UNIVERSITY OF CALIFORNIA

Los Angeles

Investigating the Relationship Between Perceptual Learning and Statistical Learning in Human Vision

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Psychology

by

Carolyn Ann Bufford Funk

2019

ABSTACT OF THE DISSERTATION

Investigating the Relationship Between Perceptual Learning and Statistical Learning in Human Vision

by

Carolyn Ann Bufford Funk

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2019

Professor Philip Kellman, Chair

Perceptual learning (PL) and statistical learning (SL) both seek to explain how humans learn through experience. However, research lacks clear, consistent definitions, causing collective confusion about the relationship between SL and PL: some researchers view SL and PL as distinct learning processes, while others view them as part of a unified learning process. We describe two distinct learning concepts that cohere with most work: PL is improving perception and, in a psychophysical sense, improving sensitivity. SL is recording co-occurrences in memory, and perhaps, in a psychophysical sense, changing criterion or bias. These concepts allowed us to test the relationship between PL and SL. We developed a novel psychophysical assessment to measure incidental PL in a well-known visual SL paradigm of shape pairs presented simultaneously. Experiments 1 and 3 used the paradigm's SL familiarization and SL familiarity test, then our PL assessment. We also tested incidental SL in PL: Experiment 2 used

PL training on stimuli from the SL paradigm, then a brief SL familiarity test for an incidental pair, then the assessment. Experiments 1 and 3 showed familiarity, but Experiment 2 showed no evidence of familiarity in the SL test. Experiments 1 and 3 showed PL effects: transfer in increased accuracy and sensitivity and decreased false alarming relative to baseline, evidence of incidental PL in an SL paradigm. Reducing SL by reducing familiarization session length caused weaker PL, observed only for the longest exposure duration. Experiment 2 showed stronger PL effects on the assessment, but no SL. We discuss several ways SL and PL could be related and compare each possibility to our results. Several of our results are consistent with SL and PL being part of a unified learning process, or at least occurring under overlapping conditions, but the relationship may be more nuanced and asymmetric: PL may occur more under conditions designed for SL, but SL may be less likely to occur during focused PL tasks. These results help clarify and unite rich but separate literatures on perceptual and statistical learning.

The dissertation of Carolyn Ann Bufford Funk is approved.

Gregory A Bryant

Hongjing Lu

Ladan Shams

Philip Kellman, Committee Chair

University of California, Los Angeles

2019

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to acknowledge my co-authors, Drs. Everett Mettler and Phil Kellman. Dr. Kellman was the PI, overseeing the project, and particularly contributing to the design of Experiment 1, particularly the psychophysical assessment, the design of Experiment 2, and interpretation of findings. Dr. Mettler also helped design the psychophysical assessment, wrote the initial version of the program for Experiment 1, advised on revisions to the code for presenting stimuli, and managed the code for accessing the data. I have acknowledged the contributions of co-authors throughout this dissertation by using "we" for sections to which my co-authors substantially contributed (as described above), and "I" only for sections that are primarily my own individual work. A more condensed version of this dissertation is in preparation for publication:

Bufford Funk, C. A., Mettler, E., & Kellman, P. J. (in preparation). *Investigating the relationship between perceptual learning and statistical learning in human vision.*

I would also like to acknowledge Dr. Hongjing Lu for helpful discussion of the psychophysical assessment (specifically recommending not showing gridlines). Thanks also to members of the UCLA Human Perception Lab for help and support, and to family and friends for their support.

BIOGRAPHICAL SKETCH

Education

2014-2017	C.Phil. in Psychology**,** University of California, Los Angeles
2012-2014	M.A. in Psychology, University of California, Los Angeles
2008-2012	B.S. in Cognitive Science with a Specialization in Computing, with College Honors, magna cum laude, University of California, Los Angeles

Publications

Chiang, J.N., Rosenberg, M. H., **Bufford, C.A**., Stephens, D., Lysy, A., & Monti, M.M. (2018). The language of music: Common neural codes for structured sequences in music and natural language. *Brain and Language, 185,* 30-37.

Bufford, C.A., Thai, K.P., Ho, J., Xiong, C., Hinges, C.A., & Kellman, P.J. (2016). Perceptual learning of abstract musical patterns: Recognizing composer style. In T. Zanto, T. Fujioka, P. Janata, J. Johnson, J. Berger, J. Slater, & C. Chafe (Eds.), *Proceedings of the 14th Annual Conference of the International Conference for Music Perception and Cognition* (pp. 8-12).

Bufford, C.A., Mettler, E., Geller, E.H., & Kellman, P.J. (2014). The psychophysics of algebra expertise: Mathematics perceptual learning interventions produce durable encoding changes. In P. Bellow, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 272-277). Austin, TX: Cognitive Science Society.

Bufford, C.A., Mettler, E., & Kellman, P.J. (in prep). Perceptual learning in mathematics produces durable changes in perception.

Presentations

Kellman, P.J., **Bufford, C.A**., & Mettler, E. (2018). *The psychophysics of algebra: Mathematics perceptual learning interventions produce measurable and lasting changes in the perceptual encoding of mathematical objects.* Oral presentation made by first author at the meeting of the Vision Sciences Society, St. Pete Beach, Florida.

Kellman, P.J., **Bufford, C.A**., & Mettler, E. (2017).*The psychophysics of algebra: Mathematics perceptual learning interventions produce measurable and lasting changes in the perceptual encoding of mathematical objects*. Oral presentation by first author at the meeting of the Psychonomic Society, Vancouver, British Columbia, Canada.

Bufford, C.A., Thai, K.P., Ho, J., Xiong, C., Hinges, C.A., & Kellman, P.J. (2016). Perceptual learning of abstract musical patterns: Recognizing composer style. Poster presented at the meeting of the International Conference for Music Perception and Cognition, San Francisco, CA.

Bufford, C. A. & Kellman, P.J. (2015). *Learning in mathematics produces durable encoding improvements.* Poster presented at the meeting of the Cognitive Science Society, Pasadena, California.

Bufford, C.A. & Kellman, P.J. (2015). *Mathematics perceptual learning causes lasting encoding gains.* Poster presented at the meeting of the Association of Psychological Science, New York, New York.

Bufford, C.A., Mettler, E., Geller, E.H., & Kellman, P.J. (2014). The psychophysics of algebra expertise: Mathematics perceptual learning interventions produce durable encoding changes. Oral presentation at the meeting of the Cognitive Science Society, Quebec City, Quebec, Canada.

Bufford, C.A., Mettler, E., Geller, E.H., & Kellman, P.J. (2014). *Capturing mathematics perceptual learning through psychophysics.* Poster presented at the meeting of the Association of Psychological Science, San Francisco, California.

Teaching Experience and Professional Service

2014-2019	Teaching Assistant/Associate/Fellow, **UCLA Psychology Department**
2015-2019	Volunteer, **Psychology Undergraduate Research Conference**
2017-2018	**UCLA Alumni Mentoring Program Mentor**
2016-2018	Judge, **California State Science Fair**; **Los Angeles County Science Fair** (2017)
2012-2015	**Psychology in Action** blogger and symposium committee member

CHAPTER 1: INTRODUCTION

Investigating the relationship between perceptual learning and statistical learning in human vision

Humans learn about the world through experience. This is possible because the world is not chaotic and entirely random, but instead has regularities which we use to guide behavior. There are at least two general kinds of regularities that have special importance. One involves the statistical structure in objects, arrangements, and events. Another is the recurring importance of features and relations that determine important classifications and which may be useful to come to fluently encode and discriminate. Both the registration of regularities, statistical learning (SL), and the improved sensitivity of perceptual encoding, perceptual learning (PL), can allow us to perform better on tasks over time. Understanding how do these learning processes relate would inform how best to structure learning.

SL and PL differ in how we use regularities in the world. SL researchers often assume a memory mechanism through which familiar patterns are recognized (e.g. Fiser, 2009). PL is about improving pickup of task-relevant patterns, category boundaries, etc. and/or suppression of irrelevant information (e.g. Kellman, 2002; Kellman & Garrigan, 2009), so it is about perception and transfer to new stimuli (e.g. Bufford, Thai, Ho, Xiong, Hines, & Kellman, 2016; Goldstone, 1998).

Definitions PL posits improved information pickup as the mechanism for learning through experience: Eleanor Gibson pioneered modern PL, and she defined perceptual learning as changes in the encoding of information due to experience or practice (Gibson, 1953; Gibson, 1969; Gibson & Gibson, 1955; these papers combined have been cited over 6,000 times, according to Google Scholar). Under this definition, expertise is a change in sensitivity through perceptual information pickup attuned to patterns and structures imperceptible or not attended to by novices. Active training with feedback is known to accelerate the pace of PL relative to unstructured or less-structured learning without consistent feedback (e.g. Gibson, 1947; Kellman, Massey, & Son, 2010), though feedback is not required (e.g. Gibson & Gibson, 1955).

In contrast, Saffran and colleagues (1996), whose SL paper is the best-known (cited almost 5000 times, according to Google Scholar), never defined statistical learning or cited prior psychology or computer science research on SL, but implied[1] the following definition of SL: the automatic and implicit learning and recall of statistical relationships in stimuli, especially between neighboring items. Under this definition, the encoding of information does not change as expertise increases, only the ability to track and expect statistically reliable sequences and/or co-occurrences – the bias. Thus, SL theorizes a memory mechanism for learning through experience. Because SL is automatic, no training is expected or needed, but more exposure to patterns might make memory traces stronger.

Connection to Signal Detection Theory. We can relate these two concepts of learning to the fundamental goal of signal detection theory (SDT). SDT analyses separate sensitivity in the acquisition or use of information presented from criterion or bias, which relates to response tendencies, apart from the signal presented at a given moment, that reflect stimulus probabilities, incentives, etc. (Wickens, 2002). The distinction we are proposing between SL and PL maps onto this framework directly. Perceptual learning is an improvement in the sensitivity to signal; it would be apparent in many tasks as a change in an SDT measure of sensitivity such as d'. Statistical learning - registering the statistics of features or co-occurrences in displays over

[1] The fact that this seminal paper did not include a definition of SL likely has contributed to confusion around the term.

some amount of experience with them - would provide useful knowledge about the world. In an SDT framework, it could allow more accurate guessing in the absence of a signal. Variables such as statistical information about signal probability are known in SDT experiments to affect *criterion* or *bias,* which refers to the overall probability of use of the response options.

Possible Relationships Between SL and PL

But what is the relationship between SL and PL? SL and PL might be distinct learning processes with distinct mechanisms, contributing separately to learning regularities: SL to familiarity and PL to improved information pickup. The distinct ways in which PL and SL can be defined (as they have here) suggest this might be the case. But perhaps phenomena of perceptual learning and statistical learning are actually effects of a single unified learning mechanism for learning many kinds of regularities through experience, in which conditions leading to familiarity also produce changes in perceptual encoding. Recognizing which of two patterns has been encountered before (statistical learning) is a different outcome from becoming faster or more selective in encoding stimuli, but perhaps in human cognition, these processes go together. It is also possible that SL is a subtype of PL or a step in the process of PL – tracking which elements go together might be one of the ways of identifying regularities to then selectively pick up, which would be a more nuanced relationship. My dissertation explores these possible relationships empirically.

Structure of the Introduction

I will use the above definitions throughout my dissertation, as our guide to investigating the relationship between SL and PL because they are consistent with the most influential research and allow clear distinctions to be made between SL and PL. However, in this introduction, I will explore historical and recent research explicitly using the terms “statistical learning" and/or "perceptual learning" in ways consistent and inconsistent with our definitions. In so doing, I will trace changes in usage of these terms over time to help contextualize the current understandings of SL and PL in the literature, compare SL and PL in the context of the literature, and illustrate the current collective confusion about the relationship between SL and PL that this dissertation addresses empirically. Finally, I will introduce the empiric approach I have taken to addressing the question about the relationship between SL and PL in this dissertation.

Perceptual Learning

Eleanor Gibson pioneered the modern view of perceptual learning[2]. She defined perceptual learning as the process of changing the pickup of information from the world because of experience or practice (Gibson, 1953; Gibson, 1969; Gibson & Gibson, 1955). In perceptual learning, it is not the case that the learner memorizes mappings of stimulus to response or memorizes relationships between unchanged perceptual units; instead, previously ignored or imperceptible patterns become perceptible and the way the world is perceived changes – the basic encoding of stimuli is psychophysically optimized (e.g. Bufford, Geller, Mettler, & Kellman, 2014; Leek & Watson, 1988; Notman, Sowden, & Özgen, 2005).

Encoding changes may take the form of chunking, the process of forming chunks – diagnostic combinations of components that are treated as a single unit in memory (e.g. Chase & Simon, 1973; Goldstone, 1998; Goldstone, 2000) – which then enable perception of complex patterns and structures, which distinguish classes of stimuli. By improving information encoding by chunking or other processes (e.g. Epstein, 1967) to more closely match the environment

[2] The term "perceptual learning" was in use before the 1950s, but it did not mean the same thing.

(Gibson & Gibson, 1955), the learner becomes more attuned to the relevant patterns and structure in encountered stimuli. These changes to perception can be characterized as discovery and fluency effects: *discovery effects* include selecting relevant information or seeing new relationships and suppressing irrelevant information, and *fluency effects* include faster encoding and reduced cognitive load (Kellman, 2002; Kellman & Garrigan, 2009). Typical PL studies follow a pretest-posttest design with a learning phase in which participants actively attempt to classify stimuli and receive feedback, and studies often employ one or more control groups.

Eleanor Gibson and her contemporaries studied expertise with real-world stimuli such as aircraft (Gibson, 1947), morse code (Keller, 1943), and chick-sexing images (Biederman & Shiffrar, 1987) and artificial stimuli such as squiggles (Gibson & Gibson, 1955). In the spirit of this foundational work, some recent research has focused on real-world stimuli (e.g. radiological images: Kellman, 2013; Kok, de Bruin, Robben, & van Merriënboer, 2013; butterflies: Mettler & Kellman, 2015) and even extended PL to symbolic stimuli (e.g. Chinese characters: Thai, Mettler, & Kellman, 2011; mathematics: Bufford et al., 2014; Cheng, 2014; Ottmar, Landy, Weitnauer, & Goldstone, 2015). Such research has shown how technology can be used to facilitate PL in real-world and symbolic domains (e.g. Mettler & Kellman, 2015) and even provided direct evidence of encoding changes in symbolic domains (Bufford et al., 2014; Thai et al., 2011). With real-world and symbolic stimuli, participants show improved encoding of stimuli, a hallmark of PL.

In contrast to research in real-world and symbolic domains, most PL research in the last two decades has focused on tasks with simple, often artificial stimuli (*audition*: learning exact pitch duration, e.g. Karmarkar & Buonomano, 2003; phoneme discriminations, e.g. Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Norris, McQueen, & Cutler, 2003; *multimodal*:

temporal discrimination, e.g. Bratzke, Seifried, & Ulrich, 2012; Nagarajan, Blake, Wright, Byl, & Merzenich, 1996; length perception, e.g. Wagman, Carello, Schmidt, & Turvey, 2009; rhythm perception: Bakarat, Seitz, & Shams, 2015; *vision*: Vernier acuity, the alignment of line pairs, e.g. Westheimer & McKee, 1978, as cited in Kellman & Garrigan, 2009; orientated visual gratings, e.g. Song, Peng, Lu, Liu, & Li, 2007). Participants in these tasks have shown orders of magnitude improvements (e.g. Ahissar & Hochstein, 1993).

The motivation for studying basic sensory discriminations in much contemporary work has been the idea that PL may be based on sensory plasticity (change in receptive fields) in early cortical analyzers (e.g., Fahle & Poggio, 2002). However, subsequent work has shown that perceptual learning in both basic sensory and more complex domains is unlikely to depend on receptive field change (at least in vision), and is better accounted for by concepts of discovery and selection (weighting) of which analyzers provide the most useful information for a task (Petrov, Dosher & Lu, 2005; Garrigan & Kellman, 2008; for a review, see Kellman & Garrigan, 2009).

Statistical Learning

Early Usage of Statistical Learning. As with perceptual learning, the phrase *statistical learning* has been used to mean several different things. In the worlds of academic statistics and education research, it describes the learning of the school subject of statistics (e.g. Belshaw, 1951; Lajoie, 1997; Watson, 1997). The earliest known (according to Google Scholar) use of the term belongs to this group, in a lament over the state of statistical learning in the U.S. in the year before the 1880 census (Walker, 1879). This meaning of statistical learning is pertinent only to my having learned statistics and conducted statistical analyses in this dissertation, but not otherwise.

In psychology, the earliest appearance of the term statistical learning was in reference to Estes' Statistical Learning Theory (Estes, 1950). Estes' theory was part of the behaviorist movement of mapping stimuli to responses (S-R learning), in that he proposed a computational theory relating the probability of a stimulus to the way in which a response was learned, relating to target frequency matching. Estes' theory was a popular subject of research (e.g. Estes, 1962; Estes & Straughan, 1954; Suppes & Atkinson, 1960) until the early 1970s, when it was found to be "in error in so many respects" (Reber & Millward, 1971) that it fell out of favor with the research community.

The term statistical learning has been used in computer science and related fields from as early as 1988 (e.g Jannarone, Yu, & Takefuji, 1988). In this field, statistical learning describes a method by which artificial learning systems that exploit any and all reliable statistics in their input to improve output over time or training. Pednault (1997) defined a Statistical Learning Theory for this field, in which the computer learns the correct model of several competing initial models through the input data. Models under this theory and other artificial statistical learning processes are more computationally sophisticated than Estes' theory, but share the same spirit of mapping input to output (or stimulus to response).

Statistical Learning in Human Perception and Learning. The computer science idea of statistical learning – tracking reliable statistics in data to optimize responding – influenced the modern perception psychology view of statistical learning – tracking reliable statistics to guide behavior. In their seminal paper on auditory statistical learning (SL) in infants, Saffran and colleagues (1996) never stated a definition of SL but implied the following definition of SL: Statistical learning (SL) is the automatic process of tracking in memory the statistical relationships between stimuli. The encoding of information is unchanged, but as one gains more experience with stimuli, statistically reliable relationships between basic elements are recorded in memory and exploited to process stimuli more efficiently. In studies, a single group of learners are typically passively familiarized with stimuli, then tested in a 2AFC recognition test. For example, in an auditory sequence liuke that sed by Saffran et al. (1996), the syllable “bim” might immediately follow “ji” 100% of the time. The transition probabilities (TPs) from one syllable to another are controlled and compared in analyses, often setting TPs at 100% within sequences, and at chance for transitions across sequences, e.g. 25% “ji” then “ku”.

Much of the modern work in SL has followed this paradigm of sequential stimuli. In vision, real-world stimuli have seldom been studied (exceptions: e.g. Brady & Oliva, 2008, and Emberson & Rubinstein, 2016, who manipulated the order of category photographs) because controlling the statistical relationships in real-world scenes or videos would be extremely difficult. Instead, visual SL training typically consists of passive viewing of shapes, often sequences of tens or hundreds of repetitions of a small vocabulary. In such visual SL studies, statistical features are typically brief two- or three-shape subsequences, e.g. triangle-square-circle and diamond-heart-spade, repeatedly embedded in the full training sequence. The transition probabilities or TPs (see Saffran et al., 1996)[3] from one shape to another are controlled and compared in analyses, often setting TPs at 100% within sequences, e.g. triangle then square, and at chance for transitions across sequences, e.g. 25% circle then diamond.

Participants then complete a 2-alternative forced-choice posttest: participants are asked to choose which of two shape sequences is more familiar, the embedded subsequence or the same-length foil sequence that either bridged subsequences, e.g. heart-spade-triangle or never

[3] Formula for transition probability: $Y|X = \frac{frquency\ of\ XY}{frequency\ of\ X}$

appeared, e.g. square-triangle-diamond. Learners show evidence of learning predictive relationships in familiarization stimuli (auditory: e.g. Creel, Newport, & Aslin, 2004; Endress, Nespor, & Mehler, 2009; Pelucchi, Hay, & Saffran, 2009; multimodal: e.g. Barakat, Seitz, & Shams, 2015; Conway & Christiansen, 2006, 2009; Robinson & Sloutsky, 2007; Seitz, Kim, van Wassenhove, & Shams, 2007; visual: e.g.Turk-Browne, Jungé, & Scholl, 2005; Saffran et al., 1996), almost always without having been told such statistical relationships exist. The fact that participants choose subsequences from the exposure more frequently than chance in these 2AFC familiarity tests is taken as evidence of statistical learning.

Founding Modern SL (in Human Vision). Modern research in SL in adult vision traces back to two papers published in 2001 (Fiser & Aslin, 2001; Hunt & Aslin, 2001). Instead of S-R relationships, this SL literature is (according to our definition) about stimulus-stimulus relationships. Fiser and Aslin (2001) used spatial arrangements instead of temporal arrangements: pairs of shapes were presented simultaneously and always co-occurred. Pairs were arranged in grids with three pairs per grid, on each passive familiarization trial. Participants showed SL by identifying familiar pairs over foil pairs more frequently than predicted by chance. This design followed our definition of SL, and is much better known that the other paper.

The other paper investigated statistical learning of light sequences (Hunt & Aslin, 2001). This study followed the ideas of sequential statistical learning studies, but departed from them and from our definition of SL in several important ways. Instead of using passive familiarization, this study required a motor response from participants on each trial. Participants clicked on lighted buttons, which were lit in specific sequences to form visual “words”. Instead of testing familiarity, the authors measured reaction time (RT) to touch the buttons and tracked RT change across learning, which is not a test of memory, the essence of SL, under our definition. Instead of learning in a single, brief session, participants learned for an hour per day for eight days. Across three experiments, the authors found that participants showed both general speed up on the task but also particularly speeded responses for predictable lights (later parts of each “word”). Analyses of individual performances showed that some participants learned both the statistics for three-part “words” as intended in addition to the statistics of pairs in those words, while others learned just pairs and not complete triples (“words”). This paper, though less frequently cited than the other~~s~~ from the same year, is especially important because it foreshadowed the broadening of visual SL that has occurred in recent years.

Modern Visual SL: Changing Methods, and Broadening Definitions. Adding to the confusion, much of the most recent work in SL has moved away from conforming to prior methods (typically brief passive familiarization followed by a 2AFC familiarity test) and findings, and from the prior conception - and thus, our definition - of SL. Researchers have tested the automaticity of SL and its unconsciousness and have even moved away from SL as item correlations. They have shown the insufficiency of such definitions to fully explain SL-related phenomena. In so doing, recent research contributed to collective confusion about SL and its relationship to PL.

For example, Turk-Browne et al. (2005) built upon the typical sequential presentation methodology to investigate the role of automaticity in statistical learning in a series of experiments. They assigned half of their shape vocabulary to the color red and half to green, and defined subsequences within each color individually, e.g. triangle-square-circle in red and diamond-heart-spade in green, to create two experimental sequences. Then they presented the two sequences pseudorandomly interleaved and gave participants a cover task in one color only.

Participants completed 2AFC familiarity judgments for each color's subsequences and foils separately, with all test shapes presented in black.

Participants demonstrated typical familiarity for sequences in the attended color, but no familiarity – no learning – in the unattended color. Because this was a novel finding, the authors conducted several follow-up experiments to determine if any reasonable modification of their experiment design would show learning in the unattended color. They increased the training, restored training colors in the test, reversed the test colors, and even replaced the familiarity test with a serial reaction-time test, which is an implicit behavioral measure of learning. The authors concluded that statistical learning did not occur automatically in that it did not occur without attention: selective attention gated what statistical information was available for learning – only the information in the attended color. Within the attended color, participants learned subsequences despite being engaged in a subsequence-irrelevant cover task. So only within attended stimuli, the regularities of the subsequences were automatically learned. Thus, the authors found using both familiarity tests and reaction time in a rapid serial visual presentation (RSVP) task that statistical information cannot be learned when it is ignored or task-irrelevant, so SL is only automatic to for stimuli that are not ignored. This is an important deviation from the definition of SL as being completely automatic without caveats.

