this approval, please feel free to contact me at 603-862-2003 or Julie.simpson@unh.edu. Please refer to the IRB # above in all correspondence related to this study. The IRB wishes you success with your research.

For the IRB,


Julie F. Simpson
Manager

cc: File

Ware, Colin
Ferguson, Arlei

University of New Hampshire

Research Conduct and Compliance Services, Office of Sponsored Research
Service Building, 51 College Road, Durham, NH 03824-3585
Fax: 603-862-3564

07-Oct-2008

Gilman, Anne
Psychology, Conant Hall
327 Forest Park
Dover, NH 03824

IRB #: 4109
Study: Multimodal Binding Capacity in Working Memory
Review Level: Expedited
Approval Expiration Date: 30-Oct-2009

The Institutional Review Board for the Protection of Human Subjects in Research (IRB) has reviewed and approved your request for time extension for this study. Approval for this study expires on the date indicated above. At the end of the approval period you will be asked to submit a report with regard to the involvement of human subjects. If your study is still active, you may apply for extension of IRB approval through this office.

Researchers who conduct studies involving human subjects have responsibilities as outlined in the document, *Responsibilities of Directors of Research Studies Involving Human Subjects*. This document is available at <http://www.unh.edu/osr/compliance/irb.html> or from me.

If you have questions or concerns about your study or this approval, please feel free to contact me at 603-862-2003 or Julie.simpson@unh.edu. Please refer to the IRB # above in all correspondence related to this study. The IRB wishes you success with your research.

For the IRB,



Julie F. Simpson
Manager

cc: File
Ware, Colin
Ferguson, Ariel

University of New Hampshire

Research Conduct and Compliance Services, Office of Sponsored Research
Service Building, 51 College Road, Durham, NH 03824-3585
Fax 603-862-3564

29-Oct-2008

Gilman, Anne
Psychology, Conant Hall
327 Forest Park
Dover, NH 03824

IRB #: 4109

Study: Multimodal Binding Capacity in Working Memory

Approval Expiration Date: 30-Oct-2009

Modification Approval Date: 28-Oct-2008

Modification: Addition of tasks.

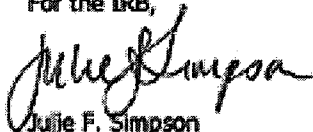
The Institutional Review Board for the Protection of Human Subjects in Research (IRB) has reviewed and approved your modification to this study, as indicated above. Further changes in your study must be submitted to the IRB for review and approval prior to implementation.

Approval for this protocol expires on the date indicated above. At the end of the approval period you will be asked to submit a report with regard to the involvement of human subjects in this study. If your study is still active, you may request an extension of IRB approval.

Researchers who conduct studies involving human subjects have responsibilities as outlined in the document, *Responsibilities of Directors of Research Studies Involving Human Subjects*. This document is available at <http://www.unh.edu/osr/compliance/irb.html> or from me.

If you have questions or concerns about your study or this approval, please feel free to contact me at 603-862-2003 or Julie.simpson@unh.edu. Please refer to the IRB # above in all correspondence related to this study. The IRB wishes you success with your research.

For the IRB,



Julie F. Simpson
Manager

cc: File

Ware, Colin
Ferguson, Ariel

References

- Allen, R. J., Baddeley, A. D., & Hitch, G. J. (2006). Is the binding of visual features in working memory resource-demanding? *Journal of Experimental Psychology: General*, *135*(2), 298-313.
- Allen, R. J., Hitch, G. J., & Baddeley, A. D. (2009). Cross-modal binding and working memory. *Visual Cognition*, *17*(1/2), p83 - 102.
- Allen, S. (2004). Designs for learning: Studying science museum exhibits that do more than entertain. *Science Education*, *88*(S1), S17-S33.
- Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*(2), 106-111.
- Alvarez, G. A., & Cavanagh, P. (2008). Visual short-term memory operates more efficiently on boundary features than on surface features. *Perception & Psychophysics*, *70*(2), 346-364.
- Alvarez, G. A., & Oliva, A. (2007). The role of global layout in visual short-term memory. *Visual Cognition*, *15*(1), 70-73.
- Alvarez, G. A., & Thompson, T. W. (2009). Overwriting and rebinding: Why feature-switch detection tasks underestimate the binding capacity of visual working memory. *Visual Cognition*, *17*(1/2), p141 - 159.
- Amedi, A., Kriegstein, K., Atteveldt, N. M., Beauchamp, M. S., & Naumer, M. J. (2005). Functional imaging of human crossmodal identification and object recognition. *Experimental Brain Research*, *166*(3), 559-571.