Other researchers have explored other assumptions of early (and our) definitions of SL, such as the degree to which SL is implicit. Some have concluded that SL is implicit (e.g. Kim, Seitz, Feenstra, & Shams, 2009), while others have concluded that SL is at least partially explicit for at least some participants (Bertels, Franco, & Destrebecqz, 2012). Researchers have also examined the extent to which the learning in SL paradigms is purely driven by the statistics of the stimuli, through examining the effects of perceptual grouping on SL and applying computational models to SL experiment data.

Researchers have studied connectedness (Baker, Olson, & Behrmann, 2004), similarity (Glicksohn & Cohen, 2011), and common fate (Fiser, Scholl, & Aslin, 2007) in SL paradigms. In all of these studies, the Gestalt grouping influenced what was learnable: When the connected shape was task-relevant to the target shape, it was learned (Baker et al., 2004). Embedded shape pairs in the same color were learned when other non-colored or cross-colored embedded pairs were not (Glicksohn & Cohen, 2011). The speed at which two shapes moved toward and two different shapes moved away from an occluder (in a stream-bounce illusion paradigm) influenced which shapes participants paired together – participants saw streaming at slower speeds and bouncing at faster speeds. Under our definition of SL, only the statistics of the stimuli should influence learning, and all statistics should be learned, but research on grouping effects indicates that this is not the case.

Computational modeling studies of visual SL can compare models based on transition probabilities (a unit of analysis in SL - see footnote # 3) to other kinds of models, such as chunking models. This comparison can illuminate whether or not human data fits the SL definition of recording correlations between elements in memory without changing perception (i.e. transition probability-based models), or if, instead, perception is changed (the definition of PL) to pick up correlated elements as a single unit (i.e. chunking models).

Computational modelling studies of SL also challenge our definition of SL. In a sequential shape design with infants (Slone & Johnson, 2018), transition probability models were found to fit the data less well than a chunking model. Computational models have also been fit to adult data with simultaneous shape arrays. Fiser and Aslin (2001) followed up on their adult study of arrays of paired shapes in grids by investigating larger units: shape triples, quads, and sextuples. They used the same shapes as in their original study, and similar (but larger) grids. Participants learned the larger units, but not embedded pairs within the units (Fiser & Aslin, 2005), contrary to the idea from our definition of learning all statistically reliable relationships. This result was replicated by Lu and Lee (2013). When computationally modelling data from their studies of larger arrangements of shapes, Orban and colleagues found that transition probability models fit the data less well than chunking models (Orbán, Fiser, Aslin, & Lengyel, 2008). These findings contradict the assumptions of our definition SL, but suggest that SL might be related to PL because chunking is part of PL.

SL assumes that learning should be tied to the learned stimuli in their exact learned relationship (spatial or temporal), but research on transfer with SL also hints that SL may not be independent of PL. Turk-Browne and Scholl (2009) trained participants on spatial arrays and found transfer to sequentially presented stimuli. They also found the opposite – transfer from sequential stimuli to spatial arrays of stimuli. They even found transfer from sequential stimuli triplets to backwards-ordered sequential triplets. Similarly, Otsuka, Nishiyama, Nakahar, and Kawaguchi (2013) found some transfer to reversed triplets, as well as transfer from line drawings to words (category names of the objects in the drawings). If temporal and spatial arrangements can transfer to each other, temporal order can be reversed, and learning can transfer from visual stimuli to semantic versions of those stimuli (which are visually very different), then another assumption of our definition of SL does not hold.

Finally, Bakarat and colleagues (2013) used a reaction-time measure to compare learning on first and second items of sequentially presented shape pairs. They found faster reaction times for the second item of each pair. Improved reaction time could be due to quickly retrieving the second item from memory (SL) or improved pickup of information (PL). The fact that the second items showed higher sensitivity than the first items suggests PL occurred in this SL paradigm.

SL As Any Reliable Statistic. Perhaps in response to the assumptions of early definitions (including mine) of SL not capturing all SL-related phenomena, modern definitions of SL have been broadened to showing learning in any way of any reliable statistic, including global statistics. For example, Jones and Kaschak (2012) made a location in their visual search task more likely to be the correct location, and found learning of a statistical bias for that location in more initial saccades to that location. Cosman and Vecera (2014) also found that participants learned a bias in their attentional capture design. Participants showed a bias in reaction time of faster responding to valid cues in the color in which targets appeared more often in training, and slower responding to invalid cues. These studies manipulated and found learning of a global statistic as opposed to manipulating relationships between individual stimuli. As such, this is a departure from our definition of SL, and may or may not be the same phenomenon.

Bays, Turk-Browne, and Seitz (2016) found evidence of two separate learning processes for environmental statistics. Participants showed improvements in accuracy on particular stimuli in their detection task and improvements in reaction time on other stimuli in an RSVP task. The dissociation of the tasks and learning effects seen is further evidence that there may be multiple visual SL processes, including learning of element relationships and global statistics.

Comparing PL and SL

Both perceptual learning and statistical learning intend to explain how we learn from regularities in the world, yet SL and PL differ in how they theorize that we use these regularities, proposing different mechanisms for learning through experience: SL (as defined in this dissertation) proposes a memory mechanism, whereas PL has a perceptual mechanism.

Recent research has challenged the assumptions of our definition of SL and broadened the definition of SL in the literature to any reliable statistic (e.g. Cosman & Vecera, 2014). Computational modeling (e.g. Orbán et al., 2008) and transfer (e.g. Otsuka et al., 2013) findings suggest that PL and SL may be closely related, for learning of statistical relationships between items, because transfer is a hallmark of PL and because chunking is one of many possible encoding changes due to PL. However, it is essential to maintain distinct definitions and concepts of SL and PL to be able to investigate the relationship between them, so I will continue with our definitions (as stated above).

The Problem of Different Methodologies. SL and PL tend to be studied in different ways. PL research tends to involve active training over a longer period of time and test for improvements on a task and often include control group(s), whereas SL research tends to have a very brief passive learning phase and a single familiarity test with a single group[4]. Both the differences in paradigms and the differences make direct comparison of SL and PL studies difficult at best.

Additionally, PL and SL have been studied with different stimuli. PL studies have more often employed real-world stimuli than SL studies. SL is almost exclusively studied with artificial stimuli in vision. SL researchers favor artificial stimuli because researchers easily can control and manipulate statistical features of such stimuli. PL researchers interested in real-world expertise are less concerned with computing statistical probabilities between stimuli and parts of stimuli than with improving performance on real-world tasks. Different methodologies do not

[4] There is, however, evidence of "fast learning" in PL in which large learning gains are demonstrated in a single session (e.g. Ahissar & Hochstein, 1993; Hawkey, Amitay, & Moore, 2004). SL has also been studied over a relatively long time scale (e.g. Hunt & Aslin, 2001).

necessarily imply different learning processes, but they may contribute to the variety of opinions as to SL and PL in the literature.

Different methodologies are also due, in part, to different research goals. SL researchers, understandably, want to learn how stimuli statistics influence learning, generally in language learning and other higher cognitive processes. PL researchers are interested in a broader array of questions around improving expertise and learning in many different domains, and in understanding neural mechanisms underlying PL. The abundance of research into mechanisms of PL using simple, often artificial stimuli has advanced the neuroscience of visual perceptual learning, but its very productivity has become a problem because it has given at least some researchers outside the field of PL (e.g. Fiser, 2009) the impression that PL is not also involved in higher cognitive processes (which, by their nature, involve more complex real-world and/or symbolic stimuli).

Researchers' Conflicting Ideas in the Literature. Amongst researchers, there is collective confusion about the relationship of SL and PL. Many will write about one or the other, and just include a token citation or two from the other literature. Because of limited cross-talk, researchers have many conflicting understandings of SL and PL and their relationship.

Hypothesis of different kinds of learning Some consider SL and PL to be distinct kinds of learning. For example, Fiser and colleagues claim that SL "refers to a particular type of implicit learning … fundamentally different from traditional perceptual learning" (Fiser, Berkes, Orbán, & Lengyel, 2010). In his review, Fiser (2009) asserts that the goal of statistical learning is to capture "useful aspects of the low-level sensory input for further processing". He further asserts that "sensory input … is fairly clear but ambiguous" because "[i]t can support far too many possible combinations of elements that all could be potentially relevant higher order features".

Fiser expresses a common misunderstanding of perception as only providing basic elements that higher cognition must associate into meaningful relationships, instead of perception being about picking up meaningful structure in the world at all levels of visual and task complexity.

Hypothesis of a unified learning process Others, perhaps the majority, consider SL and PL to be part of a unified learning process. For example, Bao (2015) describes three categories of PL, unsupervised, supervised, and reinforcement learning, and describes SL as unsupervised PL. Conway and colleagues (2007) assert that "VSL may be more closely related to perceptual processing – specifically, perceptual learning – than to associative learning phenomena". In a review of SL, Frost and colleagues (2015) state that SL studies "suggest that there are independent modality constraints in learning distributional information, pointing to modality specificity, and further to stimulus specificity akin to perceptual learning".

Researchers also demonstrate this view when they use the terms SL and PL interchangeably, characterizing other researchers' work using different terminology than the original authors, typically SL as PL. For example, Series and Seitz (2013) refer to work with the term "statistical learning" in the title as both PL and SL: "Turke-Browne and Scholl (2009) provide evidence for transfer of perceptual learning across space and time, suggesting that statistical learning leads to flexible representations". Similarly, Toro and colleagues (2005) refer to a different paper with the phrase "statistical learning" in the title as being PL: "a recent study on visual perceptual learning by Baker, Olson, and Behrmann (2004) suggests that the extraction of statistical correlations among elements of separate objects can be prevented when attention is not directed to both objects".

Other researchers consider SL and PL distinct but related learning processes. This is a variation on the hypothesis of a unified learning process, in which the relationship is weaker.

Researchers have approached studying SL and PL together with this assumption of different but related processes. If one could attempt to "isolate the effect of perceptual learning on statistical learning" (Glicksohn & Cohen, 2011), then PL and SL must be distinct but related learning processes. This weaker view may be correct in that learning a global bias would not be considered PL, but it does not seem to follow from most of the recent SL work on element relationships. Given that SL research is best modeled in chunks and not in element-preserving transition probabilities and the transfer of SL, the hypothesis of a unified learning process is more likely to be correct than the hypothesis of different kinds of learning[5].

Dissertation Problem and Approach

Dissertation Problem. Researchers differ in their ideas about SL and PL and the relationship of SL and PL, causing collective confusion. Important differences in experimental paradigms contribute to this confusion because existing SL and PL studies cannot be directly compared. SL and PL researchers support competing hypotheses of distinct learning processes and of a unified learning process, but both cannot be correct. Thus, the nature of the relationship is between SL and PL in human vision is not clear.

To date, no one has investigated the relationship of PL and SL in the same experiment. Thus, it is unknown whether these conceptually distinct proposed processes are in fact different learning processes in humans or are actually a single learning mechanism. If statistical learning and perceptual learning are different kinds of learning, then conditions that lead to familiarity may have no effect on encoding (and vice versa). If, instead, there is only one general learning mechanism here, not two, or if these two tend to co-occur, then whenever one finds familiarity

[5] This is likely for our definition of SL, but the learning of global statistics may be unrelated to PL. This question is outside of our scope.

improvements (SL), these may be accompanied by detectable improvements in encoding (PL). New research is needed to address these possibilities, researchers' confusion, and clarify the relationship of SL and PL. My dissertation addressed this theoretical gap.

Dissertation Approach. This dissertation addressed the confusion in the literature of the relationship of statistical learning and perceptual learning in humans. We investigated the relationship in a series of experiments using visual stimuli, in which SL and PL paradigms and effects could be directly compared.

One way to compare SL and PL is to use a paradigm known to engender statistical learning and then added a test to see if perceptual learning (encoding improvements) has also occurred (Experiments 1 and 3). I directly tested for perceptual learning via encoding changes in a statistical learning paradigm: I developed and used a novel psychophysical task following a replication of a statistical learning paradigm to assess the presence of encoding changes, which are not theorized to exist in statistical learning. For example, such encoding changes would be improved discrimination of whether two stimuli are the same or different or improved visual search. If improvements were found via the psychophysical task, I would then conduct additional signal detection analyses to assess whether improvements on psychophysical tasks were in sensitivity – additional evidence for PL, or bias – evidence for SL instead. Prior research (Bufford et al., 2014; Thai et al., 2011) has demonstrated that encoding changes due to PL in vision can be directly measured.

A complementary approach to testing for PL effects following SL training is to test for SL effects following PL training (Experiment 2). Relationships between stimuli in PL paradigms are typically not manipulated the way that statistics are controlled in SL studies, so to make this a viable approach, I developed a PL training based on an SL paradigm. This also allowed me to use the same psychophysical task following training as in Experiment 1, which enabled cross-experiment analyses.

CHAPTER 2: EXPERIMENT 1 INTRODUCTION AND METHODS

Experiment 1: Does PL accompany visual SL?

The goal of Experiment 1 is to address the relationship of SL and PL by testing for PL effects caused by an SL paradigm. Specifically, we psychophysically tested for encoding changes caused by a well-known visual statistical learning paradigm of visual arrays (Fiser & Aslin, 2001, Experiment 1). The hypothesis of different kinds of learning (that PL and SL are different learning processes) would predict only SL effects on the psychophysical assessment. In contrast, the hypothesis of a single learning process would predict PL effects following an SL familiarization.

We chose to study visual arrays as opposed to visual sequences because the former are more ecologically valid visual stimuli: humans constantly process scenes, and less frequently process sequences of objects without motion. Experiments directly measuring encoding changes before and after PL-based interventions in complex visual domains such as Chinese characters (Thai et al., 2011) and algebraic equations (Bufford et al., 2014) have shown that psychophysical tasks such as visual search and same/different tasks can capture encoding changes due to PL. Other researchers have documented psychophysical differences in performance between novices and experts because of experts' PL (e.g. Cheng, 2014; Chase & Simon, 1973). Thus, it is likely that a psychophysical approach would capture any encoding changes following SL.

Fiser and Aslin (2001) overall followed the typical scheme of passive familiarization followed by a familiarity test, but not the typical stimuli arrangement of sequentially presented simple shapes with statistically manipulated transitions between shapes. Instead, the authors created 12 new shapes, which were randomized into 6 pairs for each participant individually. Each participant's 6 pairs consisted of two vertical pairs, two horizontal pairs, and two diagonal pairs (one diagonal in each direction). These pairs were assembled into all possible 3 by 3 grids consisting of one vertical pair, one horizontal pair, and one diagonal pair, for a total of 144 grids. Shapes always appeared in their pairs (100% spatial probability of second shape, given first, and vice versa). Participants passively viewed each of their 144 unique grids in a 7 minute period. In the following familiarity test, participants identified their pairs as more familiar than foil pairs at above-chance levels. Thus participants demonstrated spatial statistical learning.

Psychophysically Testing for PL

With its arrays of shapes, this paradigm was well-suited to the psychophysical task of target detection. The arrays functioned as scenes to learn, and pairs could be used as targets following the arrays, to probe for participants' perceptions of the pair objects in the scenes. For this reason, we created a novel psychophysical detection task to assess PL effects following the familiarization.

Performance on shape pairs in the familiarization might have been influenced by learning the shape pairings, perhaps including their exact arrangement, or learning the types of spatial arrangements in which shapes could appear, or both. SL would predict learning of shape pairings and perhaps pair orientations. PL would predict some improvement in encoding: this could consist of selective pickup of pairs or improved discrimination of components (shapes), shown in selective encoding or more rapid detection in short exposures.

However, PL and SL are theoretically distinct and make distinct predictions of what would be learned in Fiser and Aslin's paradigm (2001). SL is learning of correlations between studied items, so it would predict no transfer of learning to new conditions except, perhaps, a bias to expect to continue seeing what was learned. In our psychophysical task, we would see this bias in an elevated false alarm rate for one of the comparison conditions that uses familiarized pairs as targets for target-absent trials. In contrast, PL is improvements in encoding – which could be unitization (chunking) of correlated items, learning of pairs' possible spatial arrangements and/or learning how to segment grids into pairs – causing improved sensitivity in a psychophysical sense.

To distinguish between PL and SL in the experiment, our assessment had three conditions: Familiar, which used the pairs from familiarization, and two comparison conditions. The Shuffled condition used the same shapes as in the familiarization but shuffled into new pairings, to test for near transfer, and for bias to expect what was seen previously. The third condition, New Shapes, utilized a new set of 12 never-before-seen shapes that we created arranged in pairs like the other conditions. New Shapes primarily served as a control, because only learning of how to segment arrays into pairs could transfer to this condition.

Hypotheses

Results would support SL as a separate learning mechanism from PL if participants showed learning on familiarized stimuli (Familiar) but little or none on other stimuli, and showed a psychophysical bias for familiarized stimuli with partial information (elevated false alarm rate in Shuffled). In contrast, results would indicate the presence of PL following an SL paradigm if participants showed transfer to new, closely related stimuli (Shuffled) and possibly to new, distantly related stimuli (New Shapes), and if participants increased their pychophysical sensitivity (in any or all conditions, but especially Familiar).

Method

Participants

90 (67 female, 23 male) undergraduates in psychology and linguistics courses at the University of California, Los Angeles were recruited through the Psychology Department subject

pool. Participants had normal or corrected-to-normal vision and were compensated with partial course credit.

Materials & Procedure

Statistical Learning Paradigm. Experiment 1 replicated Fiser and Aslin (2001)'s paradigm: We re-created the shapes and their pairings and arrangements for the 144 grids presented in 7 minutes in the familiarization, viewed passively. We built the recognition task, with one trial per target (as the number of trials was not specified in the original study). We also replicated the creation of a unique shape set – randomization of shapes to pairs for each individual participant.

We did make two changes: In our version, the 2IFC familiarity test did not include grid lines. The familiarity test had one trial per target, in random order (Fiser & Aslin, 2001 did not specify their number of familiarity test trials). We tripled the familiarization from one block of 7 minutes with the 144 grids to three to 21 minutes, in three 7 minute blocks each with all 144 grids in a different random order, with breaks between. Piloting had showed that increasing familiarization increased recognition test performance – increased learning – to make the learning more detectable.

After participants completed the familiarization and the recognition test, then they participated in our psychophysical assessment[6]. We developed a novel psychophysical detection task to assess perceptual learning in this SL paradigm. In the detection task, participants were

[6] Given that SL occurs quickly, in as little as 7 minutes, we decided against a pretest-posttest design. We did not want participants to learn pairs before the familiarization, and we kept the assessment short (about 10 minutes) to minimize learning during the assessment.

tested on their ability to determine whether a target pair was present or absent in a previously presented grid. Finally, participants completed a brief survey.

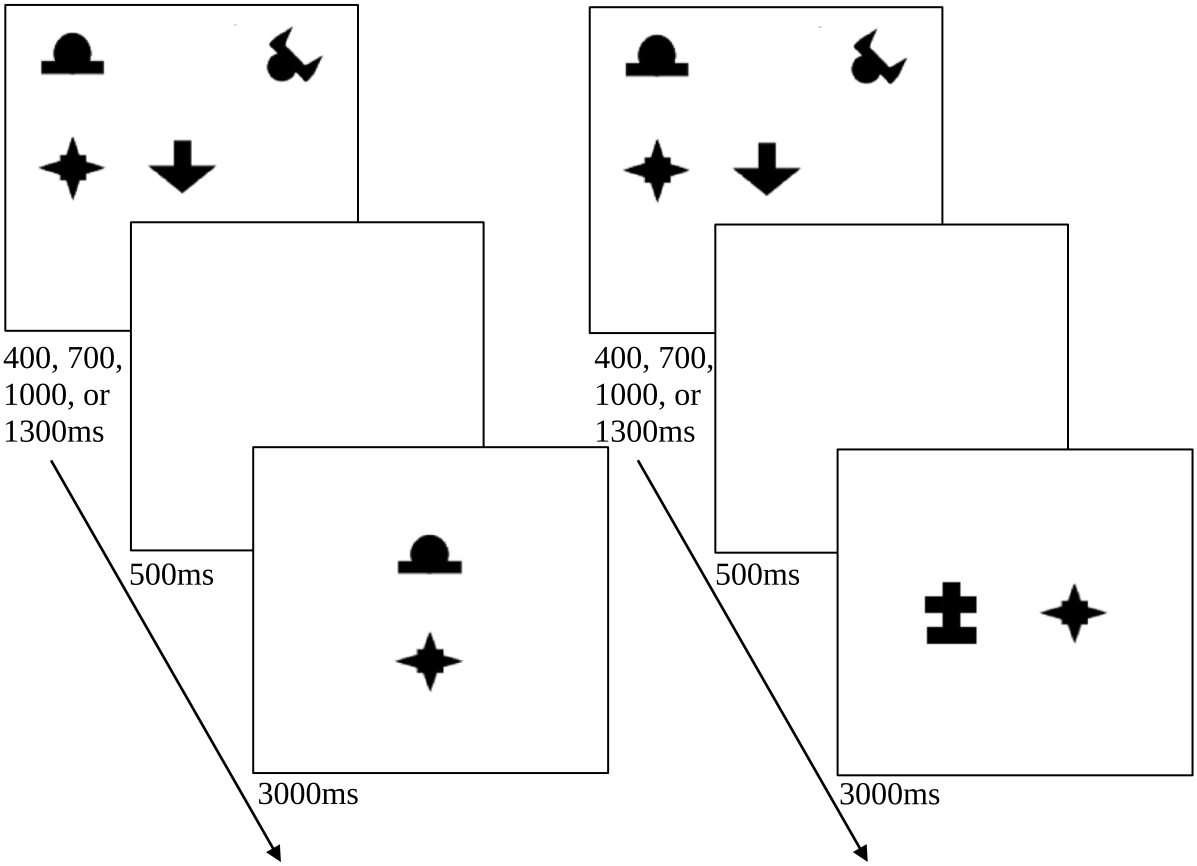

Figure 1. Schematic of psychophysical assessment for a target-present trial (left) and a target absent trial (right).

Psychophysical Detection Task. Each of 80 psychophysical test trials (4 grid Exposure Durations x 2 Target Presence levels (Present, Absent) x 10 trials per exposure duration-presence combination) consisted of a grid, a blank screen (500ms), and the target (3000ms). Exposure duration for grid presentation followed the method of constant stimuli, using the Exposure

Durations 400ms, 700ms, 1000ms, and 1300ms. Piloting indicated that these exposure durations covered a range in which performance showed neither floor nor ceiling effects.

Grids consisted of two differently-oriented pairs of shapes (e.g. one vertical and one diagonal pair), arranged so that the pairs were adjacent in the 3x3 grid, and gridlines were not shown. Targets were also shown without gridlines, centered in the screen. On half of the trials, target shape pairs were present. We ensured that participants could not achieve high accuracy by simply looking at a single target shape: on target-absent trials, targets shared exactly one shape with the shape pairs in the grid. Thus, participants had to identify whether the target pair as a whole was present in the grid. They pressed ‘z’ for present and ‘/’ for absent.

Participants were assigned to one of three versions of the assessment: Familiar (*n* = 29), Shuffled (*n* = 32), or New Shapes (*n* = 29). In the Familiar condition, grids contained pairs presented in familiarization. Targets were either one of the pairs in the grid (target-present trials), or a pair that shared exactly one shape with the grid but belonged to a Shuffled shape set (target-absent trials).

In the Shuffled condition, shapes were randomized to a new shape set of 6 pairs with the following constraints: 1) Shapes could not be paired with the same shape; 2) Shapes could not be in the same position (e.g. the left side of a horizontal pair); 3) The shapes from any two original pairs must be divided into at least three new pairs, so that every possible grid could have a target pair that shared exactly one shape with the grid. Targets for target-absent trials were not from the new shape set, but all 6 pairs from the (participant’s individualized) shape set from familiarization. By using familiarization pairs for target-absent trials, we tested for a bias due to SL: a bias to expect to continue seeing pairs from familiarization. This bias could have manifested as simply choosing targets that had been seen in familiarization without reference to the grid for the trial[7].

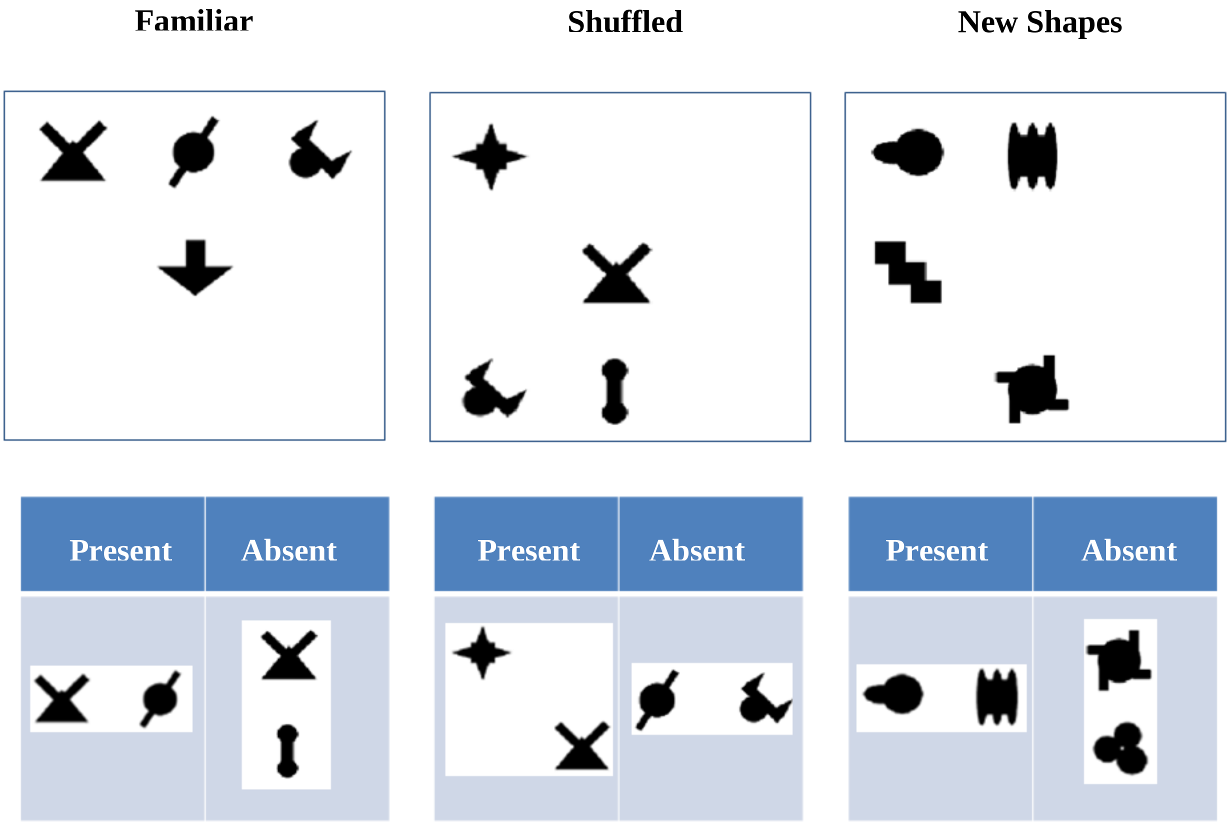

Figure 2. Schematic of relationships between arrays and targets within and across conditions in the psychophysical assessment. For target-present trials, the target pair appeared in the array. For target-absent trials, one of the two shapes in the target pair appeared in the array, but the pair belonged to a different condition – Familiar arrays with Shuffled targets; Shuffled arrays with Familiar targets.

[7] Note that this is one way bias could be detected in this study. Another was the false alarming to displays containing only component of a familiarized pair in the Familiar condition.

In the New Shape trials, twelve new shapes we created were arranged into pairs for grids and target pairs after the fashion of Fiser and Aslin (2001)'s grids. The shapes in the familiarization shape set were replaced with new shapes, and randomization for the Shuffled condition was applied, to create a new shape set of 6 pairs to serve as targets for target-absent trials. Thus, the New Shape condition had 6 target pairs, and 6 target-absent pairs that overlapped with the shapes in grids by exactly one shape, as in the other conditions.

Survey. After the psychophysical task, participants completed a brief survey using Google Forms. The first three items were mandatory: 1. Participants entered their name (to allow their survey to be matched to their behavioral data). 2. Participants reported their hours of sleep the previous night on a 5-point Likert-type scale as "less than 7", 7, 8, 9, or "10 or more". 3. Participants reported how alert they felt from 1 ("not alert at all") to 5 ("extremely alert"). These items were followed by several optional open-ended questions (see Appendix A): "What did you notice during the experiment?"; "Did you have any strategies? If you had any strategies, please describe them."; and "Did you notice any patterns? If you noticed patterns, please describe them." When participants completed the survey, they were released.

Dependent Measures

Familiarity Test. We collected accuracy on each trial, and averaged accuracy across trials, for a single value for each participant. Analyses were conducted on the aggregated data.

Psychophysical Assessment. On each trial, we collected accuracy. The raw trial-by-trial values were averaged for trials with the same Exposure Duration with the same level of Target Presence (Present, Absent). Main analyses were conducted on the aggregated accuracies. I also calculated average accuracy on the assessment to use as the dependent measure for exploratory analyses with survey data and other factors.

Additionally, I calculated raw hit and false alarm rates for each participant from the aggregated data. Then I calculated corrected hit and false alarm rates using the log-linear correction of adding .5 to each cell and dividing by n + 1 (Hautus, 1995; Knoke & Burke, 1980), then calculated sensitivity (d') and bias (criterion[8]) for each participant. Analyses were conducted on (raw) hit rate, (raw) false alarm rate, d', and criterion. A positive value for criterion meant that participants tended to respond that the target was present more often; a negative value meant that participants responded absent more frequently.

Additional Predictors

Survey. I recoded the endpoints of the sleep measure from descriptions into numbers, e.g. from "less than 7" to 6, to allow for quantitative analyses.

I was interested in whether any particular strategy was associated with better performance, so I read participant responses to the strategy question. I found that more than half of participants (*n* = 49) gave a response that indicated naming of the shapes (e.g. "i [*sic*] tried associating each shape with a word or thing that reminded me of the shape"). Each participant was coded dichotomously as either employing the linguistic strategy or not (*n* = 30).

I was also interested in whether explicit knowledge of the pair structure might have influenced performance. In their open-ended responses as to what they noticed and to other optional questions, some participants (*n* = 24) directly mentioned the pair structure or described a pair, e.g. "the star object always comes [with] a triangle shape object". These participants were coded as 1 for Noticing. Others did not explicitly report noticing the pairs, but they reported noticing that certain shapes could only appear in certain locations in the arrays, or that shapes

[8] I used this formula for criterion: criterion = .5*(z(H)+z(FA)).

appeared with other shapes or near certain other shapes. These properties were consequences of the pair structure because shapes were paired and certain shapes could not appear in certain locations. For example, the lower shape in a vertical pair could not appear in the top row. This group of participants (*n* = 11) may have noticed the pairs as well. These participants were coded as 0.5 for Noticing. The remaining participants did not report anything related to the pair structure (*n* = 53) and were coded as 0 for Noticing, or did not report anything about what they noticed at all (*n* = 2, excluded from analyses of noticing).

Other Factors I also noted a few other variables from the experiment log. I categorized the time of day at which the experiment started: morning (before 11am, *n* = 42), midday (at or after 11am but before 2pm, *n* = 36), or later (*n* = 12). I coded dichotomously whether the experiment occurred on a weekday (*n* = 76) or on a weekend (*n* = 14). I also recorded participants' gender to confirm that there were no differences in performance by gender.

CHAPTER 3: EXPERIMENT 1 RESULTS AND DISCUSSION

Results

Recognition

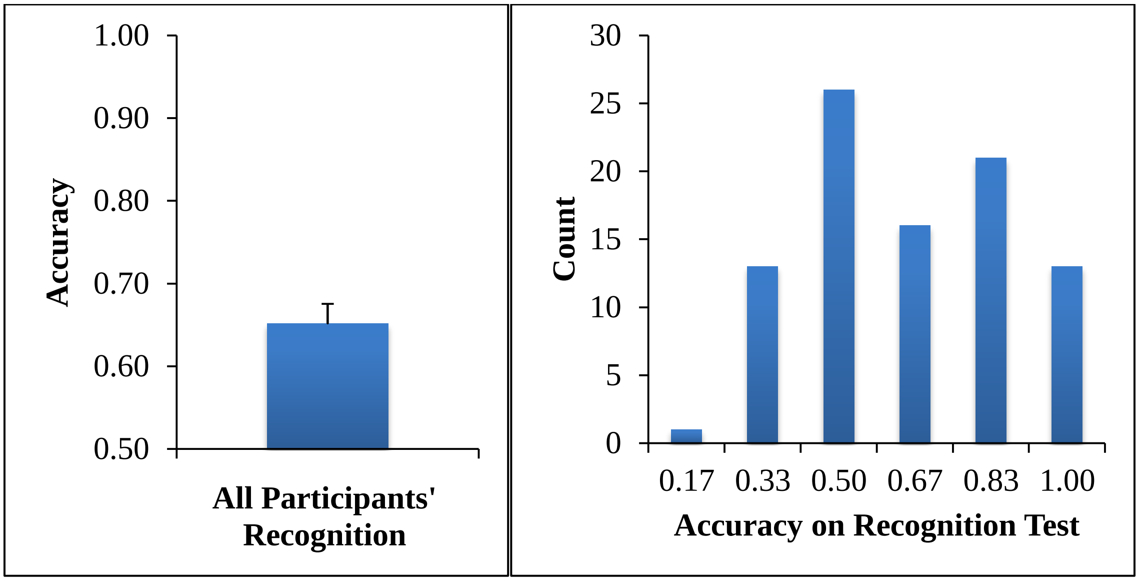

Figure 3. Recognition Accuracy (bar graph, left) and frequency histogram (right). The left panel shows recognition accuracy for all participants. The error bar indicates the standard error of the mean. The right panel is a histogram of accuracies on the recognition test, showing the high variability in accuracy.

Figure 3 shows average recognition performance (left panel) and the distribution of recognition accuracy (right panel). Recognition appears to be above chance, but highly variable across participants. Statistical analyses showed that participants demonstrated accuracy significantly greater than chance (*M* = 0.65, *SE* = 0.02) on recognition of pairs in familiarization in the recognition test, *t*(89) = 6.46, *p* < .001, Cohen's *d* = 0.68. Inspecting the distribution of accuracy for participants (in Figure 3, right panel) suggested that there might have been two populations of participants – those that successfully passed the recognition test and those that did not. To test this hypothesis of two populations, participants were divided into two groups by their accuracy on recognition: “Recognizers” (n = 50) scored above 50% and “Nonrecognizers” (n = 40) scored at or below 50%. An independent-samples t-test of recognition group on recognition accuracy revealed that Recognizers (M = 0.82, SE = 0.02) had significantly higher accuracy than Nonrecognizers (M = 0.44, SE = 0.01) and this was an extremely large effect, $t(88) = 16.147$, $p < .001$[9], Cohen’s d = 3.48.

An ANOVA of Condition on recognition accuracy showed that conditions did not differ in recognition ($p = .75$). Participants in all conditions recognized pairs seen in familiarization equally well. This was unsurprising because all participants received the same recognition test (relative to their familiarization shape set), but it indicated that there were not significant differences across participants in different conditions by chance.

Psychophysical Assessment: Main Analyses

Accuracy. Figure 4 shows participants’ accuracy for each assessment condition at each search array Exposure Duration for both target-present and target-absent trials, and it appears to show a main effect of Condition such that participants performed better in the Familiar and Shuffled conditions than in New Shapes. A three-way ANOVA of Condition (Familiar, Shuffled, New Shapes) by Exposure Duration (400ms, 700ms, 1000ms, 1300ms) by Target Presence (Present, Absent) on accuracy tested the observed pattern of results and revealed a large main

[9] For planned comparisons – tests involving both hypothesis-relevant independent variables (Condition, ExperimentVersion) and hypothesis-relevant dependent variables (recognition accuracy; psychophysical assessment accuracy, false alarm rate, and sensitivity) – the standard uncorrected alpha level of .05 was used throughout this dissertation, and also for other ANOVAs and ANCOVAs. For all other tests, the alpha level of .001 was used (corrected alpha level, approximated to three decimal places).

effect of Condition, *F*(2,87) = 6.70, *p* = .002, *partial-eta-squared* = 0.13, a marginal main effect of Exposure Duration *F*(3,261) = 2.44, *p* = .07, *partial-eta-squared* = 0.02, an interaction effect of Exposure Duration and Target Presence, *F*(3,261) = 7.87, *p* < .001, *partial-eta-squared* = 0.08, and a three-way interaction, *F*(6,261) = 2.37, *p* = .03, *partial-eta-squared* = 0.05. No other effects were significant (all *p*'s > .49).

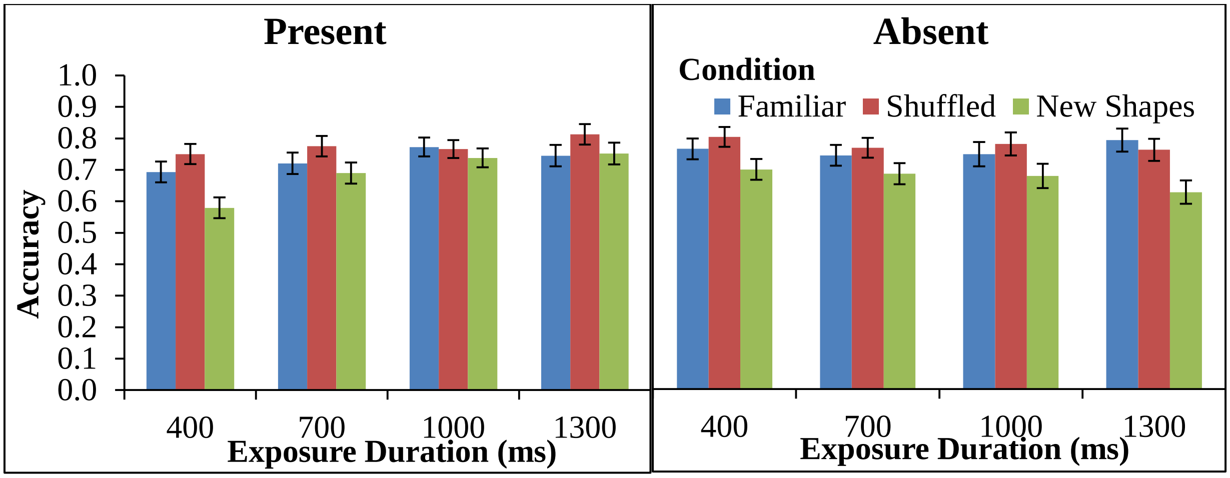

Figure 4. Condition by Exposure Duration by Target Presence on accuracy. Error bars indicate standard error of the mean.

To investigate the marginal main effect of Exposure Duration, I conducted custom hypothesis tests in ANOVA using SPSS. I first compared 1000ms (*M* = 0.75, *SE* = 0.01) and 1300ms (*M* = 0.75, *SE* = 0.02) exposure durations, which did not significantly differ in accuracy (*p* = .93). Then I compared 400ms (*M* = 0.72, *SE* = 0.01) to the average of the 1000ms and 1300ms exposure durations and found that 400ms had marginally lower accuracy, *F*(1,87) = 7.76 *p* = .007, *partial-eta-squared* = 0.08. No other comparisons were significant (all *p*'s > .23). Because the shortest exposure duration had marginally lower accuracy than the longest two, and

Exposure Durations increased numerically in accuracy, accuracy increased monotonically across Exposure Durations. Thus, the results of the assessment are interpretable.

To investigate the main effect of Condition via custom hypothesis tests, I first compared the Familiar (*M* = 0.75, *SE* = 0.02) condition to the Shuffled (*M* = 0.78, *SE* = 0.02) condition which were not significantly different (*p* = .27). Then I compared the combination of the Familiar and Shuffled conditions to the New Shapes (*M* = 0.68, *SE* = 0.02) condition, which revealed that the Familiar and Shuffled conditions were significantly more accurate than the New Shapes condition and that this was a large effect, *F*(1,87) = 11.94, *p* = .001, *partial-eta-squared* = 0.12. Participants were more accurate in the conditions with shapes they saw during familiarization, and learning transferred from Familiar to Shuffled.

To follow up on the significant three-way interaction of Condition, Exposure Duration, and Target Presence, I divided the data by exposure duration to test each simple two-way interaction of Condition and Target Presence using custom hypothesis tests in ANOVA. For 400ms, 700ms, and 1000ms, Condition and Target Presence did not interact (all *p*'s >.13). For 1300ms, Condition and Target Presence interacted, *F*(2,87) = 3.76, *p* = .03, *partial-eta-squared* = 0.08, so I broke the data by Target Presence to investigate the simple simple effect of Condition at 1300ms to follow up on the simple interaction. These custom hypothesis tests at 1300ms revealed a simple simple effect of Condition for Absent, *F*(2,87) = 3.676, *p* = .03, *partial-eta-squared* = 0.08. For when the target was Absent at 1300ms, I compared the Familiar (*M* = 0.79, *SE* = 0.04) and Shuffled (*M* = 0.76, *SE* = 0.04) conditions and found that they were not different from each other (*p* = .55), but when I compared them in combination to the New Shapes (*M* = 0.63, *SE* = 0.04) condition, I found that Familiar and Shuffled were significantly more accurate than New Shapes, *F*(1,87) = 11.17, *p* = .001, *partial-eta-squared* = 0.11. There was no simple simple effect of Condition when the target was Present (*p* = .29). The three-way interaction of Condition and Exposure Duration and Target Presence was driven by the Condition effect at 1300ms for Absent.

I examined the two-way interaction of Exposure Duration and Target Presence using custom hypothesis tests, breaking on Exposure Duration. For 400ms, accuracy was marginally higher when the target was Absent (*M* = 0.76, *SE* = 0.02) than when it was Present (*M* = 0.67, *SE* = 0.02), *F*(1,87) = 8.54, *p* = .004, *partial-eta-squared* = 0.09. Accuracy did not differ on TargetPresence for the other exposure durations (all *p*'s > .10), so the interaction of Exposure Duration and Target Presence was driven by higher accuracy for Absent than Present at 400ms.

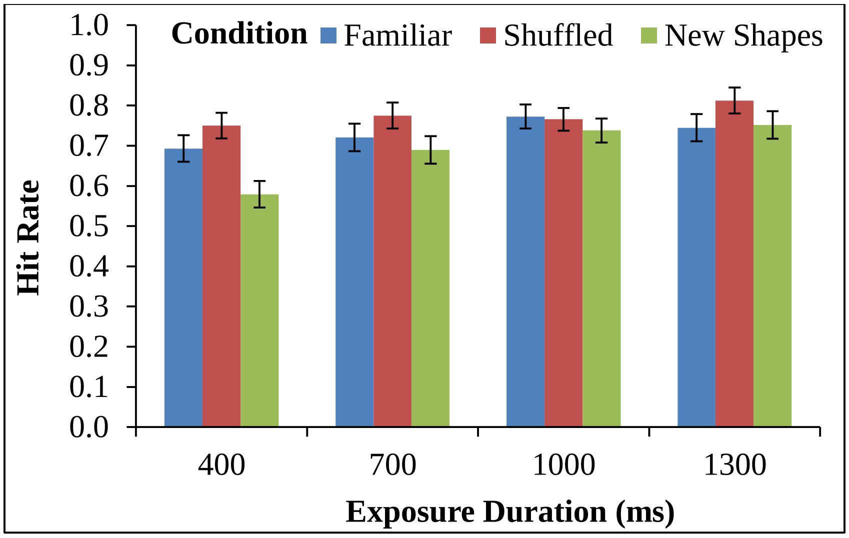

Figure 5. Condition by Exposure Duration on hit rate. Error bars indicate standard error of the mean.

Hit Rate. Figure 5 showed effects of Condition and Exposure Duration on hit rate, and appeared to show higher hit rates for Familiar and Shuffled across exposure durations, especially for Shuffled except at 1000ms. To test these apparent effects, I conducted an ANOVA of

Condition by Exposure Duration on hit rate, which showed that there was a significant effect of Exposure Duration, *F*(3, 261) = 9.70, *p* < .001, *partial-eta-squared* = 0.10, an interaction of Condition and Exposure Duration, *F*(6, 261) = 2.19, *p* = .04, *partial-eta-squared* = 0.05, and a marginal main effect of Condition, *F*(2,87) = 3.04, *p* = .05, *partial-eta-squared* = 0.07.

Using custom hypothesis tests in ANOVA, I followed up on significant effects. For the main effect of Condition, I conducted all pairwise comparisons. There were no significant differences (all *p*'s > .01).

For the interaction of Condition and Exposure Duration, I examined the simple effect of Condition at each search grid exposure duration. There was a significant simple effect of Condition at 400ms, *F*(2,87) = 7.05, *p* = .001, *partial-eta-squared* = 0.14. At 400ms, I compared Familiar (*M* = 0.69, *SE* = 0.03) and Shuffled (*M* = 0.75, *SE* = 0.03), which did not differ (*p* = .22), but when I next compared them together to New Shapes (*M* = 0.58, *SE* = 0.03), Familiar and Shuffled in combination they showed significantly more hits than New Shapes, *F*(1,87) = 12.32, *p* = .001, *partial-eta-squared* = 0.12. There were no simple effects of Condition for the other exposure durations (all *p*'s > .18), so the interaction of Condition and Exposure Duration was driven by the simple Condition effect of transfer from Familiar to Shuffled at 400ms.

I also examined the main effect of Exposure Duration on hit rate via custom hypothesis tests. When compared, I found 1300ms (*M* = 0.77, *SE* = 0.02) and 1000ms (*M* = 0.76, *SE* = 0.02) did not differ (*p* = .54). I then compared 400ms (*M* = 0.67, *SE* = 0.02) to 1000ms and 1300ms together, and found that 400ms had a lower hit rate than the combination of 1000ms and 1300ms, *F*(1,87) = 31.67, *p* < .001, *partial-eta-squared* = 0.27. I found that the combination of 1000ms and 1300ms did not differ in hit rate from 700ms (*p* = .046) when I tested this comparison. I also found that 400ms and 700ms (*M* = 0.73, *SE* = 0.02) did not differ (*p* = .02).

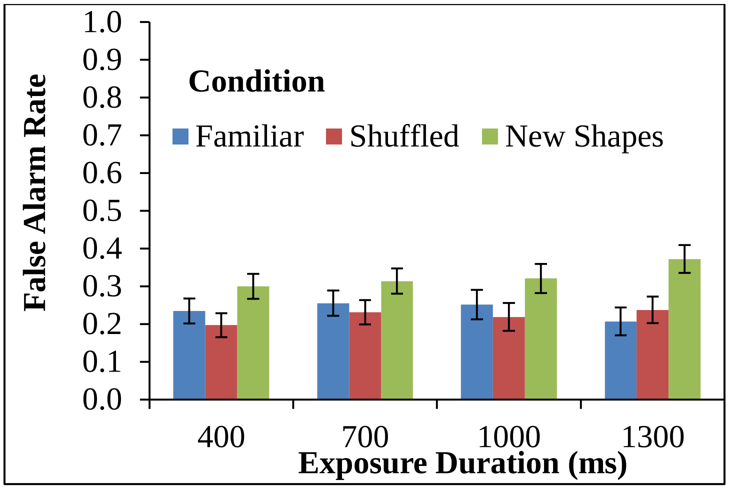

Figure 6. Condition by Exposure Duration on false alarm rate. Error bars indicate standard error of the mean.