- Attneave, F., & Arnoult, M. D. (1956). The quantitative study of shape and pattern perception. *Psychological Bulletin*, *53*, 452–471.
- Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*(7), 622–628.
- Baddeley, A. D. (2007). *Working memory, thought, and action*. Oxford, UK: Oxford University Press.
- Bavelier, D., Newport, E. L., Hall, M., Supalla, T., & Boutla, M. (2008). Ordered short-term memory differs in signers and speakers: Implications for models of short-term memory. *Cognition*, *107*, 433–459.
- Bever, T. G., & Chiarello, R. J. (1974). Cerebral dominance in musicians and non-musicians. *Science*, *185*(4150), 537–539.
- Borenstein, E., & Ullman, S. (2002). Class-specific, top-down segmentation. In *European conference on computer vision* (Vol. 2351, pp. 109–124). Springer-Verlag.
- Brady, T. F., Konkle, T., & Alvarez, G. A. (2008). Efficient coding in visual short-term memory: Evidence for an information-limited capacity. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the Cognitive Science Society* (pp. 887–892). Washington, DC: Cognitive Science Society, Austin, TX.
- Broadbent, D. E., Cooper, P. F., FitzGerald, P., & Parkes, K. R. (19820201). The cognitive failures questionnaire (cfq) and its correlates. *British Journal of Clinical Psychology*, *21*(1), p1 - 16.
- Butcher, K. R. (2006). Learning from text with diagrams: Promoting mental model development and inference generation. *Journal of Educational Psychology*, *98*(1), 182–197.

Callanan, M. A., Jipson, J. L., & Soennichsen, M. S. (2002). Maps, globes, and videos: Parent-child conversations about representational objects. In S. G. Paris (Ed.), *Perspectives on object-centered learning in museums*. Mahwah, US: Laurence Erlbaum.

Calvert, G. A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex*, *11*(12), 1110–1123.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*, 42–45.

Demany, L., Trost, W., Serman, M., & Semal, C. (2008). Auditory change detection: Simple sounds are not memorized better than complex sounds. *Psychological Science*, *19*(1), 85-91.

Donaldson, W. (1993). Accuracy of d' and A' as estimates of sensitivity. *Bulletin of the Psychonomic Society*, *31*, 271–274.

Dux, P. E., Ivanoff, J., Asplund, C. L., & Marois, R. (2006). Isolation of a central bottleneck of information processing with time-resolved fMRI. *Neuron*, *52*, 1109-1120.

Eng, H. Y., Chen, D., & Jiang, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin & Review*, *12*, 1127-1134.

Fabiani, M., Kazmerski, V. A., Cycowicz, Y. M., & Friedman, D. (1996). Naming norms for brief environmental sounds: Effects of age and dementia. *Psychophysiology*, *33*, 462–475.

Feng, J., Spence, I., & Pratt, J. (October 2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science*, *18*(10), 850-855(6).

Forster, S., & Lavie, N. (2007). High perceptual load makes everybody equal: Elimi-

nating individual differences in distractibility with load. *Psychological Science*, 18(5), 377–382.

Fort, A., Delpuech, C., Pernier, J., & Giard, M.-H. (2002). Early auditory–visual interactions in human cortex during nonredundant target identification. *Cognitive Brain Research*, 14(1), 20–30.

Fougnie, D., & Marois, R. (2006). Distinct capacity limits for attention and working memory: Evidence from attentive tracking and visual working memory paradigms. *Psychological Science*, 17(6), 526–534.

Furnham, A., & Walker, J. (2001). Personality and judgments of abstract, pop art, and representational paintings. *European Journal of Personality*, 15(1), 57–72.

Gajewski, D. A., & Brockmole, J. R. (2006). Feature bindings endure without attention: Evidence from an explicit recall task. *Psychonomic Bulletin & Review*, 13(4), 581–587.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction*, 15, 313–331.

Goldin-Meadow, S., & Wagner, S. M. (2005). How our hands help us learn. *Trends in Cognitive Science*, 9, 234–241.

Green, C. S., & Bavelier, D. (2003). Action video games modify visual attention. *Nature*, 423, 534–537.

Green, C. S., & Bavelier, D. (2006). Effect of action video games on the spatial resolution of visuospatial attention. *Journal of Experimental Psychology: Human Perception and Performance*, 32(6), 1465–1478.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, 75, 424–429.