False Alarm Rate. Figure 6 showed the effects of Condition and Exposure Duration on false alarm rate, and it appeared that New Shapes had the highest false alarm rate across exposure durations. An ANOVA of Condition by Exposure Duration on false alarm rates confirmed this pattern, by revealing a main effect of Condition, *F*(2,87) = 4.17, *p* = .02, *partial-eta-squared* = 0.09 and no other effects (all *p*'s > .38). Using custom hypothesis tests, I compared Familiar (*M* = 0.24, *SE* = 0.03) to Shuffled (*M* = 0.22, *SE* = 0.03) and found that they did not differ in false alarm rate (*p* = .68), but when I compared their average to New Shapes (*M* = 0.33, *SE* = 0.03), I found that Familiar and Shuffled combined had significantly lower false alarm rates than New Shapes, *F*(1,87) = 8.09, *p* = .006, *partial-eta-squared* = 0.09. Learning in terms of decreased false alarming transferred from Familiar to Shuffled.

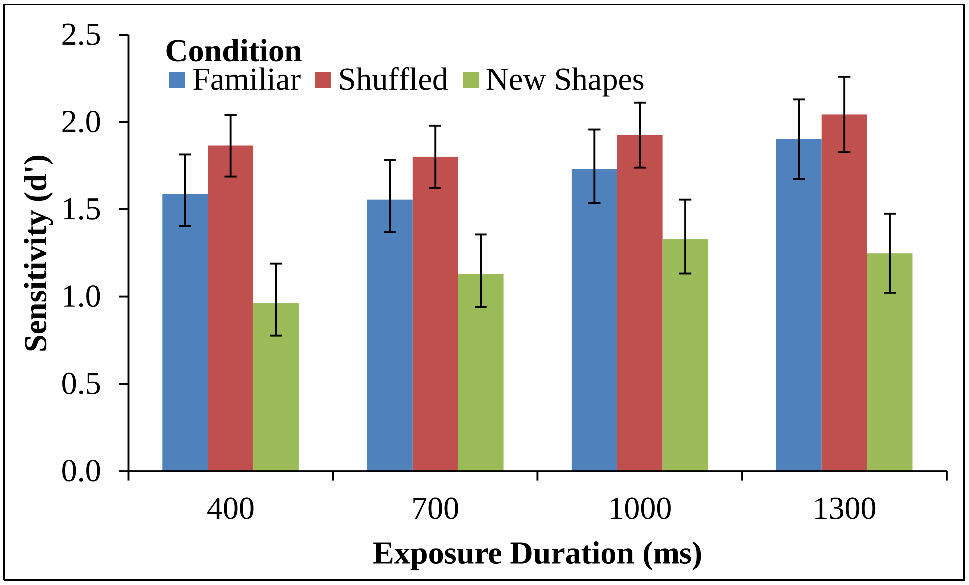

Figure 7. Condition by Exposure Duration on sensitivity. Error bars indicate standard error of the mean.

Sensitivity. Figure 7 showed the effects of Condition and Exposure Duration on sensitivity. There appeared to be a strong main effect of higher sensitivity for Familiar and Shuffled than for New Shapes. To statistically test this pattern of results, I conducted an ANOVA of Condition by Exposure Duration on sensitivity, which revealed a main effect of Condition, *F*(2,87) = 6.39, *p* = .003, *partial-eta-squared* = 0.13, and a marginal main effect of Exposure Duration, *F*(3,261) = 2.50, *p* = .06, *partial-eta-squared* = 0.03, but no interaction (*p* = .79). I used custom hypothesis tests in ANOVA to follow up on the main effect of Condition. I first compared Familiar (*M* = 1.39, *SE* = 0.12) and Shuffled (*M* = 1.56, *SE* = 0.11) and found that they did not differ (*p* = .29). When I next compared Familiar and Shuffled combined to New Shapes (*M* = 0.98, *SE* = 0.12), I found that sensitivity was significantly higher for the combination of Familiar and Shuffled than for New Shapes, *F*(1,87) = 11.45, *p* = .001, *partial-*

eta-squared = 0.12. Learning transferred from Familiar to Shuffled. I investigated the significant main effect of Exposure Duration via custom hypothesis tests of all pairwise comparisons. These revealed no differences (all *p*'s > .01).

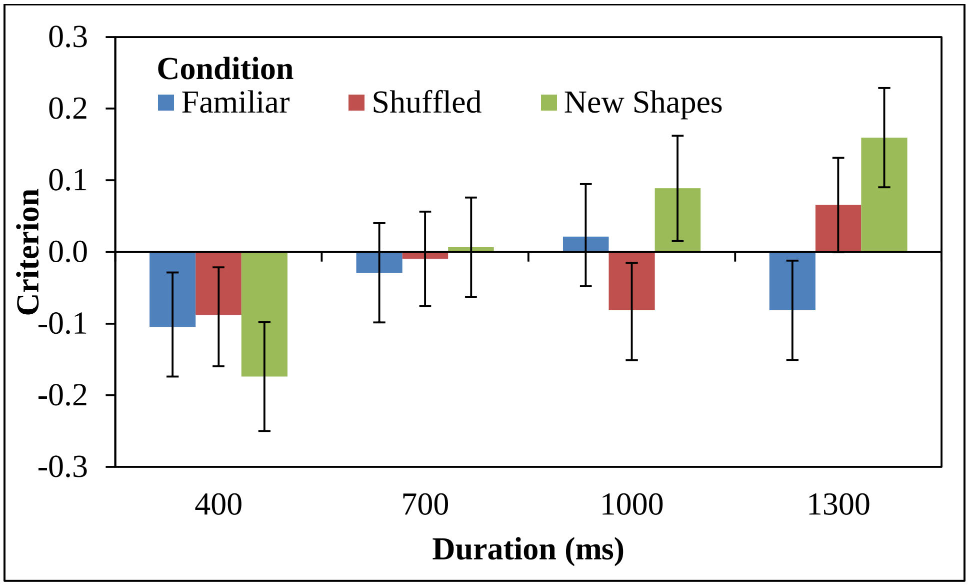

Figure 8. Condition by Exposure Duration on criterion. Error bars indicate standard error of the mean.

Bias. Figure 8 showed the effects of Condition and Exposure Duration on bias as measured by criterion. In the psychophysical assessment, bias was a response tendency to say absent or present. Criterion measured this tendency in terms of the distance in standard deviations from the unbiased criterion (equal numbers of present and absent responses), with negative criterion indicating more absent responses. Looking at Figure 8, it appeared that participants' criterion varied with Exposure Duration. A two-way mixed ANOVA of Condition and Exposure Duration on criterion confirmed the apparent pattern by revealing a main effect of

Exposure Duration, *F*(3,261) = 5.95, *p* = .001, *partial-eta-squared* = 0.06, and an interaction, *F*(6,261) = 2.40, *p* = .03, *partial-eta-squared* = 0.05, and no main effect of Condition (*p* > .66).

To investigate the interaction, I conducted custom hypothesis tests of the simple effect of Condition at each Exposure Duration. No simple effect of Condition was found for 400ms (*p* = .69), 700ms (*p* = .94), or 1000ms (*p* = .24). For 1300ms, there was a marginal simple effect of Condition, *F*(2,87) = 3.08, *p* = .05, *partial-eta-squared* = 0.07. At 1300ms, I compared conducted all pairwise comparisons using custom hypothesis tests, and found that Familiar (*M* = -0.08, *SE* = 0.07) showed significantly more bias to say absent than New Shapes (*M* = 0.16, *SE* = 0.07), *F*(1,87) = 6.06, *p* = .02, *partial-eta-squared* = 0.07. Shuffled (*M* = 0.07, *SE* = 0.07) at 1300ms did not differ in bias from Familiar (*p* = .13) or New Shapes (*p* = .33). The interaction of Condition and Exposure Duration was driven by the transfer of learning from Familiar to Shuffled at 1300ms.

Custom hypothesis tests in ANOVA were conducted to follow up on the main effect of Exposure Duration. I first compared 700ms (*M* = -0.01, *SE* = 0.04) and 1000ms (*M* = 0.01, *SE* = 0.04) and found that they did not differ in criterion (*p* = .15). Then I compared the middle exposure durations combined to 1300ms (*M* = 0.05, *SE* = 0.04) and found that the longest three exposure durations did not differ (*p* = .18). Finally, I compared 400ms (*M* = -0.12, *SE* = 0.04) to the other exposure durations combined, and found that 400ms showed significantly more bias to respond absent than the longer exposure durations combined, *F*(1,87) = 14.34, *p* < .001, *partial-eta-squared* = 0.14.

Psychophysical Assessment: Additional Analyses

Recognition Accuracy and Assessment Accuracy. A Pearson correlation was used to investigate the relationship of recognition accuracy and average accuracy on the psychophysical assessment. Assessment accuracy across conditions and recognition accuracy were not correlated *r*(89) = 0.11, *p* = .31. When data were divided by participants' assessment condition, there was no correlation between assessment accuracy and recognition accuracy for Familiar, *r*(28) = .18, *p* = .36; or Shuffled, *r*(31) = .04, *p* = .81; or New Shapes, *r*(28) = .22, *p* = .26.

Comparison to Baseline. 49 additional participants (41female, 8 male) participated in just the psychophysical detection assessment (no familiarization or familiarity test) to establish baseline performance on the task. 18 were in the Familiar condition, 16 were in the Shuffled condition, and 17 were in the New Shapes condition. I re-ran the above ANOVA analyses including this sample, with Experiment Version (Baseline, Experiment 1) as a between-subjects factor, for the hypotheses-relevant dependent variables (accuracy, false alarm rate, and sensitivity).

Accuracy.. Figure 9 showed effects of Experiment Version, Condition, and Exposure Duration on accuracy for Experiment 1 and Baseline. Performance appeared to be higher for Experiment 1 for Familiar and Shuffled than for Baseline. An ANOVA of Experiment Version by Condition by Exposure Duration by Target Presence on accuracy confirmed this apparent pattern, and revealed a significant main effect of Experiment Version, such that accuracy was significantly higher for Experiment 1 (*M* = 0.74, *SE* = 0.01) than for the Baseline sample (*M* = 0.70, *SE* = 0.02), *F*(1,133) = 4.60, *p* = .03, *partial-eta-squared* = 0.03. Participants demonstrated learning in terms of higher accuracy in the full experiment relative to those participants who only completed the assessment. The ANOVA also revealed a trending interaction of Experiment Version and Condition, *F*(2,133) = 2.67, *p* = .07, *partial-eta-squared* = 0.04, a significant main effect of Exposure Duration, *F*(3,399) = 6.10, *p* <.001, *partial-eta-squared* = 0.04, and an interaction of Exposure Duration and Target Presence, *F*(3,399) = 8.19, *p* <.001, *partial-eta-squared* = 0.06. No other effects were significant (all *p*'s > .12).

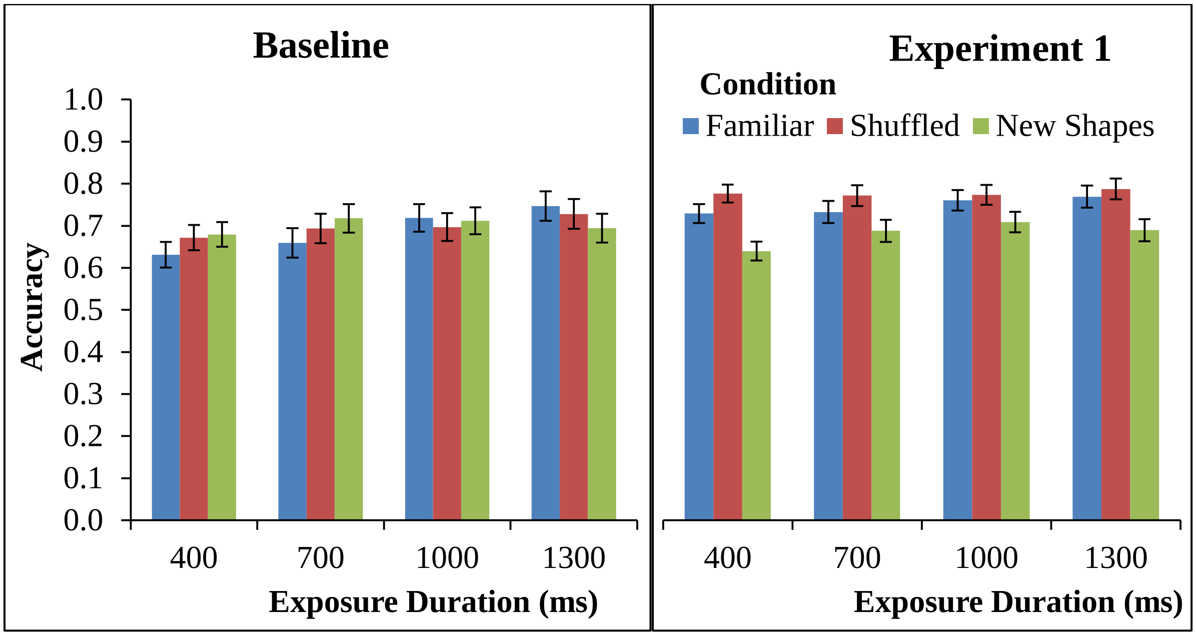

Figure 9. Experiment Version by Condition by Exposure Duration on accuracy (collapsed across Target Presence). Error bars indicate standard error of the mean.

To understand the above significant effects, I conducted custom hypothesis tests. For the trending interaction of Experiment Version and Condition, I looked at the simple effect of Experiment Version on accuracy for each condition. For Familiar, participants were marginally more accurate in Experiment 1 (*M* = 0.75, *SE* = 0.02) than Baseline assessment performance (*M* = 0.69, *SE* = 0.03), *F*(1,133) = 3.28, *p* = .07, *partial-eta-squared* = .02. For Shuffled, participants were also more accurate in Experiment 1 (*M* = 0.78, *SE* = 0.02) than at Baseline (*M* = 0.70, *SE* = 0.03), *F*(1,133) = 6.23, *p* = .01, *partial-eta-squared* = .05. Accuracy for New Shapes in Experiment 1 did not differ from Baseline (*p* = .55). The marginal interaction of Experiment

Version and Condition was driven by learning in the Familiar and Shuffled conditions relative to Baseline and no learning relative to Baseline for New Shapes.

To test the interaction of Exposure Duration and Target Presence, I looked at the simple effect of Target Presence at each exposure duration individually. At 1300ms, Present (*M* = 0.77, *SE* =0.02) was significantly more accurate than Absent (*M* = 0.70, *SD* = 0.02), *F*(1,133) = 11.15, *p* = .001, *partial-eta-squared* = .08. No differences were found at 400ms, 700ms, or 1000ms (*p*'s > .07). The interaction of Exposure Duration and Target Presence was driven by the simple effect of Target Presence at 1300ms.

Following up on the main effect of Exposure Duration, I first compared 1000ms (*M* = 0.73, *SE* = 0.01) and 1300ms (*M* = 0.74, *SE* = 0.01), which did not differ in accuracy (*p* = .51). Then I compared 400ms (*M* = 0.69, *SE* = 0.01) to the combined longer exposure durations and found that 400ms had significantly lower accuracy than the combined longest exposure durations, *F*(1,133) = 20.74, *p* < .001, *partial-eta-squared* = 0.14. Then I compared the two highest exposure durations to 700ms (*M* = 0.71, *SE* = 0.01), and found no difference in accuracy (*p* = .08). Comparing the shortest exposure durations revealed no differences (*p* = .06).

False Alarm Rate. Figure 10 shows the difference in false alarm rates between Experiment 1 and Baseline by Condition and Exposure Duration. It appeared that participants in Experiment 1 showed a lower false alarm rate in Familiar and Shuffled than at Baseline, but did not differ for New Shapes. I tested these apparent effects in an ANOVA of Experiment Version by Condition by Exposure Duration on false alarm rate, which revealed a significant main effect of Experiment Version, such that Experiment 1 (*M* = 0.26, *SE* = 0.02) had a significantly lower false alarm rate than Baseline (*M* = 0.33, *SE* = 0.02), *F*(1,133) = 5.63, *p* = .02, *partial-eta-squared* = 0.04. It also revealed a marginal main effect of Condition, *F*(2,133) = 2.75, *p* = .06, *partial-eta-squared* = 0.04, and no other effects (all *p*'s >.11), so there was no interaction of Experiment Version and Condition.

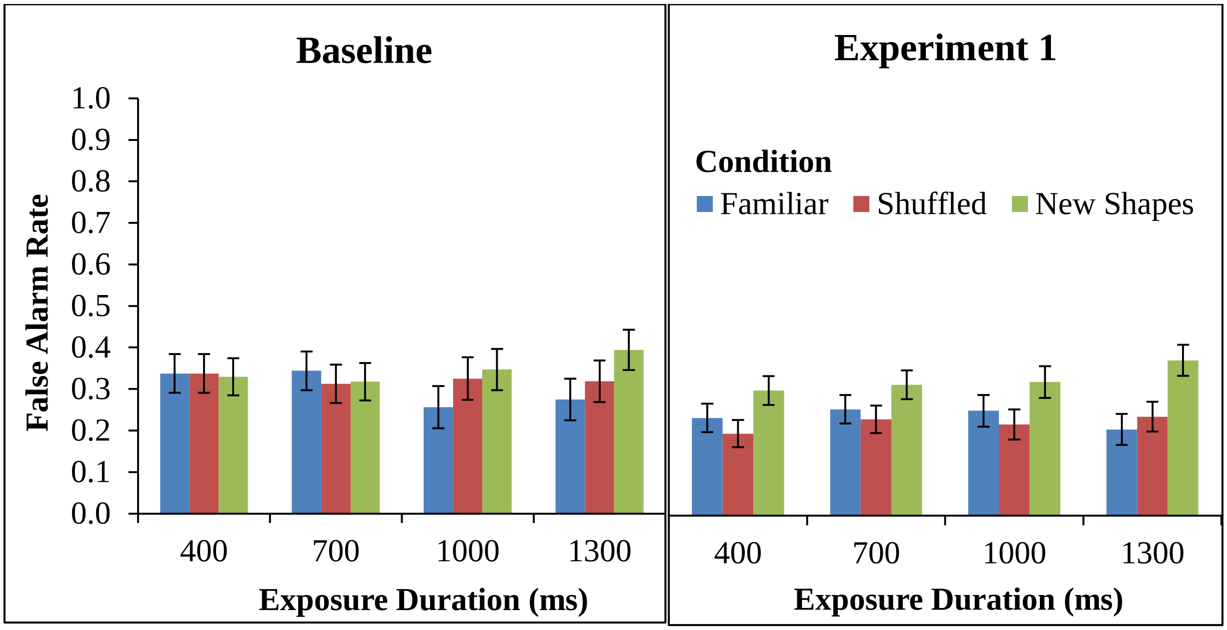

Figure 10. Experiment Version by Condition by Exposure Duration on false alarm rate. Error bars indicate standard error of the mean.

I used custom hypothesis tests in ANOVA to investigate the main effect of Condition. When I compared Familiar (*M* = 0.27, *SE* = 0.02) and Shuffled (*M* = 0.27, *SE* = 0.02), I found that they did not differ in false alarm rate (*p* = .95). When I combined them and compared them to New Shapes (*M* = 0.34, *SE* = 0.02), Familiar and Shuffled demonstrated significantly lower false alarm rates than New Shapes, *F*(1,133) = 5.50, *p* = .02, *partial-eta-squared* = 0.04. Learning in terms of reduced false alarming transferred from Familiar to Shuffled.

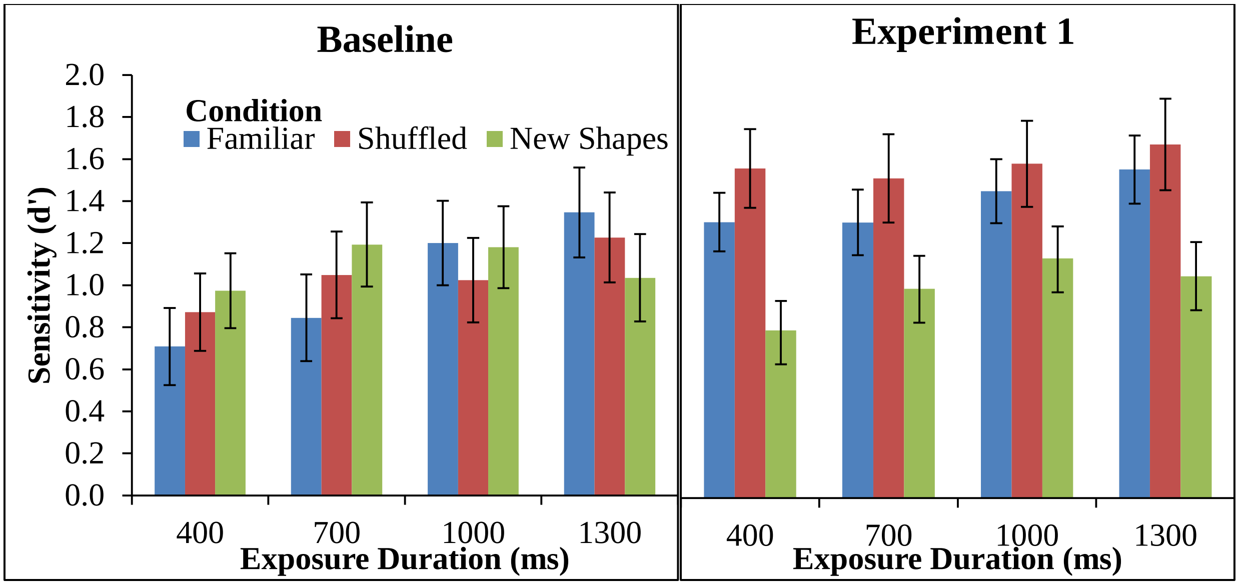

Figure 11. Experiment Version by Condition by Exposure Duration on sensitivity. Error bars indicate standard error of the mean.

Sensitivity. Figure 11 shows sensitivity by Experiment Version and Condition and Exposure Duration. Sensitivity in Experiment 1 for Familiar and Shuffled appeared to be higher than sensitivity at Baseline, and an ANOVA analysis of Experiment Version by Condition by Exposure Duration on d’ confirmed this: it revealed a main effect of Experiment Version, such that participants demonstrated significantly higher sensitivity in Experiment 1 (*M* = 1.31, *SE* = 0.07) than at Baseline (*M* = 1.06, *SE* = 0.09), *F*(1,133) = 8.16, *p* = 0.03, *partial-eta-squared* = 0.04. It also revealed a main effect of Exposure Duration *F*(3,399) = 5.67, *p* = .001, *partial-eta-squared* = 0.04, and a marginal interaction between Experiment Version and Condition, *F*(2,133) = 2.96, *p* = .06, *partial-eta-squared* = 0.04. No other effects were significant (all *p*’s > .14).

For the marginal interaction of Experiment Version and Condition, I tested the simple effect of ExperimentVersion for each condition. For Familiar, there was a marginal simple effect of Experiment Version, such that Experiment 1 (*M* = 1.39, *SE* = 0.12) was marginally more sensitive than Baseline (*M* = 1.03, *SE* = 0.16), *F*(1,133) = 3.39, *p* = .07, *partial-eta-squared* = 0.03. Similarly, there was a significant simple effect of Experiment Version for Shuffled: participants in the Experiment 1 (*M* = 1.56, *SE* = 0.12) were more sensitive than those in the Baseline group (*M* = 1.04, *SE* = 0.16), *F*(1,133) = 7.26, *p* = .008, *partial-eta-squared* = .05. In contrast, there was no simple effect for New Shapes (*p* = .54). The marginal interaction of Experiment Version and Condition was driven by Experiment 1 showing learning relative to baseline in both Familiar and Shuffled.

To investigate the interaction and the main effect of Exposure Duration, I conducted custom hypothesis tests in ANOVA. I first compared the shorter two exposure durations, 400ms (*M* = 1.03, *SE* = 0.07) and 700ms (*M* = 1.04, *SE* = 0.07), and found that they did not differ in sensitivity (*p* = .12). The longer exposure durations, 1000ms (*M* = 1.25, *SE* = 0.07) and 1300ms (*M* = 1.31, *SE* = 0.08), also did not differ in sensitivity (*p* = .54). However, participants showed significantly greater sensitivity on the long exposure durations combined than on the short exposure durations combined, *F*(1,133) = 12.59, *p* = .001, *partial-eta-squared* = 0.09.

Survey Data

Noticing the Pairs. Noticing the pair structure could have improved participants' performance, if noticing them helped them parse the grids. If, however, explicit noticing interfered with automatic, implicit noting, then noticing the pairs could have worsened participants' performance. In fact, a Pearson correlations showed no relationship between noticing and recognition accuracy (*p* = .52) or noticing and psychophysical assessment accuracy (*p* = .04).

Strategy.. Even though Linguistic Coding was the most popular strategy (see Dependent Measures) and it could have made encoding the grids more efficient, Linguistic Coding did not impact performance on either the recognition (*p* = .81) or the assessment (*p* = .42).

Alertness and Sleep. Participants reported a medium level of alertness (*M* = 3.03, *SE* = 0.09). Alertness did not correlate with recognition (*p* = .87) or assessment (*p* = .18) accuracy. The modal response for hours slept the prior night was "less than 7" (*n* = 50). Given the lack of variability, it was not surprising that no correlation of sleep and recognition accuracy (*p* = .90) or sleep and psychophysical assessment accuracy (*p* = .50) accuracy was found.

Other Factors

We confirmed via independent-samples t-tests that gender had no effect on recognition performance (*p* = .71) and no effect on average accuracy in the assessment (*p* = .88). There was also no effect of whether the experiment was administered on a weekday or weekend on either recognition accuracy (*p* = .87) or assessment accuracy (*p* = .35). One-way ANOVAs of time of day on recognition accuracy (*p* = .58) and assessment accuracy (*p* = .27) also showed no effects.

Discussion

Experiment 1 was designed to examine the relation between perceptual learning (PL) and statistical learning (SL) in a well-known SL paradigm (Fiser & Aslin, 2001). We successfully replicated the method and finding of the paradigm (Fiser & Aslin, 2001), and we developed a novel psychophysical detection task to follow it. We found a decreased – not increased – false alarm rate in the Familiar and Shuffled conditions, transfer from Familiar to Shuffled, and improved sensitivity in both Familiar and Shuffled.

Comparing our Data to the Hypotheses

On our task, the hypothesis that SL and PL are different kinds of learning predicted SL patterns of results following an SL familiarization. These SL patterns of results would be showing familiarity on the recognition test, and possibly also an elevated false alarm rate with the same shapes randomized into new pairs (Shuffled condition). In contrast, the hypothesis that SL and PL are a single, integrated learning process predicted a high correlation of recognition and assessment accuracy, transfer to Shuffled from Familiar, and increased sensitivity..

Assessment Performance. Participants showed higher accuracy, higher hit rates, higher sensitivity, and lower false alarm rates for the Familiar and Shuffled conditions than for shapes introduced in the assessment (New Shapes condition). Participants were able to transfer their learning of the familiarization pairs to the Shuffled condition. Participants showed this learning across all exposure durations, but especially in the lower exposure durations and on target-absent trials, where there was more room for improvement. Participants showed similar higher levels of performance on Familiar and Shuffled trials relative to New Shapes trials for target-present and target-absent trials, except at the longest array on-screen exposure duration (1300ms), when the Condition effect was only observed for target-absent trials. Participants improved most in hit rate at the lower exposure durations, but showed improvement in false alarm rate across all exposure durations. With longer exposure durations, participants showed higher sensitivity (d').

It was a bit surprising that performance on Familiar was not better than performance on Shuffled, because PL predicts the most learning on what was trained. Perhaps target-absent trials were more difficult than target-present trials. If so, this may have explained why performance was not different on Shuffled and Familiar trials. The six familiarization pairs were used as targets in the Shuffled condition for target-absent trials. All participants learned the pairs in familiarization, so those in the Shuffled condition may have more easily identified these pairs as absent than those in the Familiar condition ruled out unfamiliar targets on their target-absent trials and thus successfully completed the harder trials at a higher rate. It is also possible that participants improved their encoding of shapes but not pairs via familiarization, and so showed similar improvements in encoding in both Familiar and Shuffled. This experiment cannot distinguish these hypotheses of what was learned.

Recognition Test. Almost half of participants (40 of 90) scored at or below chance on the traditional test of statistical learning, the recognition test. However, there were no differences between those who passed the recognition test and those who did not in assessment performance: PL effects were found for all participants. Additionally, recognition accuracy did not correlate with assessment accuracy. This suggests that more-explicit recognition tests might not capture learning as well as implicit tasks, such as our psychophysical assessment. If PL and SL always co-occurred, then we would have seen a correlation. The fact that we did not find a correlation but that we did observe PL effects following a SL paradigm suggests that the relationship between SL and PL might be nuanced.

Familiar and Shuffled Performance above Baseline. Analyses of the full experiment strongly suggested learning and PL in particular, but we could not rule a difference in difficulty of the different shapes or decreased performance on New Shapes on that basis alone. To confirm that our effects were increased learning, we compared participants' performance on the assessment to Baseline group performance – performance of a separate sample of participants who only completed the assessment without any familiarization. These analyses revealed that: 1. Baseline performance did not differ across conditions, 2. New Shapes performance following familiarization was not different from Baseline and 3. Familiar and Shuffled following familiarization had significantly higher performance than Baseline. Participants showed learning on the conditions with shapes and/or pairs in the familiarization but no learning on the novel shapes and their pairings. We ruled out the alternative explanations of difficulty differences for the shapes and of decreased New Shapes performance, and confirmed our interpretation of the results: because participants were more sensitive to the shapes and/or pairs following familiarization, they were better able to encode the arrays and complete the psychophysical assessment in the Familiar and Shuffled conditions than in the New Shapes condition. Performance on the New Shapes condition was no different than Baseline because participants' encoding of learned shapes and/or pairs did not benefit them on unfamiliar shapes.

Exploratory Analyses

I conducted several exploratory analyses, to investigate possible roles of other factors, including strategies and knowledge of the pair structure. The most commonly reported strategy was giving names or descriptions to the shapes, but using the strategy was not associated with a change in performance on either the recognition test or the psychophysical assessment. Because the shape pairs were what participants were to learn, explicitly noticing them or noticing an effect of them on the arrangement of shapes might be expected to cause higher performance, but noticing was not significantly correlated with accuracy in either the recognition test or in the psychophysical assessment.

Summary

In sum, we found direct psychophysical evidence for PL in encoding changes due to an SL familiarization paradigm. Our findings were more consistent with the hypothesis of a unified learning process than with the hypothesis of distinct kinds of learning. However, our findings were not entirely consistent with either hypothesis, and suggested that the relationship between

SL and PL may be more nuanced. Because of the confusion in the literature and lack of prior research directly addressing the relationship between SL and PL, this is a novel and important finding.

Next Steps

It is important to replicate and expand on Experiment 1 because of its theoretical importance and novel finding, and to better understand the scope of effects observed in this experiment. It would be interesting to learn under what conditions PL effects in an SL paradigm might strengthen, weaken or do not hold. We tripled the familiarization used in the original paper (Fiser & Aslin, 2001), so it is possible that weaker effects might be found with less familiarization, and, perhaps, stronger effects with more familiarization. Given that unitization is part of PL and that recording reliable statistical relationships between base elements (SL) could support unitization, it is possible that when learning is weak or incomplete, we might find SL effects instead of PL effects, weaker PL effects, or a mix of both.

We might expect stronger PL effects, which might allow us more insight as to whether shapes or shape pairs were learned, under several different circumstances. One would be longer familiarization, as previously mentioned. We would also expect stronger PL effects following a PL-based intervention, which would provide converging evidence about the relationship between SL and PL.

CHAPTER 4: EXPERIMENT 2 INTRODUCTION AND METHODS

Experiment 2: How does a PL intervention compare to SL?

Experiment 2 expanded on Experiment 1 in several ways. We investigated potentially strengthening effects found in Experiment 1 by creating a PL intervention using the pairs in the Fiser and Aslin (2001) paradigm, and then administering our psychophysical assessment. Additionally, we investigated whether PL and SL might co-occur in a PL paradigm by including a non-target (SL) pair constructed like targets in the training.

In Experiment 1, we found evidence of perceptual learning (PL) following a statistical learning (SL) paradigm. Following a passive familiarization, participants demonstrated transfer to the Shuffled condition and the increase in sensitivity in both the Familiar and Shuffled conditions, relative to participants who did not have the familiarization and relative to the New Shapes condition. Transfer and increased sensitivity are not consistent with SL (or the hypothesis of different kinds of learning), but are signatures of PL.

Because this was a novel finding addressing an important gap in the human learning literatures, we investigated the conditions under which the effect may strengthen or weaken. One possible way to strengthen the effect would be to give participants PL training instead of just passive familiarization, as in typical SL paradigms including Experiment 1. Experiment 2 was designed to test whether PL training would yield quantitatively larger but not qualitatively different effects, to seek further evidence that SL and PL are, in fact, a unified learning process.

PL in the real world does not require PL training. However, PL-based interventions can speed the process of PL. To this end, we designed and built a PL intervention to train participants on shape pairs in the familiarization. Our PL training consisted of a visual search task with feedback: participants were shown a target shape pair, then searched for it in a search array, and then received feedback. In contrast to a passive SL familiarization of fixed length, our PL intervention required participants to interact with domain structure and receive feedback on each learning trial, until objective mastery criteria were reached. Due to the active nature and mastery criteria of the PL condition, we might find quantitatively larger effects following the PL training than following SL familiarization. However, we would expect the PL training to produce qualitatively the same pattern of results as found with familiarization - PL effects. If we find what we expect, then this would constitute converging evidence towards clarifying the relationship of SL and PL.

Method

Participants

70[10] participants[11] (54 female, 16 male) undergraduates in psychology and linguistics courses at the University of California, Los Angeles were recruited through the Psychology Department subject pool. Participants had normal or corrected-to-normal vision and were compensated with partial course credit. Eighteen participants were excluded due to programming errors: 10 with missing or incomplete data, and 8 because their percentage of SL pair trials was increased[12].

[10] Power analyses in GPower 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007) based on pilot data indicated that a sample size of 20-25 per between-subjects condition would be sufficient.

[11] Data from 11 of these participants, all in the Familiar condition, were collected later to balance the conditions. They were needed because the 7 participants that were excluded due to a program bug that crashed the assessment were all in the Familiar condition.

[12] This may have increased the difficulty of the training, as seen in more trials and time required to complete training (see below, and Appendix B). Analyses including these 8 gave the same overall pattern of results.

Materials & Procedure

Experiment 2 employed the same assessment as Experiment 1, with the same associated dependent measures. I replaced the familiarization and familiarity test in Experiment1 with a perceptual learning (PL) condition and a new familiarity test for the PL condition. I also made a small change to the survey.

PL Training. In the PL condition, participants searched for target pairs in search grids. Each participant had a shape set of generated for them as in Experiment 1, and the pairs in their shape set each became a target to learn. We used target detection because tasks involving recurrent search for particular stimuli typically lead to PL effects of faster and more accurate extraction of these stimuli (e.g., Schneider & Shiffrin, 1977; Karni & Sagi, 1993). Each participant had a shape set of generated for them as in Experiment 1, and the pairs in their shape set each became a target to learn. On each trial, participants fixated, then saw a target pair for 1s, saw a search grid generated for that trial, indicated by key press whether the target was present (‘z’) or absent (‘/’) in the grid, received accuracy feedback (correct/incorrect), and finally advanced[13] to the next trial. Targets were present on half of the trials.

[13] 32 of 70 participants advanced from one trial to the next by clicking the mouse; 38, by pressing the spacebar. Participants who clicked did not differ from those who used the spacebar in number of trials (*p* = .99). Click advancement did, however, require marginally more time (*M* = 26.02 minutes, *SE* = 0.64) than spacebar advancement (*M* = 23.49, *SE* = 0.48), *t*(68) = 2.20, *p* = .002. Advancement method did not influence the psychophysical assessment accuracy (*p* = .31) or familiarity accuracy (*p* = .16).

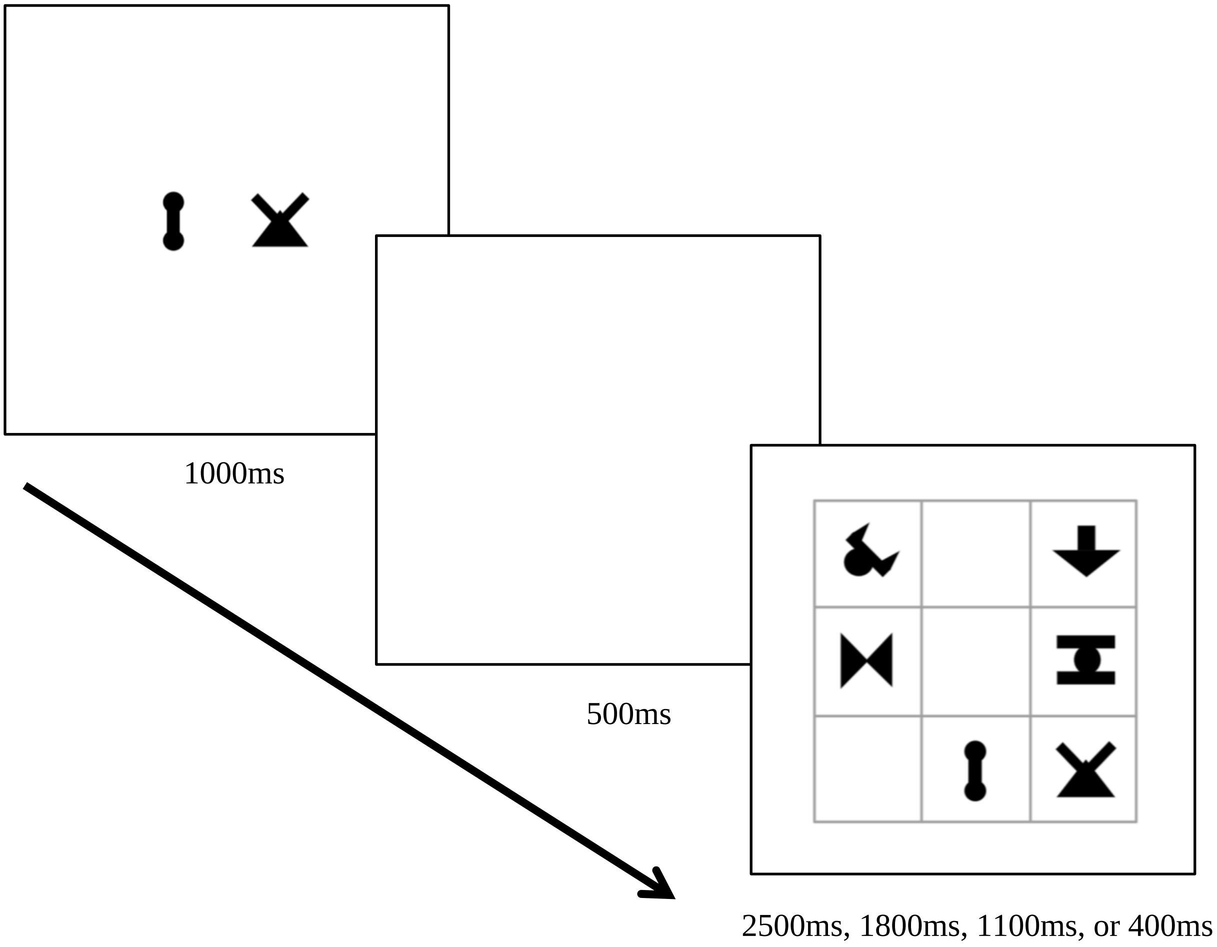

Figure 12 Schematic of a PL training trial showing the pair to learn, the pause, and the search grid.

Search grids consisted of six shapes from Fiser and Aslin (2001)'s shapes (see Figure 12). Half of grids contained a target pair. Non-target shapes were not arranged in target pairs because prior research (Schneider & Shiffrin, 1997) indicated that targets were only learnable when distinct from the distractor set. No shape could appear more than twice in a grid. Only one shape from a trial's target pair could appear in a target-absent search grid. On a target-present trial, one shape in the target pair could appear (once) in the distractors in the search grid. Half of target-present trials and half of target-absent trials contained a statistical learning pair. The SL pair consisted of two shapes selected from two different targets, and randomly assigned to either a horizontal or vertical orientation. Randomly selected shapes (not in pairs) filled in search grids so that each grid consisted of six shapes. Target pairs and the SL pair could not appear accidentally, only when specified for the trial.

Unlike in familiarization which lasted for a fixed session length, the PL condition ended when participants reached objective mastery criteria. Participants needed to answer 8 of the last 10 trials of each target correct, for each of four different search grid durations per target. Accuracy counts were reset for each target for each search grid duration. Search grid durations were 3000ms, 2100ms, 1200ms, and 300ms. Reaching the accuracy criterion for a given target (one of the 6) and a given search grid duration (one of the 4) was considered a level of mastery. Every 45 trials, participants were given a brief break and saw their current mastery level. To make training as brief as possible, each target was removed from training once the final (4th) mastery level for that target was achieved. When participants reached mastery level 24[14] (6 targets x 4 search grid durations = 24), they saw the break and mastery level feedback screen for a final time, and then a completion screen. Then they continued to the one-trial familiarity test. (Participants who had not reached mastery after 40 minutes would also have seen the completion screen, and advanced to the familiarity test, but all participants reached mastery.)

One-Trial Familiarity Test. The familiarity test for the PL condition was the same as the familiarity test in Experiment 1, except in the following ways: there was only a single trial, and the target was the SL pair from the PL training.

[14] Mastery level was, on average, 24.11 levels (*SE* = 0.13), and 60 of 70 participants or 86% of participants completed exactly 24 mastery levels. A coding error caused 10 participants to have more or fewer mastery levels.

Survey Because the most common response to the sleep hours question on the survey was "7 or less" hours, we changed the response options to separate out different amounts of sleep that fell into "7 or less" in Experiment 1. In Experiment 2 the sleep hour question response options were "0 (no sleep)", "1-3", "4", "5", "6", "7", "8", "9", and "10 or more".

Dependent Measures

PL Training. We collected trial-by-trial accuracy and reaction time (RT). These were primarily used for the purpose of structuring training. We also used this data to calculate the mean difference in accuracy for trials with and without the SL pair, and the mean difference in RT (correct trials only) for trials with and without the SL pair. We also collected the total number of PL trials per participant and the total length of PL training time.

One-Trial Familiarity Test. We collected accuracy, and responses were coded as either correct (1) or incorrect (0).

Psychophysical Assessment. (See Experiment 1 Method, in Chapter 2.)

Additional Predictors

Survey. I recoded the endpoints of the sleep measure from descriptions into numbers, e.g. from "1-3" to 3, to allow for quantitative analyses. I also examined alertness in Experiment 2. I was (again) interested in whether any particular strategy was associated with better performance, so we read participant responses to the strategy question, looking for linguistic strategy responses. Each participant was coded dichotomously as either employing the linguistic strategy (*n* = 39) or not (*n* = 31).

Finally, I also investigated noticing of the pairs in Experiment 2. Participants were much more likely to notice the pairs in Experiment 2 than Experiment 1, given that each pair was presented multiple times in isolation as the target of PL training trials. 44 participants reported noticing the pairs. One participant likely noticed the pairs. Twenty participants' responses did not indicate that they noticed the pairs or any consequences of the pair structure. (Five participants declined to respond to the noticing question, and were excluded from noticing analyses.)

Other Factors Because the other factors examined in Experiment 1 did not influence performance on either familiarity or the psychophysical assessment, we did not look at them in Experiment 2.

CHAPTER 5: EXPERIMENT 2 RESULTS & DISCUSSION

Results

Perceptual Learning Condition

Participants needed 24.65 minutes (*SE* = 0.42) on average to complete an average of 227.03 trials (*SE* = 2.84).They showed high accuracy (*M* = 0.92, *SE* = 0.01) on training trials, and responded in an average of 889.18ms (*SE* = 22.45) on trials answered correctly. Training time and number of trials were positively correlated such that as participants required more training trials, they also required more training time, *r*(70) = .50, *p* < .001. Mean difference in accuracy for trials with and without the SL pair was 0.00 (*SE* = 0.00). Mean difference in reaction time for (correct) trials with and without the SL pair was 7.89 seconds (*SE* = 10.73).

Effects of SL Pair Percentage. Eight participants (excluded from these Results, except for here) had a higher percentage of SL trials (75% for 7 and 64% for 1) than the other participants. Independent-samples *t*-tests showed that the 8 participants with a higher percentage of SL pair trials did not differ from the other participants in recognition accuracy (*p* = .28), average assessment accuracy (*p* = .23), time to complete training (*p* = .79), accuracy on training trials (*p* = .65), response time on trials answered correctly (*p* = .48), mean difference in accuracy for trials with and without the SL pair (*p* = .48), or mean difference in reaction time for correct trials with and without the SL pair (*p* = .43). However, these 8 participants (*M* = 413.25, *SE* = 22.70) required significantly more trials to complete assessments than those with 50% SL pair trials (*M* = 227.03, *SE* = 2.84), *t*(76) = 8.14, *p* < .001.

One-Trial Familiarity Test

Participants' accuracy did not differ from chance (*M* = 0.45, *SE* = 0.06) on recognition of the statistical learning pair in the familiarity test (*p*=.40, see Figure 13). 38 (of 69)[15] or 55% of participants did not indicate that the statistical learning pair seemed more familiar to them than the foil pair in the single 2AFC trial.

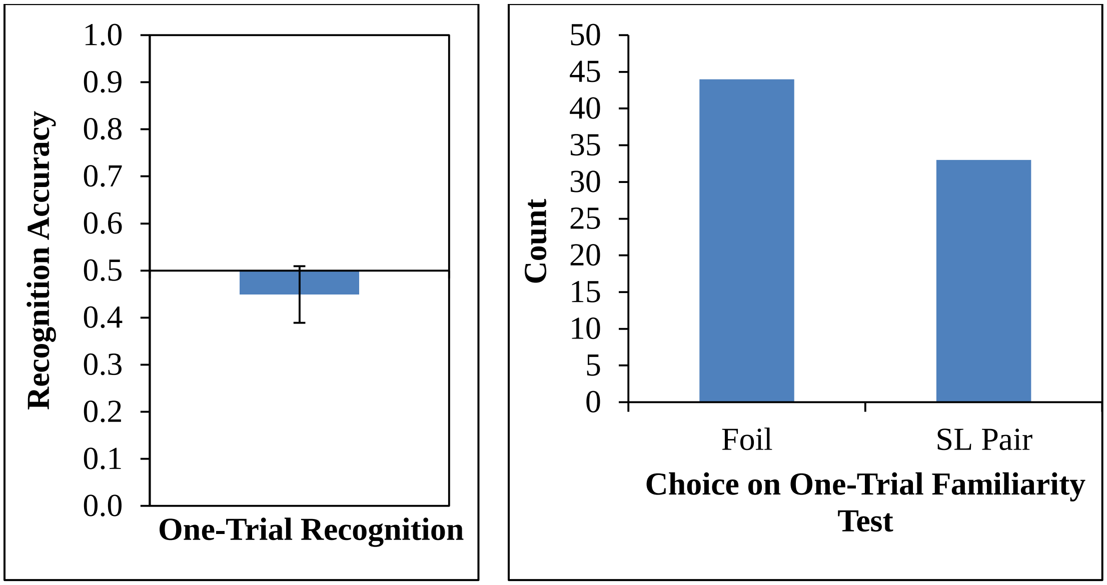

Figure 13. Average accuracy on the one-trial recognition test (left) and histogram of accuracy (right). Error bar (left) shows standard error of the mean.

An ANOVA of Condition on recognition accuracy showed that conditions did not differ in recognition (*p* = .82). Participants in all condition recognized the statistical learning pair equally poorly. This is unsurprising because all participants received the same recognition test

[15] One participant experienced an error and did not complete the recognition trial.

(relative to their familiarization shape set). It indicated that there were not significant differences across participants in different conditions by chance, an important manipulation check.

PL Training and Psychophysical Assessment Accuracy. However, participants who did recognize the SL pair might have performed differently from those who did not on the assessment. Perhaps those who recognized the SL pair learned targets less well from looking at distractors, or, conversely, perhaps they noticed the SL pair and were able to eliminate it more quickly from their searches and more efficiently find targets than other participants. To investigate this possibility, I conducted a Pearson correlation of average psychophysical assessment accuracy and familiarity test accuracy. I found no significant correlation of familiarity and assessment accuracy (*p* = .69)[16].