- Griffin, M. M., & Robinson, D. H. (2005). Does spatial or visual information in maps facilitate text recall? Reconsidering the conjoint retention hypothesis. *Educational Technology Research and Development*, 53(1), 23–36.
- Hairston, W. D., Hodges, D. A., Casanova, R., Hayasaka, S., Kraft, R., Maldjian, J. A., et al. (2008). Closing the mind's eye: Deactivation of visual cortex related to auditory task difficulty. *Neuroreport*, 19(2), 151-154.
- Halit, H., Haan, M. de, Schyns, P. G., & Johnson, M. H. (2006). Is high-spatial frequency information used in the early stages of face detection? *Brain Research*, 1117(1), 154–161.
- Halverson, J. (1992). Paleolithic art and cognition. *Journal of Psychology: Interdisciplinary and Applied*, 126(3), 221–236.
- Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, 458, 632–635.
- Hess, C. W., Sefton, K. M., & Landry, R. G. (1986). Sample Size and Type-Token Ratios for Oral Language of Preschool Children. *J Speech Hear Res*, 29(1), 129-134.
- Holmes A, E. M., Winston J. S. (2005). The role of spatial frequency information for ERP components sensitive to faces and emotional facial expression. *Brain Research: Cognitive Brain Research*, 25(2), 508–520.
- Hommel, B., & Colzato, L. S. (2009). When an object is more than a binding of its features: Evidence for two mechanisms of visual feature integration. *Visual Cognition*, 17(1/2), p120 - 140.
- Horowitz, T. S., Klieger, S. B., Fencsik, D. E., Yang, K. K., Alvarez, G. A., & Wolfe, J. M. (2007). Tracking unique objects. *Perception & Psychophysics*, 69(2), 172–184.
- Huang, L., Holcombe, A. O., & Pashler, H. (2004). Repetition priming in visual search: Episodic retrieval, not feature priming. *Memory & Cognition*, 32(1), 12–20.

- Johnson, W. (1944). Studies in language behavior: I. a program of research. *Psychological Monographs*, 56, 1-15.
- Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology*, 59(1), 193-224.
- Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology*, 92, 126-136.
- Kawahara, J.-i. (in press). Auditory-visual contextual cueing effect. *Perception & Psychophysics*.
- Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (submitted). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects.
- Konkle, T., & Oliva, A. (2007). Normative representation of objects: Evidence for an ecological bias in perception and memory. In D. McNamara & J. Trafton (Eds.), *Proceedings of the 29th annual conference of the cognitive science society* (pp. 407-413). Nashville, TN: Cognitive Science Society, Austin, TX.
- Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1(4), 476-490.
- Low, R., & Sweller, J. (2005). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 147-158). Cambridge, UK: Cambridge University Press.
- Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 57(3), 203-220.
- Mayer, R. E. (2001). *Multimedia learning*. New York, US: Cambridge University Press.

- Mayer, R. E. (Ed.). (2005a). *The Cambridge handbook of multimedia learning*. Cambridge, UK: Cambridge University Press.
- Mayer, R. E. (2005b). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). Cambridge, UK: Cambridge University Press.
- Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*(1), 187–198.
- McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, 37*(4), 547–559.
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*(2), 81–97.
- Miyake, A., & Shah, P. (Eds.). (1999a). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge, UK: Cambridge University Press.
- Miyake, A., & Shah, P. (1999b). Toward unified theories of working memory: Emerging general consensus, unresolved theoretical issues, and future research directions. In A. Miyake & P. Shah (Eds.), (pp. 442–481). Cambridge, UK: Cambridge University Press.
- Moreno, R. (2006). Does the modality principle hold for different media? a test of the method-affects-learning hypothesis. *Journal of Computer Assisted Learning, 22*(3), 149–158.
- Mueller, S. T. (2006). *PEBL: The Psychology Experiment Building Language*. (<http://pebl.sourceforge.net/>)

- Muller, D. A., Lee, K. J., & Sharma, M. D. (2008). Coherence or interest: Which is most important in online multimedia learning? *Australasian Journal of Educational Technology, 24*(2), 211-221.
- Noice, H., & Noice, T. (2001). Learning dialogue with and without movement. *Memory and Cognition, 29*(6), 820-827.
- Olsson, H., & Poom, L. (2005). Visual memory needs categories. *Proceedings of the National Academy of Sciences, 102*(24), 8776-8780.
- Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review, 76*(3), 241-263.
- Park, S., Kim, M., & Chun, M. M. (2007). Concurrent working memory load can facilitate selective attention: Evidence for specialized load. *Journal of Experimental Psychology: Human Perception and Performance, 33*(5), 1062-1075.
- Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics, 44*(4), 369-378.
- Pexman, P. M., Hargreaves, I. S., Edwards, J. D., Henry, L. C., & Goodyear, B. G. (2007). The neural consequences of semantic richness: When more comes to mind, less activation is observed. *Psychological Science, 18*(5), 401-406.
- Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin & Review, 15*, 161-167.
- Pickering, S. J. (Ed.). (2006). *Working memory and education*. Burlington, US: Academic Press.
- Pollack, D. A., I.; Norman. (1964). A non-parametric analysis of recognition experiments. *Psychonomic Science, 1*, 125-126.