PL Training and Familiarity. I conducted analyses to investigate whether the PL training influenced performance on the familiarity test. Familiarity test accuracy did not correlate with the number of PL training trials completed (*p* = .15). PL training time did not correlate with familiarity test accuracy (*p* = .37).

Psychophysical Assessment: Main Analyses

Accuracy. Figure 14 shows the effects of Condition, Exposure Duration, and Target Presence on accuracy. It appeared that Familiar showed the highest accuracy across exposure durations and across Target Presence. Shuffled appeared to have numerically higher accuracy than New Shapes in all but one case. An ANOVA of Condition by Exposure Duration by Target Presence on accuracy tested these apparent patterns and revealed a main effect of Condition, *F*(1,

[16] When I broke the data on Condition, I found no correlation of familiarity and assessment accuracy for Familiar, r(23) = .46, p = .16; Shuffled, r(22) = .35, p = .10; but there was a significant negative correlation for New Shapes, r(21) = -.54, p = .01, such that choosing the SL pair predicted lower assessment accuracy.

67) = 16.46, *p* < .001, *partial-eta-squared* = 0.33, a main effect of Exposure Duration, *F*(3, 201) = 5.41, *p* = .001, *partial-eta-squared* = 0.08, and no other effects (all *p*'s > .24).

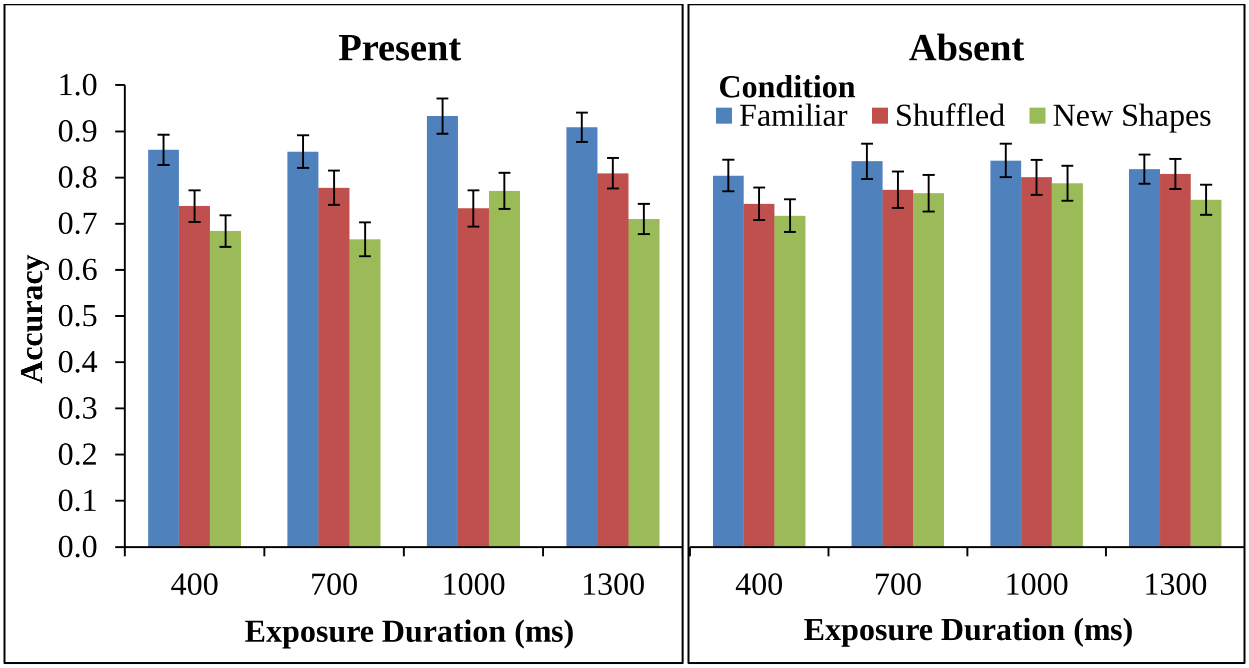

Figure 14. Condition by Exposure Duration by Target Presence on accuracy. Error bars indicate standard error of the mean.

Custom hypothesis tests were used to investigate the significant effects. These tests for Condition revealed that the Familiar condition (*n* = 24[17], *M* = 0.88, *SE* = 0.02) was significantly more accurate than Shuffled (*n* = 23, *M* = 0.78, *SE* = 0.02), *F*(1,67) = 13.13, *p* = .001, *partial-eta-squared* = 0.16. Similarly, Familiar was significantly more accurate than New Shapes (*n* = 23, *M* = 0.72, *SE* = 0.02), *F*(1,67) = 31.95, *p* < .001, *partial-eta-squared* = 0.32. Shuffled was

[17] The 13 original and 11 added participants did not differ on average recognition accuracy (*p* = .24), assessment accuracy (*p* = .41), number of training trials (*p* = .53) or training time (*p* = .65).

also more accurate than New Shapes, *F*(1,67) = 4.03, *p* = .05, *partial-eta-squared* = 0.06. Learning was strongest in the trained condition, but did transfer to Shuffled as well.

For Exposure Duration, I first compared 1000ms (*M* = 0.81, *SE* = 0.02) and 1300ms (*M* = 0.81, *SE* = 0.02) and found no difference in accuracy (*p* = .56). Then I compared 400ms (*M* = 0.76, *SE* = 0.02) to 700ms (*M* = 0.79, *SE* = 0.02) and found that did not differ (*p* = .10). Then I compared the shorter exposure durations to the longer exposure durations and found that 400ma and 700ms combined were significantly less accurate than 1000ms and 1300ms combined, *F*(1,67) = 16.20, *p* < .001, *partial-eta-squared* = 0.20. As in Experiment 1, the shorter exposure durations showed less accuracy than the longer ones, so the results are (again), interpretable.

Hit Rate. Figure 15 showed the effects of Condition and Exposure Duration on hit rate. It appeared that Familiar had the highest hit rate across exposure durations, followed by Shuffled. To test these apparent effects, I conducted an ANOVA of Condition by Exposure Duration. This analysis revealed a main effect of Condition *F*(1,67) = 16.94, *p* < .001, *partial-eta-squared* = 0.34, and a main effect of Exposure Duration *F*(3,201) = 4.12, *p* = .007, *partial-eta-squared* = 0.06, and no interaction (*p* = .13).

Custom hypothesis tests were used for all pairwise comparisons between Conditions. Familiar (*M* = 0.90, *SE* = 0.02) condition showed a significantly higher hit rate than Shuffled (*M* = 0.76, *SE* = 0.03), *F*(1,67) = 14.94, *p* < .001, *partial-eta-squared* = 0.18. Familiar also had significantly more hits than New Shapes (*M* = 0.70, *SE* = 0.03), *F*(1,67) = 32.34, *p* < .001, *partial-eta-squared* = 0.33. Shuffled had a marginally higher hit rate than New Shapes, *F*(1,67) = 3.25, *p* = .08, *partial-eta-squared* = 0.08. Learning was strongest in the trained condition, but did transfer to Shuffled as well.

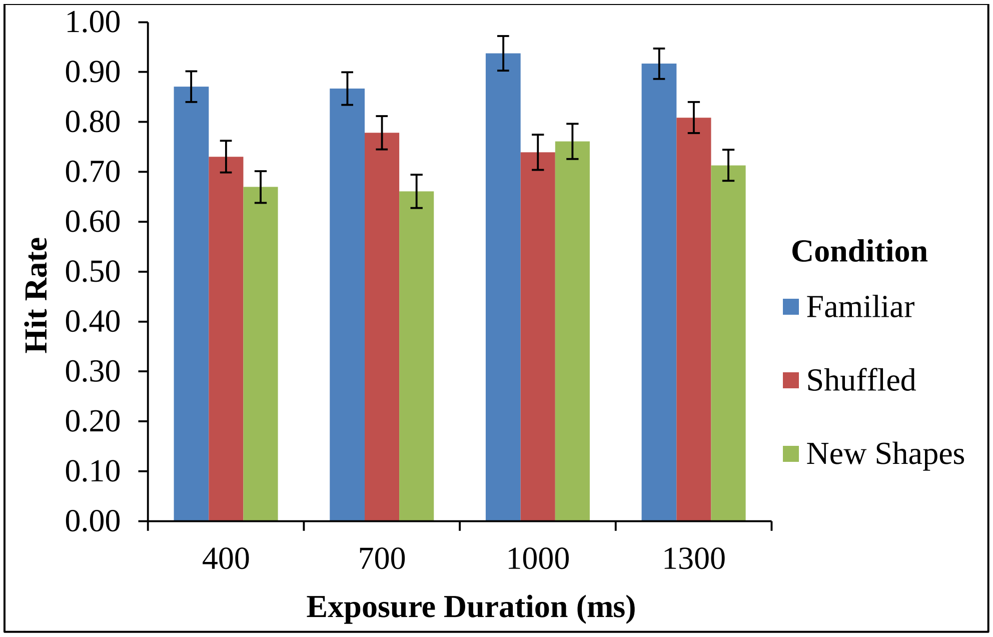

Figure 15. Condition by Exposure Duration on hit rate. Error bars indicate standard error of the mean.

I conducted custom hypothesis tests to investigate the main effect of Exposure Duration. I first compared the shorter exposure durations and found that 400ms (*M* = 0.76, *SE* = 0.02) and 700ms (*M* = 0.77, *SE* = 0.02), did not differ in hit rate (*p* = .56). Similarly, when I compared 1000ms (*M* = 0.81, *SE* = 0.02) and 1300ms (*M* = 0.81, *SE* = 0.02), I found that they also did not differ in hit rate (*p* = .99). Finally, I compared the short exposure durations to the long exposure durations and found that the longer exposure durations combined had significantly higher hit rates than the shorter exposure durations combined, *F*(1,67) = 11.77, *p* = .001, *partial-eta-squared* = 0.15.

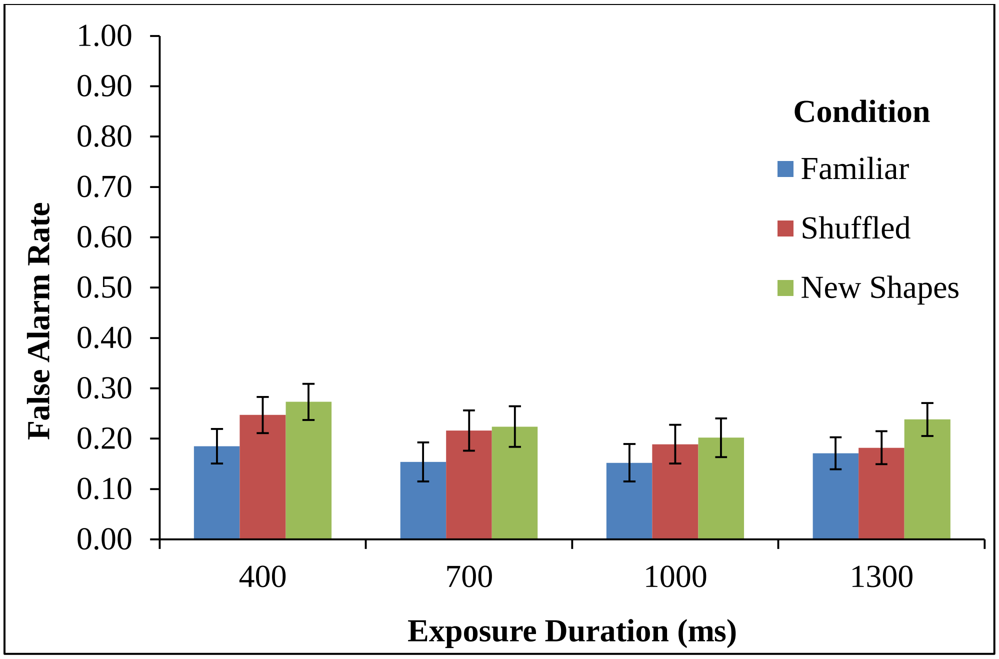

Figure 16. Condition by Exposure Duration on false alarm rate. Error bars indicate standard error of the mean.

False Alarm Rate. Figure 16 showed effects of Condition and Exposure Duration on false alarm rate. It appeared that false alarm rate was neatly stair-stepped across exposure durations, with Familiar showing the lowest rates, then Shuffled, and New Shapes showing the highest. An ANOVA of Condition by Exposure Duration on false alarm rate tested these apparent effects and revealed a main effect of Condition, *F*(2,67) = 4.98, *p* = .01, *partial-eta-squared* = 0.13, and a main effect of Exposure Duration, *F*(3,201) = 2.78, *p* = .04, *partial-eta-squared* = 0.04, and no interaction (*p* = .78).

I followed up on the main effect of Condition with custom hypothesis tests in ANOVA of all pairwise comparisons. I found that Familiar (*M* = 0.13, *SE* = 0.03) had a significantly lower false alarm rate than New Shapes (*M* = 0.27, *SE* = 0.03), *F*(1,67) = 9.84, *p* = .003, *partial-eta-*

squared = 0.13. Shuffled (*M* = 0.21, *SE* = 0.03) had a marginally higher false alarm rate than Familiar, *F*(1,67) = 3.39, *p* = .07, *partial-eta-squared* = 0.05. Shuffled did not differ from New Shapes (*p* = .20). The lowest false alarm rate was in the trained condition.

For the marginal effect of Exposure Duration, I conducted custom hypothesis tests of pairwise comparisons. These analyses revealed that 400ms (*M* = 0.24, *SE* = 0.02) had a marginally higher false alarm rate than 1000ms (*M* = 0.18, *SE* = 0.02), *F*(1,67) = 8.51, *p* = .005, *partial-eta-squared* = 0.11. No other comparisons were significant (all *p*'s > .03).

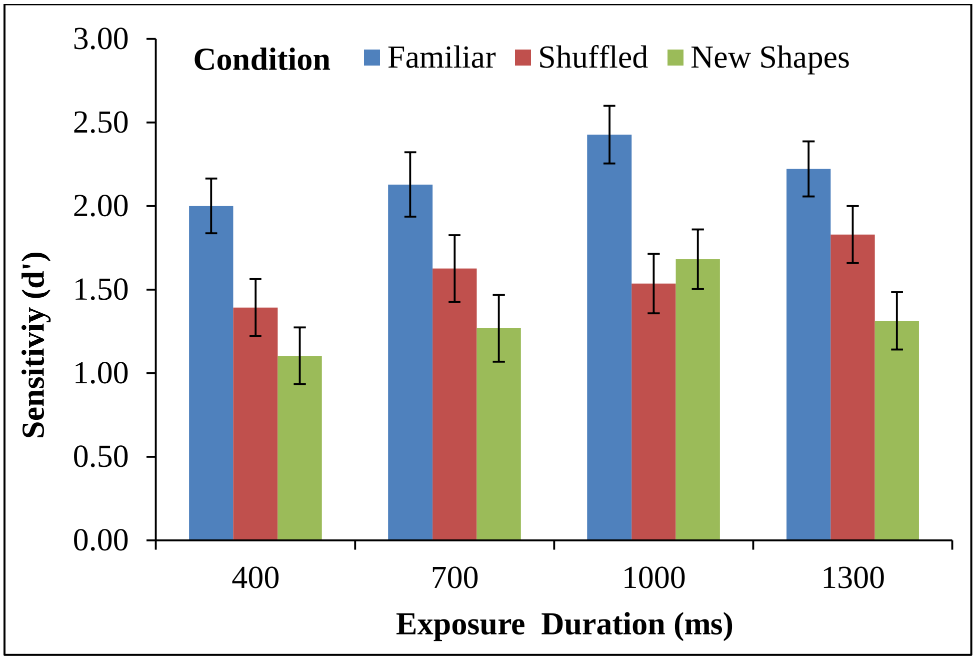

Figure 17. Condition by Exposure Duration on sensitivity. Error bars indicate standard error of the mean.

Sensitivity. Figure 17 showed effects of Condition and Exposure duration on sensitivity. Familiar appeared to have the highest sensitivity across exposure durations, and Shuffled appeared to show higher sensitivity than New Shapes at three of four exposure durations. An

ANOVA of Condition by Exposure Duration on sensitivity tested these patterns of results. I found a main effect of Condition *F*(2,67) = 16.40, *p* < .001, *partial-eta-squared* = 0.33, and a main effect of Exposure Duration *F*(3,201) = 6.32, *p* < .001, *partial-eta-squared* = 0.09, and no interaction (*p* = .15).

I followed up on significant effects via custom hypothesis tests in ANOVA. For the main effect of Condition, I conducted all pairwise comparisons. I found that Familiar (*M* = 2.33, *SE* = 0.14) had significantly higher sensitivity than Shuffled (*M* = 1.57, *SE* = 0.14), *F*(1,67) = 14.20, *p* < .001, *partial-eta-squared* = 0.18. Familiar also had significantly higher sensitivity than New Shapes (*M* = 1.20, *SE* = 0.14), *F*(1,67) = 31.41, *p* < .001, *partial-eta-squared* = 0.32. Shuffled showed marginally higher sensitivity than New Shapes, *F*(1,67) = 3.30, *p* = .07, *partial-eta-squared* = 0.05. Learning was strongest in the trained condition, but did transfer to Shuffled as well.

I conducted all pairwise comparisons using custom hypothesis tests in ANOVA to follow up on the main effect of Exposure Duration. 400ms (*M* = 1.47, *SE* = 0.10) was less sensitive than 1000ms (*M* =1.86, *SE* = 0.10), *F*(1,67) = 21.08, *p* < .001, *partial-eta-squared* = 0.24. 400ms was also less sensitive than 1300ms (*M* =1.79, *SE* = 0.10), *F*(1,67) = 16.65, *p* < .001, *partial-eta-squared* = 0.20. No other comparisons were significant (all *p*'s > .05).

Bias. Figure 18 showed the effects of Condition and Exposure Duration on bias (the tendency to respond present or absent more often regardless of stimulation) as measured by criterion. It appeared that participants showed a more positive criterion for Familiar across exposure durations. To test this apparent effect, I conducted an ANOVA of Condition by Exposure Duration on criterion (see Figure 18), and found no significant effects (all *p*'s > .23).

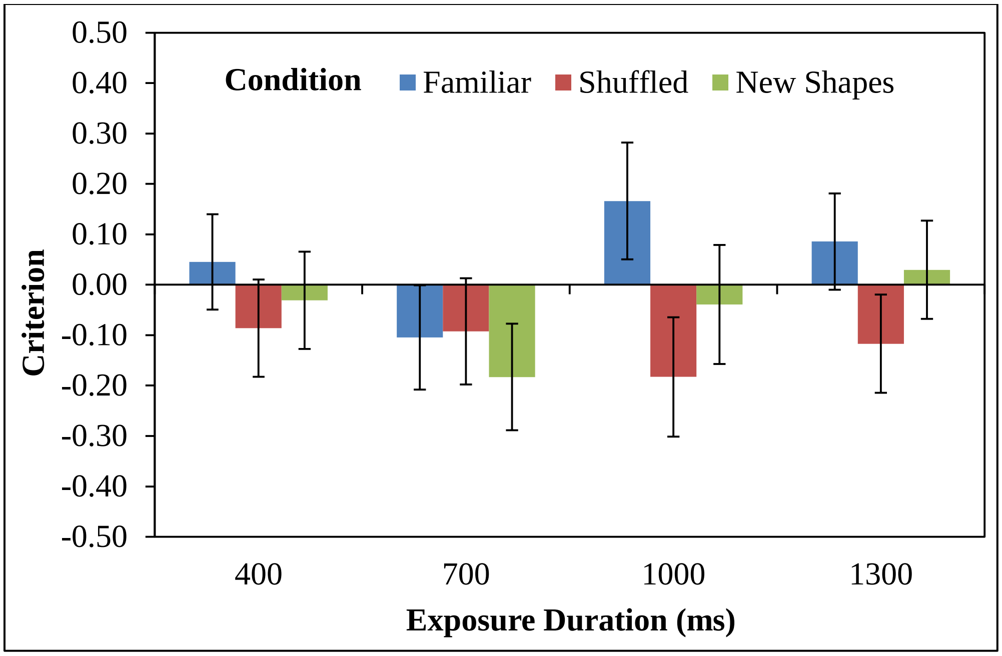

Figure 18. Condition by Exposure Duration on criterion. Error bars indicate standard error of the mean.

Effects of Training & Survey Variables

Training. As my first step in exploring if PL training time or number of trials influenced psychophysical assessment performance, I conducted Pearson's correlations. PL training time did not correlate with assessment accuracy (p = .11). The number of PL trials completed negatively correlated with assessment accuracy, $r(69) = -.53$, $p < .001$. Participants who completed more trials were generally less accurate.

Survey Data: Noticing the Pairs. A Pearson correlation of noticing with one-trial recognition accuracy showed no relationship (p = .97). Noticing did marginally correlate with assessment accuracy, $r(64) = .32$, $p = .009$. Noticing was associated with higher accuracy. I also tested the relationship between noticing and number of PL trials, because number of PL trials also negatively correlated with assessment accuracy. Noticing and number of PL trials had a negative but non-significant relationship, *r*(64) = -.29, *p* = .02.

Survey Data: Strategy. Even though linguistic coding was the most popular strategy (see Dependent Measures) and it could have made encoding the pairs and search grids more efficient, linguistic coding did not impact performance on either the recognition (*p* = .48) or the assessment (*p* = .56).

Survey Data: Alertness and Sleep. Participants reported a medium level of alertness (*M* = 3.27, *SE* = 0.09). Alertness did not correlate with recognition (*p* = .41) or assessment (*p* = .72) accuracy. The average number of hours slept was 6.46 (*SE* = 0.17). No correlation of sleep and recognition accuracy (*p* = .29) or sleep and psychophysical assessment accuracy (*p* = .03) accuracy was found.

Psychophysical Assessment: Effects of Number of Trials and Noticing the Pairs

Accuracy. I followed up on the significant correlations of Noticing the pairs and assessment accuracy and Number of PL training trials and accuracy with ANCOVA analyses, statistically controlling for these effects. Figure 19 showed the effects of Condition, Exposure Duration, and Target Presence on accuracy, covarying out the effects of Noticing and Number of PL trials. Marginal means were estimated at means of the covariates: Noticing (*M* = 0.69), Number of Trials (*M* = 227.88). In Figure 19, it appeared that Familiar showed higher accuracy than the other conditions across exposure durations, and especially for Present. I tested this pattern of results by conducting an ANCOVA of Condition by Exposure Duration by Target Presence on accuracy, covarying out the Number of PL trials completed and the effect of Noticing the pairs. There was a main effect of Condition, *F*(2,60) = 9.19, *p* < .001, *partial-eta-squared* = 0.24; a significant effect of the covariate of Number of trials, *F*(1,60) = 10.15, *p* =

.002, *partial-eta-squared* = 0.15; a marginal effect of the covariate of Noticing the pairs, *F*(1,60) = 2.82, *p* = .099, *partial-eta-squared* = 0.05; and an interaction of Number of trials and TargetPresence, *F*(1,60) = 12.00, *p* = .001, *partial-eta-squared* = 0.17. No other effects were significant (all *p*'s > .10).

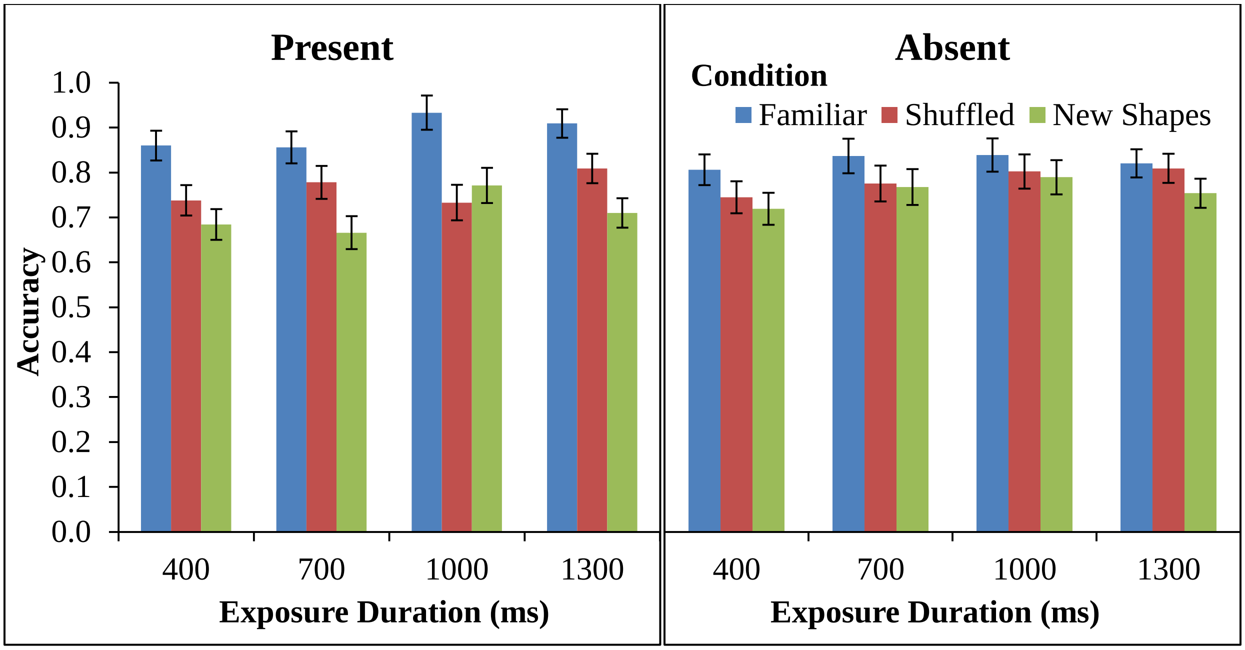

Figure 19. Condition by Exposure Duration by Target Presence on accuracy, covarying out the effect of Number of trials and Noticing the pairs. Bar heights indicate adjusted marginal means and error bars indicate standard error of the mean.

I followed up my significant main effect of Condition with custom hypothesis tests in ANCOVA on the adjusted marginal means. I first compared Shuffled (*M* = 0.78, *SE* = 0.02) and New Shapes (*M* = 0.74, *SE* = 0.02) and found that they did not differ in accuracy (*p* = .17). Then I compared Familiar (*M* = 0.86, *SE* = 0.02) to the other conditions, and found that Familiar showed significantly higher accuracy than the other two conditions combined, *F*(1,60) = 16.96, *p* < .001, *partial-eta-squared* = 0.22.

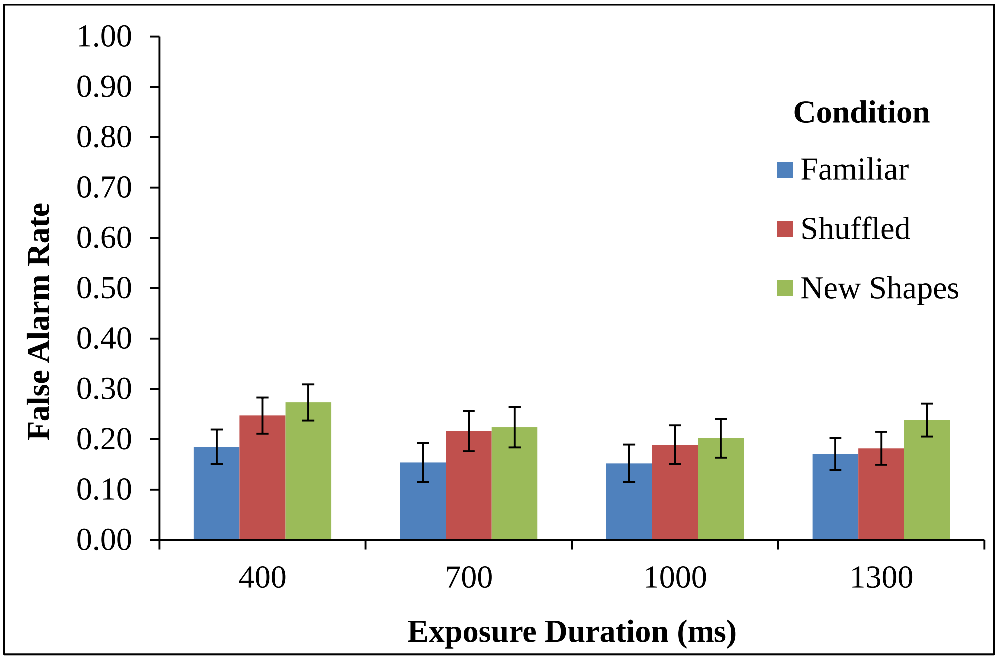

Figure 20. Condition by Exposure Duration on false alarm rate, covarying out the effect of Number of trials and Noticing. Bar heights indicate adjusted marginal means and error bars indicate standard error of the mean.

False Alarm Rate. Given that Number of PL training trials and Noticing the pairs covaried with accuracy, I followed up with ANCOVA analyses involving the other hypothesis-relevant dependent variables, false alarm rate and sensitivity. Figure 20 showed the effects of Condition, and Exposure Duration on false alarm rate, covarying out Number of PL trials and Noticing. Familiar appeared to have the lowest false alarm rate across exposure durations. I used an ANCOVA to analyze the effects of Condition and Exposure Duration on false alarm rate, covarying out the effect of Number of trials and the effect of Noticing, to test this apparent effect. There were only effects of the covariate of Number of trials, *F*(1,60) = 8.48, *p* = .005, *partial-eta-squared* = 0.12; and the covariate of Noticing the pairs, *F*(1,60) = 11.76, *p* = .001, *partial-eta-squared* = 0.16; and no other effects (all *p*'s > .26).

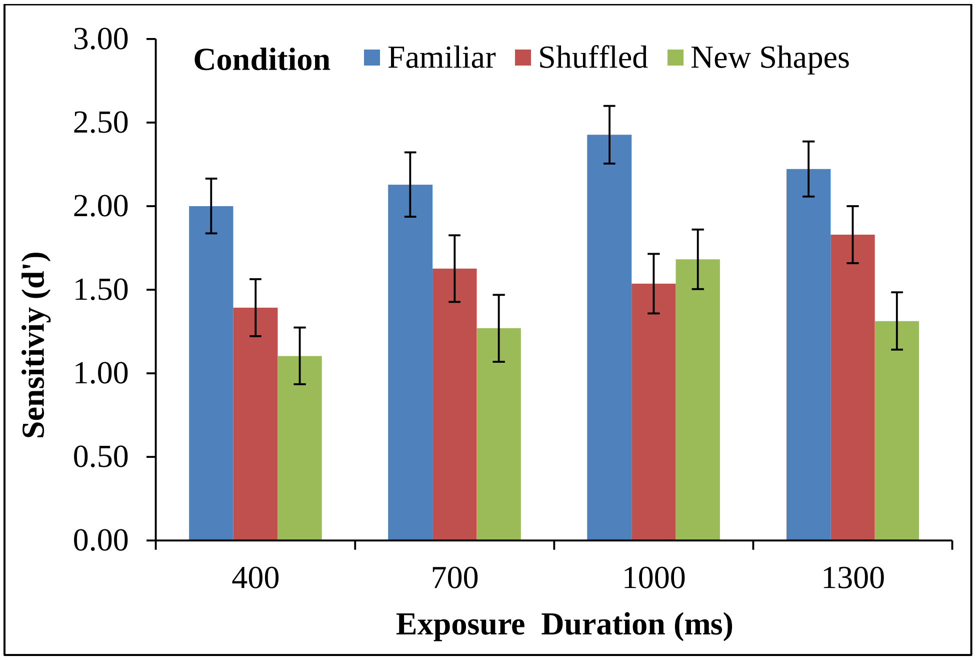

Figure 21. Condition by Exposure Duration on sensitivity (d'), covarying out the effect of Number of trials and Noticing. Bar heights indicate adjusted marginal means and error bars indicate standard error of the mean.

Sensitivity. Figure 21 showed the effects of Condition and Exposure Duration on sensitivity, statistically controlling for the effect of the Number of PL trials and the effect of Noticing the pair structure. Familiar showed the highest sensitivity across exposure durations, and Shuffled showed higher sensitivity than New Shapes at three of four exposure durations. To test these apparent effects, I conducted an ANCOVA to analyze the effect of Condition and Exposure Duration on sensitivity (d'), covarying out the effect of the Number of PL trials and the effect of Noticing the pairs. There was a significant main effect of Condition, *F*(2,60) = 9.28, *p* < .001, *partial-eta-squared* = 0.24; a significant effect of the covariate Number of trials, *F*(1,60) = 14.50, *p* = .004, *partial-eta-squared* = 0.13; a marginal effect of the covariate Noticing the pairs, *F*(1,60) = 4.93, *p* = .09, *partial-eta-squared* = 0.05; and no other effects (all *p*'s > .22).

I followed up my significant main effect of Condition with custom hypothesis tests in ANCOVA on the adjusted marginal means. I first compared Shuffled (*M* = 1.60, *SE* = 0.14) and New Shapes (*M* = 1.34, *SE* = 0.14) and found that they did not differ in sensitivity (*p* = .21). Then I compared Familiar (*M* = 2.20, *SE* = 0.14) to the Shuffled and New Shapes together, and found that Familiar showed significantly higher sensitivity than the other conditions combined, *F*(1,60) = 17.47, *p* < .001, *partial-eta-squared* = 0.23.

Discussion

Experiment 2 sought converging evidence for the hypothesis that perceptual learning (PL) and statistical learning (SL) are part of a unified learning process. We developed a PL training that we based on Fiser and Aslin (2001)'s SL paradigm used in Experiment 1 and on Schneider and Shiffrin (1977)'s paradigm, and tested for similar effects as in Experiment 1 via our novel psychophysical assessment. If SL and PL are parts of a unified learning process, then conditions leading to SL and conditions leading to PL should both yield PL effects: transfer and improved sensitivity. Our data demonstrate PL effects, and similar effects as those observed in Experiment 1.

Comparing our Data to the Hypotheses

The hypothesis of two kinds of learning would predict very different patterns of results for Experiments 1 and 2, because this hypothesis predicts SL effects of familiarity and, possibly, bias following an SL paradigm (Experiment 1) and PL effects of transfer and improved sensitivity following a PL paradigm (Experiment 2). In contrast, the hypothesis of a single, unified learning process predicts similar patterns of results for Experiments 1 and 2 – specifically a correlation between recognition and assessment accuracy and PL effects – from SL and PL paradigms. Notably, both hypotheses predict PL effects following a PL paradigm, but only differ as to their predictions for SL paradigms, and differ as to whether familiarity and assessment accuracy should be correlated. As expected by both hypotheses, we successfully induced PL in Experiment 2, as seen in improved sensitivity and transfer of learning. The largest effects in Experiment 2 were seen in the trained condition, Familiar. Familiar had significantly higher accuracy, hit rates, and sensitivity, and lower false alarm rates, than the other conditions. We also observed transfer of learning from Familiar to Shuffled: performance was lower than Familiar but higher than New Shapes for accuracy, hit rate, and sensitivity.

Qualitatively comparing the pattern of results observed in Experiments 1 and 2 on the same psychophysical assessment could give evidence as to whether results following the SL familiarization in Experiment 1 were due to PL, which was explicitly trained in Experiment 2. Because both Experiment 1 and 2 showed improved sensitivity and transfer of learning, the Experiment 2 data qualitatively show the same hallmarks of PL observed in Experiment 1, consistent with the hypothesis of a unified learning process but inconsistent with the hypothesis of two kinds of learning. However, there were no reliable overall correlations between the SL tests of familiarity and assessment accuracy for either experiment. Experiment 2 offers converging evidence that the relationship between SL and PL is nuanced.

Comparing Experiment 1 and Experiment 2 Stimuli

Experiment 1 and Experiment 2, by their different designs, gave participants different amounts of exposure to pairs to be learned, in similar amounts of training time (21 minutes and

24.65 minutes, respectively). Experiment 1 passively exposed participants to 432 grids (144 unique grids x 3 blocks of all unique grids) consisting of 3 pairs per grid, so participants were exposed to each pair an average of 216 times (432 grids x 3 pairs per grid ÷ 6 pairs). Participants in Experiment 2 saw an average of 227.03 trials, each with one target pair and one search grid (containing the target pair on half of trials). Targets were seen, on average, 56.76 times (227.03 trials ÷ 6 pairs * 1.5 exposures per trial), or 26% as often as in Experiment 1. The SL pair in Experiment 2 was seen 113.51 times on average (227.03 trials x .5 exposures per trial), or 53% as often as in Experiment 1. SL predicts better learning of relationships with more exposure to them, but PL predicts that learning depends more on the quality of training than on the number of exposures. Experiments 1 and 2 yielded similar patterns of results - more similar than predicted by the hypothesis of different kinds of learning - with many fewer exposures but more structured training in Experiment 2 than Experiment 1.

No Learning of SL Pair

Participants overall showed no learning on the one-trial recognition test of the statistical learning pair. The SL pair did not influence either accuracy or reaction time in training. Participants (with 50% of trials with the SL pair) also showed no evidence of learning of the SL pair in PL training – no higher accuracy or faster reaction times for trials with the SL pair.

However, analyses comparing participants with the intended SL pair percentage to those with a higher percentage suggests an influence of the SL pair on learning: participants who had the SL pair on a higher percentage of trials needed more trials to complete learning, but not more time. The SL pair may have increased the difficulty of learning as indicated by increasing the number of trials required to complete training. More trials were required for participants who responded incorrectly more often because of the nature of training to mastery criteria. The SL pair may have increased difficulty, but it seems also to have facilitated quicker responding because participants with a higher percentage of SL trials did not take longer to complete learning. Accuracy and response time in training did not differ across SL percentage, but the small sample of participants with a high SL percentage and the near-ceiling accuracy in training for all participants may explain why no differences were found.

Given that statistical learning is about tracking correlations, SL would predict learning of the SL pair. SL would also predict that this would be a larger effect than learning of targets because the SL pair appeared twice as frequently as each target. Lack of learning of non-attended parts of grids is more consistent with perceptual learning because learning to suppress irrelevant information is a PL discovery effect (Kellman, 2002; Kellman & Garrigan, 2009). This explanation is also consistent with the psychophysical assessment results.

The familiarity result in Experiment 2 differed from Experiment1, in which participants showed significant recognition on their recognition of the 6 pairs in their familiarization. These recognition tests differed not only in results, but also in the number of trials and trial content. Experiment 1 tested the 6 pairs to which participants were exposed in familiarization, whereas Experiment 2 tested the SL pair that appeared twice as often as target pairs to be learned but only appeared in search grids as part of the set of distractors. Another possible explanation for the different findings is the different number of exposures to tested pairs (216 on average in Experiment 1 v. 113.51 on average in Experiment 2), but this would not explain the psychophysical data.

Number of PL Trials and Noticing the Pairs

Number of PL training trials correlated with assessment accuracy, and Noticing the pairs correlated with assessment accuracy. Both were significant covariates in ANCOVA analyses statistically controlling for the effects of Number of trials and of Noticing. These analyses showed that Familiar had the highest accuracy and most sensitivity, relative to the other conditions (Shuffled, New Shapes), which did not differ. Conditions did not differ in false alarm rate with the effects of Number of training trials and of Noticing controlled. When Number of training trials and Noticing the pairs were included in the statistical model, transfer from Familiar to Shuffled was no longer observed, but participants still showed improved sensitivity, so our interpretation of our results holds.

What was Learned?. The fact that Noticing positively correlated with assessment accuracy and was a significant covariate in ANCOVA analyses demonstrates that participants with explicit knowledge of the pair structure showed more learning, and awareness of the pairs helped explain variation in learning for assessment performance. The pairs themselves (or in one case, a consequence of the pair structure) were explicit for these participants. This suggests that these participants learned the pairs - not just the shapes - and that this learning helped explain the PL effects demonstrated in the assessment.

Summary

In sum, we found PL effects following PL training. The pattern of results was similar to the pattern of results found in Experiment 1, with the exception of stronger effects in the trained condition, and much more similar than predicted by the hypothesis of different kinds of learning. Participants showed transfer of learning from Familiar to Shuffled in accuracy, hit rate, and sensitivity. Noticing the pair structure improved assessment accuracy, suggesting that participants learned pairs, as did completing the PL training in fewer trials. When Noticing and Number of PL trials were statistically controlled for in ANCOVAs, participants still showed improved sensitivity, evidence of PL. But as in Experiment 1, SL familiarity did not correlate with assessment accuracy. Because we found evidence of PL and similar effects to Experiment 1, Experiment 2 constitutes converging evidence that the relationship between SL and PL is nuanced.

Next Steps

Experiments 1 and 2 have similar results, but it would also be beneficial to quantitatively compare their results (see Ch. 8 for multi-experiment analyses). It is possible that by increasing the length of the familiarization session (Experiment 3), an SL familiarization might yield psychophysical assessment results more similar to the results of Experiment 2 – show a particular benefit for learned (Familiar) pairs, as seen after PL training. It is also possible that by decreasing the familiarization session length (Experiment 3), that effects might weaken or the data might no longer show PL effects.

CHAPTER 6: EXPERIMENT 3 INTRODUCTION AND METHODS

Experiment 3: How does session length impact learning?

Experiments 1 and 2 gave evidence that the relationship between and perceptual learning (PL) and statistical learning is nuanced, so in Experiment 3 I replicated Experiment 1 and extended it to study under what conditions the PL effects following an SL familiarization hold. Specifically, I examined the time course of learning by varying the familiarization session length across participants. I predicted stronger learning with more familiarization and weaker learning, or possibly SL effects, or a mix of SL and PL effects with reduced familiarization.

In Experiment 1, we found evidence of PL effects following an SL familiarization: participants increased sensitivity for the shapes in the relationships they learned (Familiar) and transfer in accuracy, false alarm rate, and sensitivity for arrays that contained the same shapes in new relationships (Shuffled), and decreased their relative to Baseline performance. In Experiment 2, by employing a PL intervention, we found similar effects of transfer and increased sensitivity. In Experiment 3 I aimed to gain a better understanding of the scope of these effects by exploring conditions in which the effects might strengthen, weaken, or even not hold. By doing so, I sought a deeper understanding of these novel findings.

Experiment 3 directly replicated Experiment 1 and expanded on Experiment 1's findings by investigating the effects of varying amounts of learning by varying the familiarization session length. In Experiment 3, I systematically varied the amount of familiarization, using a shorter session length, the same session length as in Experiment 1 (the replication), and a longer session length. I expected greater amounts of familiarization to increase the strength of the effects – especially increased sensitivity, and possibly more transfer – so as to be more similar to the strength of effects observed following PL training in Experiment 2. For smaller amounts of familiarization, I expected weaker PL effects, or possibly, qualitatively different effects. With minimal or incomplete learning, perhaps I would find SL effects instead of PL effects or a mix of both SL and PL effects, if finding correlations between items is a step in the process of unitization. Such SL effects could possibly be increased false alarming in Shuffled, and little or no change in sensitivity and little or no transfer of learning.

Method

Participants

205[18,19] (121 female, 81 male) undergraduates in psychology and linguistics courses at the University of California, Los Angeles were recruited through the Psychology Department subject pool. There were 23 participants in Familiar with the 7-minute session length; 23, with 21 minutes; and 23, with 35 minutes. For Shuffled, there were 24 participants with 7 minutes; 22, with 21 minutes; and 23, with 35 minutes. For New Shapes, 22 had 7 minutes; 23 had 21 minutes; and 22 had 25 minutes.

[18] Power analyses in GPower 3.1 (Faul et al., 2007) based on pilot data indicated that a sample size of 20-25 per between-subjects condition would be sufficient.

[19] 12 participants were added (from the same source) after the other participants were run to balance the conditions. The 11 original and 12 added participants in who had a session length of 21 minutes and the New Shapes version of the psychophysical assessment did not differ on average recognition accuracy (p = .49), but they did differ in average assessment accuracy – the later participants (n = 12, M = 0.77, SE = 0.11) were more accurate than the earlier participants (n = 11, M = 0.68, SE = 0.09), t(21) = 2.08, p = .05. Bootstrapping population distributions of the means of the original and added groups (using 10,000 repetitions of the average assessment accuracies and the same sample sizes, i.e. 11 for original) and calculating 99% confidence intervals in R suggested that the original [0.67, 0.71] and late[0.75, 0.79] participants were from different populations. Results excluding the later participants (see Appendix D) showed generally the same pattern of results as including them. Because adding the later participants balanced sample sizes across conditions and boosted sample size to the level recommended by power analyses without generally changing the pattern of results, analyses include these participants unless otherwise specified.

Participants had normal or corrected-to-normal vision and were compensated with partial course credit. Eleven participants were excluded because they had completed the incorrect version of the psychophysical assessment due to experimenter error. Three participants were excluded due to loss of data: a bug caused the assessment to crash for one participant, and no data were collected for two because the study was cut short by a fire drill.

Materials & Procedure

Experiment 3 directly replicated Experiment 1[20] and extended it to different amounts of familiarization. The method was identical to Experiment 1, with the following exceptions: 1) I added familiarization Session Length as a between-subjects variable. The 21-minute level was identical to Experiment 1. Other levels included 7 minutes, the length of familiarization used in the original study (Fiser & Aslin, 2001) which was 14 minutes shorter than the session length in Experiment 1; and 35 minutes, a session length 14 minutes longer than the session length used in Experiment 1. 2) The slightly modified survey from Experiment 2 was used for Experiment 3. 3) I again omitted the additional factors examined in Experiment 1.

[20] A programming error removed the gridlines in the familiarization for the third and later blocks of familiarization (impacted third block in 21 minute session length and third through fifth blocks in the 35 minute session length).

CHAPTER 7: EXPERIMENT 3 RESULTS AND DISCUSSION

Results

Recognition

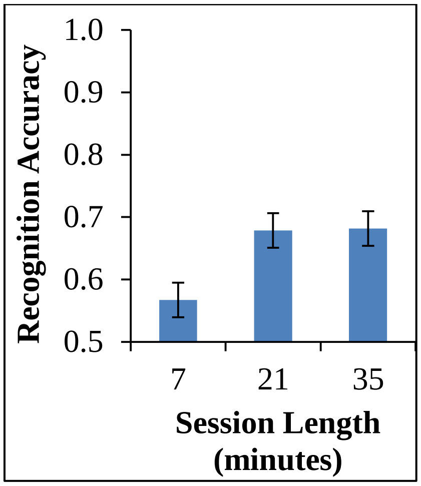

Figure 22. Session Length on recognition accuracy. Error bars indicate standard error of the mean.

Figure 22 showed that participants demonstrated accuracy significantly higher than chance (*M* = 0.64, *SE* = 0.02) on recognition of pairs in familiarization across session lengths in the recognition test, *t*(202) = 8.78, *p* < .001, Cohen’s *d* = 2.77. As apparent in the distribution of accuracy for participants (in Figure 23), many participants did not pass the recognition test, but more passed with longer session lengths. Participants were divided into two groups by their accuracy on recognition: “Recognizers” (*n* = 125) scored above 50% and “Nonrecognizers” (*n* = 80) scored at or below 50%. An independent-samples *t*-test of recognition group on recognition accuracy revealed that Recognizers (*M* = 0.79, *SE* = 0.01) had significantly higher accuracy than Nonrecognizers (*M* = 0.40, *SE* = 0.01), *t*(203) = 21.20, *p* < .001, Cohen’s *d* = 1.48.