- Postle, B. R., D'Esposito, M., & Corkin, S. (2005). Effects of verbal and nonverbal interference on spatial and object visual working memory. *Memory & Cognition*, *33*(2), 203–212.
- Pourtois, G., Dan, E. S., Grandjean, D., Sander, D., & Vuilleumier, P. (2005). Enhanced extrastriate visual response to bandpass spatial frequency filtered fearful faces: time course and topographic evoked-potentials mapping. *Human Brain Mapping*, *26*(1), 65–79.
- Prabhakaran, V., Narayanan, K., Zhao, Z., & Gabrieli, J. D. E. (2000). Integration of diverse information in working memory within the frontal lobe. *Nature Neuroscience*, *3*(1), 85–90.
- Ramachandran, V. S. (1990). In C. Blakemore (Ed.), *The utilitarian theory of perception* (pp. 346–360). Cambridge, UK: Cambridge University Press.
- Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object set: The role of surface detail in basic-level object recognition. *Perception*, *33*, 217–236.
- Saiki, J., & Miyatsuji, H. (2005). Feature binding in visual working memory evaluated by type identification paradigm. *Cognition*, *102*, 49–83.
- Schmidt, S. R. (2006). Emotion, significance, distinctiveness, and memory. In R. R. Hunt & J. Worthen (Eds.), *Distinctiveness and memory* (pp. 47–64). Oxford, UK: Oxford University Press.
- Sekuler, R., & Kahana, M. J. (2007). A stimulus-oriented approach to memory. *Current Directions in Psychological Science*, *16*(6), 305–310.
- Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(4), 683–692.
- Sweller, J. (2005). Implications of cognitive load theory for multimedia learning.

In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19–30). Cambridge, UK: Cambridge University Press.

Székely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., et al. (2003). Timed picture naming: Extended norms and validation against previous studies. *Behavior Research Methods, Instruments & Computers*, *35*(4), 621–633.

Tabbers, H. K., Martens, R. L., & Merriënboer, J. J. van. (2007). Space and time, not surface features, guide object persistence. *Psychonomic Bulletin & Review*, *14*(6), 1199–1204.

Treisman, A. M. (2006). Object tokens, binding, and visual memory. In H. D. Zimmer, A. Mecklinger, & U. Lindenberger (Eds.), *Handbook of binding and memory: Perspectives from cognitive neuroscience* (pp. 315–338). New York, US: Oxford University Press.

Treisman, A. M., & Zhang, W. (2006). Location and binding in visual working memory. *Memory and Cognition*, *34*, 1704–1719.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, *5*(7), 1–6.

Underwood, B. J., & Schulz, R. W. (1960). *Meaningfulness and verbal learning*. Chicago, US: J. B. Lippincott.

Uttal, D. H., Liu, L. L., & DeLoache, J. S. (2006). Concreteness and symbolic development. In L. Balter & C. S. Tamis-LeMonda (Eds.), *Child psychology: A handbook of contemporary issues* (2nd ed., chap. 8). New York: Psychology Press.

Vanderplas, J. M., & Garvin, E. A. (1959). The association value of random shapes. *Journal of Experimental Psychology*, *57*, 147–154.

VanRullen, R. (2009). Binding hardwired versus on-demand feature conjunctions. *Visual Cognition*, *17*(1/2), 103–119.

- Vekiri, I. (2002). What is the value of graphical displays in learning? *Educational Psychology Review*, 14(3), 261–312.
- Visscher, K. M., Kaplan, K., Kahana, M. J., & Sekuler, R. (2007). Auditory short-term memory behaves like visual short-term memory. *Public Library of Science, Biology*, 5, e56.
- Vogel, E. K., & Awh, E. (2008). How to exploit diversity for scientific gain: Using individual differences to constrain cognitive theory. *Current Directions in Psychological Science*, 17(2), 171–176.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1), 92–114.
- Wheeler, M., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, 131(1), 48–64.
- Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5, 1–7.
- Xu, Y. (2002). Limitations of object-based feature encoding in visual short-term memory. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 28(2), 458–468.
- Zacks, J., Levy, E., Tversky, B., & Schiano, D. J. (1998). Reading bar graphs: Effects of extraneous depth cues and graphical context. *Journal of Experimental Psychology: Applied*, 4(2), 119–138.
- Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrika*, 70(1), 1–10.