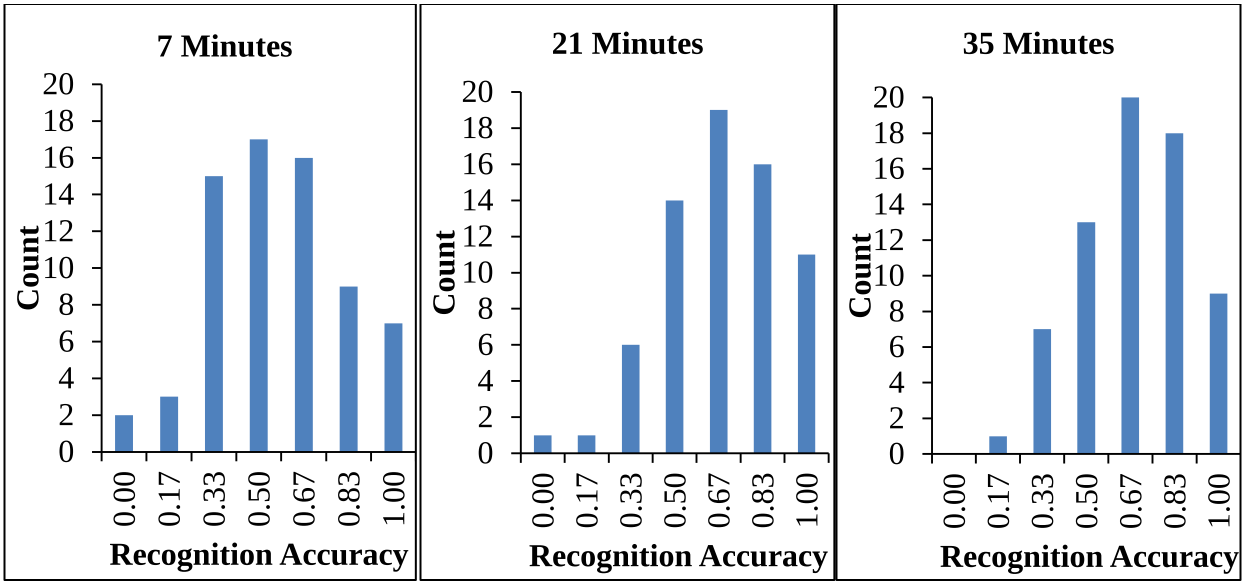

Figure 23. Frequency histograms of recognition accuracy by Session Length: 7 minutes (left panel), 21 minutes (center panel), and 35 minutes (right panel).

Condition and Session Length. Figure 22 showed effects of Session Length on recognition accuracy. It appeared that recognition was higher for the longer two session lengths, and that recognition was similar for 21 minutes and 35 minutes. I tested this apparent effect via an ANOVA of Condition and Length on recognition accuracy, which showed no main effect of Condition (p = .98), no interaction of Condition and Session Length (p = .59), and a main effect of Session Length, F(2,196) = 5.78, p = .004, *partial-eta-squared* = 0.05.

I used custom hypothesis tests in ANOVA to investigate the main effect of Session Length. I first compared 21 minutes (M = 0.68, SE = 0.03) and 35 minutes (M = 0.68, SE = 0.03) and found that they did not differ (p = .94). Then I compared the longer session lengths combined to 7 minutes (M = 0.57, SE = 0.03) of familiarization. I found that participants were significantly less accurate on the recognition test with the 7-minute session length than when they had more familiarization, F(1,196) = 11.15, p = .001, *partial-eta-squared* = 0.05. This was consistent with piloting (and why 21 minutes of familiarization was used for Experiment 1). However, participants with the 7-minute session length still showed higher recognition than chance, *t*(68) = 2.44, *p* = .02, Cohen's *d* = 0.29. Decreasing the session length decreased recognition.

Participants in all three versions of the psychophysical assessment pairs performed equally well. This was unsurprising because all participants received the same recognition test (relative to their familiarization shape set), but it indicated that there were not significant differences across participants in different assessment versions by chance. It was also important that Session Length did not interact with (assessment) Condition for recognition.

Table 1.

Session Length and Recognition Group on actual and expected counts.

	Nonrecognizers		Recognizers		
Session Length	Actual	Expected	Actual	Expected	Total
7 Minutes	37	(26.9)	32	(42.1)	69
21 Minutes	22	(26.5)	46	(41.5)	68
35 Minutes	21	(26.5)	47	(41.5)	68
Total	80		125		205

Session Length and Recognition Group. I investigated the relationship of Session Length and Recognition Group using a chi-squared test. There was a significant association between Session Length and Recognition Group, $\chi^2(2) = 9.35$, *p* = .009. Looking at Table 1, it appeared that the association of Session Length and Recognition Group was due to a higher percentage of participants in the 7 minutes group failing the recognition test than in the other session lengths.

Correlating Recognition and Assessment Accuracy. A Pearson correlation was used to investigate the relationship of recognition accuracy and average accuracy on the psychophysical assessment. Assessment accuracy and recognition accuracy were positively correlated: higher recognition accuracy predicted higher assessment accuracy, *r*(204) = 0.26, *p* < .001. ANCOVA analyses were used to follow up on this significant correlation (see Psychophysical Assessment: Effects of Noticing, Recognition, and Linguistic Coding, on page 99).

To further investigate the correlation between recognition accuracy and assessment accuracy, I split the data on both Condition and Session Length to test the correlation for each condition and session length combination. For Familiar, there was a significant positive correlation of recognition accuracy and assessment accuracy for a 21-minute session length, such that after 21 minutes, participants with higher recognition also showed higher assessment accuracy, *r*(22) = .62, *p* = .002. For Familiar, there was no correlation for 7 minutes, *r*(22) = -.20, *p* = .36; or 35 minutes, *r*(22) = .16, *p* = .47. For Shuffled, there was a significant positive correlation for 7 minutes of familiarization, such that participants with high recognition after 7 minutes of familiarization also showed higher assessment accuracy, *r*(23) = .53, *p* = .008. There was no correlation for Shuffled at 21 minutes, *r*(21) = .13, *p* = .55; or 35 minutes, *r*(22) = .22, *p* = .31. There was no correlation for New Shapes at 7 minutes, *r*(21) = .06, *p* = .79; or 21 minutes, *r*(22) = .31, *p* = .15; or 35 minutes, *r*(21) = .32, *p* = .16.

Psychophysical Assessment

Accuracy. Figure 24 showed the effects of Condition, Session Length, and Exposure Duration on accuracy. For 7 minutes of familiarization, learning appeared to be most distinct at the longest exposure duration. For 21 minutes, Familiar had the highest accuracy across exposure durations, but for 35 minutes, Shuffled showed the highest accuracy across exposure durations. To test these apparent effects, I conducted a four-way mixed ANOVA of Condition by Session Length by Exposure Duration by Target Presence on accuracy. I found significant main effects of Condition, *F*(2,196) = 3.29, *p* = .04, *partial-eta-squared* = 0.03, and Exposure Duration, *F*(3,588) = 3.05, *p* = .03, *partial-eta-squared* = 0.02. There was also a significant main effect of Target Presence, such that participants were more accurate when the target was Absent (*M* = 0.78, *SE* = 0.01) than when the target was Present (*M* = 0.72, *SE* = 0.01), *F*(1,196) = 14.66, *p* < .001, *partial-eta-squared* = 0.07. I also found three significant interactions: Condition marginally interacted with Exposure Duration *F*(6,588) = 1.94, *p* = .07, *partial-eta-squared* = 0.02; the effect of Condition depended upon the combined effects of Session Length and Exposure Duration, *F*(12,588) = 1.83, *p* = .04, *partial-eta-squared* = 0.04; and Exposure Duration interacted with Target Presence, *F*(3,579) = 5.18, *p* = .002, *partial-eta-squared* = 0.03. All other effects were non-significant (all *p*'s > .12).

Custom hypothesis tests in ANOVA were used to investigate the main effects. For the main effect of Condition, I first compared Familiar (*M* = 0.76, *SE* = 0.01) and Shuffled (*M* = 0.76, *SE* = 0.01), and found that they did not differ in accuracy (*p* = .50). Then I compared New Shapes (*M* = 0.72, *SE* = 0.01) to Familiar and Shuffled combined, and found that New Shapes showed significantly lower accuracy than Familiar and Shuffled, *F*(1,196) = 6.11, *p* = .01, *partial-eta-squared* = 0.03. The learning transferred from Familiar to Shuffled.

For the main effect of Exposure Duration, I first compared 400ms (*M* = 0.74, *SE* = 0.01) and 700ms (*M* = 0.74, *SE* = 0.01), and found that the shorter exposure durations did not show different accuracy (*p* = .52). Similarly, I next compared the longer exposure durations, 1000ms

(*M* = 0.76, *SE* = 0.01) and 1300ms (*M* = 0.76, *SE* = 0.01), and found that they did not differ in accuracy (*p* = .70). Finally, I compared 400ms and 700ms to 1000ms and 1300ms, and found that the shorter exposure durations combined showed marginally lower accuracy than the longer exposure durations combined, *F*(1,196) = 8.96, *p* = .003, *partial-eta-squared* = 0.04.

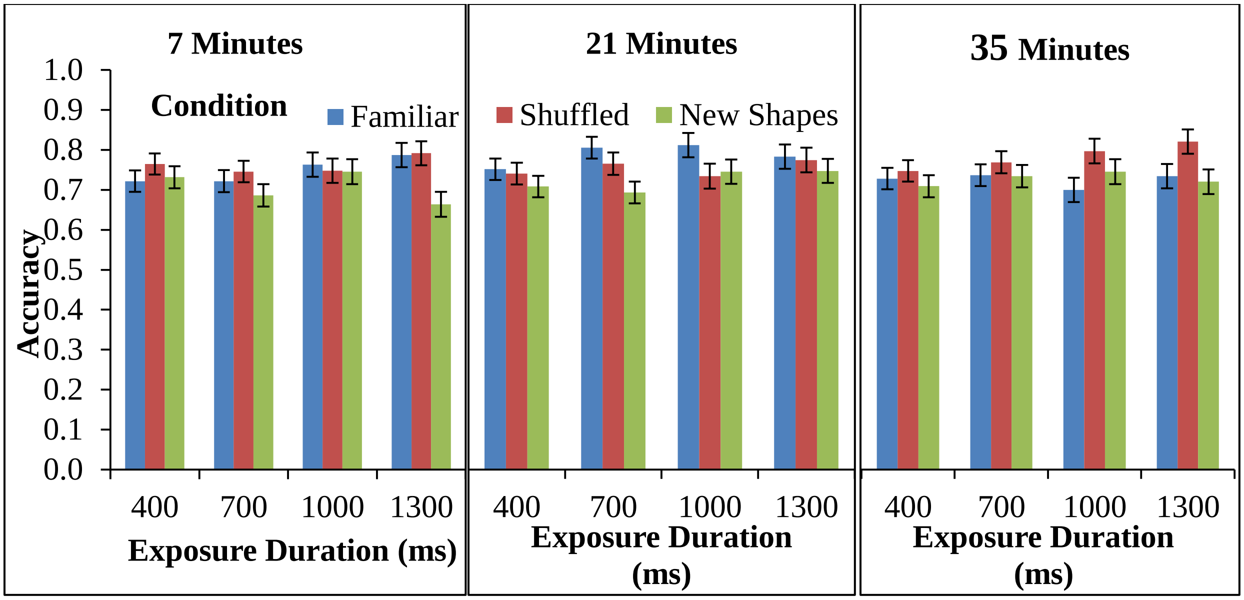

Figure 24. Condition by Session Length by Exposure Duration on accuracy (collapsed across Target Presence). Error bars indicate standard error of the mean.

Custom hypothesis tests were also used to examine the interaction of Condition and Exposure Duration, by testing the simple effect of Condition at each exposure duration. At 700ms, there was a significant simple effect of Condition, *F*(2,196) = 3.64, *p* = .03, *partial-eta-squared* = 0.04. I compared Familiar (*M* = 0.76, *SE* = 0.02) and Shuffled (*M* = 0.76, *SE* = 0.02) and found that they did not differ at 700ms (*p* = .81). Then I compared Familiar and Shuffled together to New Shapes (*M* = 0.71, *SE* = 0.02), and found that Familiar and Shuffled showed higher accuracy than New Shapes at 700ms, *F*(1,196) = 7.22, *p* = .008, *partial-eta-squared* = 0.04. There was also a simple effect of Condition at 1300ms, *F*(1,196) = 6.03, *p* = .003, *partial-eta-squared* = 0.06. Again, I first compared Familiar (*M* = 0.77, *SE* = 0.02) and Shuffled (*M* = 0.80, *SE* = 0.02) and found that they did not differ in accuracy at 1300ms (*p* = .27). When I compared New Shapes (*M* = 0.71, *SE* = 0.02) to Familiar and Shuffled combined, New Shapes showed lower accuracy than the combination, *F*(1,196) = 10.83, *p* = .001, *partial-eta-squared* = 0.05. There was no simple effect of Condition at 400ms (*p* = .30) or 1000ms (*p* = .82). The interaction of Condition and Exposure Duration was driven by the simple Condition effect at 700ms and 1300ms.

In looking at Figure 24, it appeared that the three-way interaction of Condition, Session Length, and Exposure Duration was due to Familiar showing the highest sensitivity for 21 minutes and Shuffled showing the highest sensitivity for 35 minutes, and to learning for the 7-minute exposure duration being apparent only for the longest exposure duration. To test the interaction, I broke the data on Exposure Duration and tested simple interactions for Condition and Session Length at each exposure duration. There was no simple interaction of Condition and Session Length at 400ms (*p* = .86), 700ms (*p* = .35), or 1300ms (*p* = .19). However, there was a marginal interaction of Condition and Session Length at 1000ms, *F*(4,196) = 2.16, *p* = .08, *partial-eta-squared* = 0.04. I used custom hypothesis tests to investigate the marginal interaction at 1000ms by evaluating simple simple effects of Condition for each session length. There was no simple simple effect of Condition for 7 minutes (*p* = .91) or 21 minutes (*p* = .15) of familiarization at 1000ms. However, there was a marginal simple simple effect of Condition for 35 minutes of familiarization and 1000ms, *F*(2,196) = 2.54, *p* = .08, *partial-eta-squared* = 0.03. I followed up by testing all pairwise comparisons. Familiar (*M* = 0.70, *SE* = 0.03) was less accurate than Shuffled (*M* = 0.80, *SE* = 0.03), *F*(1,196) = 5.07, *p* = .03, *partial-eta-squared* = 0.03. New Shapes (*M* = 0.75, *SE* = 0.03) did not differ in accuracy from Familiar (*p* = .30) or Shuffled (*p* = .24).

I also directly tested the apparent "flip" from Familiar having the numerically highest accuracy across exposure durations for 21 minutes to Shuffled having the numerically highest accuracy across exposure durations for 35 minutes by examining the interaction of Condition and Session Length in an ANOVA of Condition by Session Length (21, 35) by Exposure Duration by Target Presence on accuracy. There was no reliable interaction of Condition and Session Length across exposure durations and levels of Target Presence (*p* = .15).

I also directly tested the apparent effect of the learning for the 7-minute session length being restricted to the longest exposure duration. I tested the simple interaction of Condition and Exposure Duration for 7 minutes of familiarization[21]. For 7 minutes of Familiarization, there was a simple interaction of Condition and Exposure Duration on accuracy, *F*(6,198) = 2.94, *p* = .009, *partial-eta-squared* = 0.08. I followed up with custom hypothesis tests in ANOVA of simple simple effects of Condition at each exposure duration, and followed up significant simple simple effects with further custom hypothesis tests. There was a significant simple simple effect of Condition at 1300ms, *F*(2,66) = 5.59, *p* = .006, *partial-eta-squared* = 0.14. I compared Familiar (*M* = 0.79, *SE* = 0.03) and Shuffled (*M* = 0.79, *SE* = 0.03), and found that they did not differ in accuracy (*p* = .91). Then I compared Familiar and Shuffled together to New Shapes (*M* = 0.66, *SE* = 0.03), and found that the combination showed higher accuracy than New Shapes, *F*(1,66) =

[21] There was no simple interaction of Condition and Exposure Duration at 21minutes (*p* = .22) or 35 minutes (*p* = .20).

11.15, *p* = .001, *partial-eta-squared* = 0.15. There was not simple simple effect for 400ms, 700ms, or 1000ms (all *p*'s > .23). The three-way interaction was driven most clearly by the differences at 1000ms - the change from Familiar having the best performance for 21 minutes to Shuffled having the best at 35 minutes (and no learning at 7 minutes) - and by the learning only at 1300ms for 7 minutes.

Custom hypothesis tests were also used to investigate the interaction of Exposure Duration and Target Presence, by testing the simple effect of Target Presence at each exposure duration. There was a significant simple effect of Target Presence at 400ms, such that participants were more accurate on trials when the target was Absent (*M* = 0.78, *SE* = 0.01) than when the target was Present (*M* = 0.69, *SE* = 0.01), *F*(1,196) = 23.37, *p* < .001, *partial-eta-squared* = 0.11. At 700ms, participants were also more accurate for Absent (*M* = 0.77, *SE* = 0.01) than Present (*M* = 0.71, *SE* = 0.01), *F*(1,196) = 11.05, *p* = .001, *partial-eta-squared* = 0.05. 1000ms showed the same pattern of (marginally) higher accuracy for Absent (*M* = 0.78, *SE* = 0.01) than Present (*M* = 0.73, *SE* = 0.01), *F*(1,196) = 7.13, *p* = .007, *partial-eta-squared* = 0.04. In contrast, there was no simple effect of Target Presence for 1300ms (*p* = .29). The interaction of Exposure Duration and Target Presence was driven by the simple effects of Target Presence for the three shorter exposure durations.

Hit Rate. Figure 25 showed effects of Condition, Session Length, and Exposure Duration on hit rate. There appeared to be no strong effects. I tested the (lack of a clear) pattern of results with an ANOVA of Condition by Session Length by Exposure Duration on hit rate, which revealed only a significant main effect of Exposure Duration, *F*(3,588) = 7.33, *p* < .001, *partial-eta-squared* = 0.04. No other effects were significant (all *p*'s > .17).

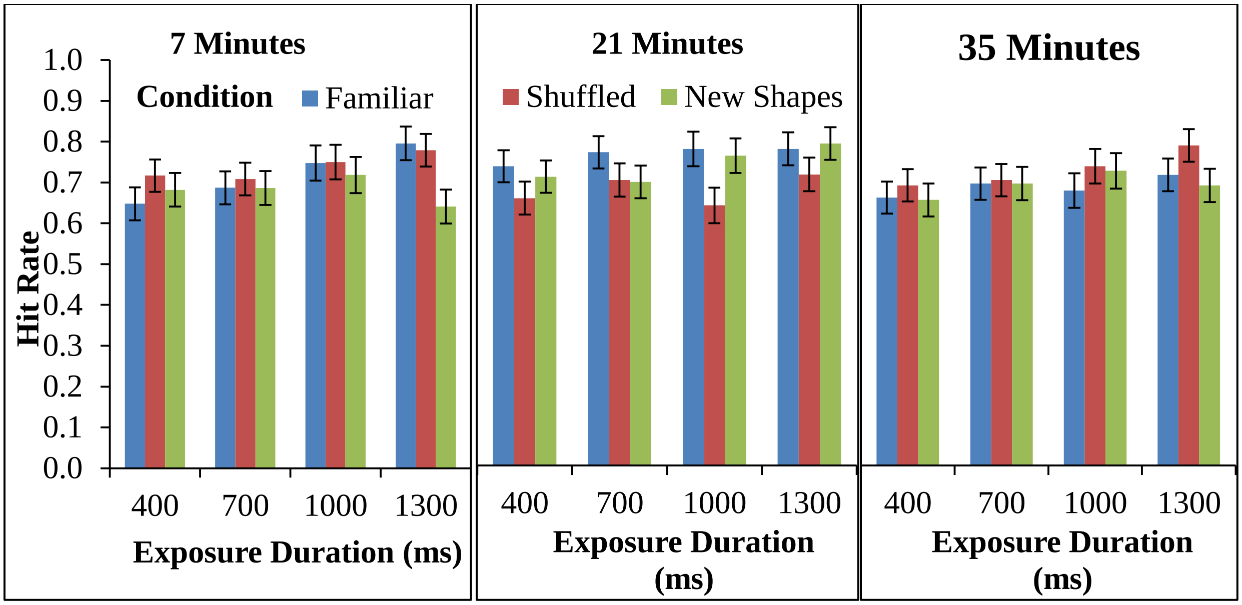

Figure 25. Condition by Session Length by Exposure Duration on hit rate. Error bars indicate standard error of the mean.

I used custom hypothesis to examine the main effect. I first compared 400ms (*M* = 0.69, *SE* = 0.01) and 700ms (*M* = 0.71, *SE* = 0.01) and found that they did not have different hit rates (*p* = .13). Next, I compared 1000ms (*M* = 0.73, *SE* = 0.01) and 1300ms (*M* = 0.75, *SE* = 0.01) and again found that they did not differ (*p* = .18). However, when I compared 400ms and 700ms to 1000ms and 1300ms, I found that the short exposure durations showed a significantly lower hit rate than the long exposure durations, *F*(1,196) = 17.41, *p* < .001, *partial-eta-squared* = 0.08.

False Alarm Rate. Figure 26 showed the effects of Condition, Session Length, and Exposure Duration on false alarm rates. Familiar and Shuffled appeared to generally have lower false alarm rates than New Shapes, but not for all exposure durations or all session lengths. To test the apparent effects, I conducted an ANOVA of Condition, Session Length, and Exposure

Duration on false alarm rate. This analysis revealed a main effect of Condition, *F*(2,196) = 5.54, *p* = .005, *partial-eta-squared* = 0.05, and no other effects (all *p*'s > .20).

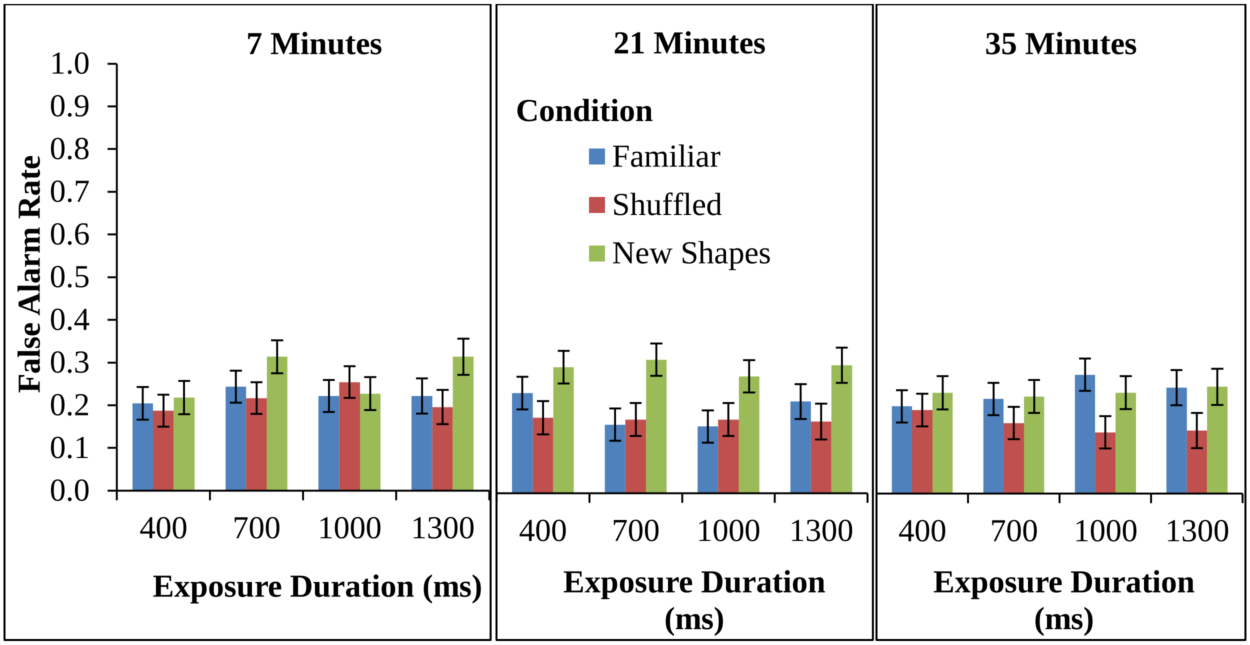

Figure 26. Condition by Session Length by Exposure Duration on false alarm rate. Error bars indicate standard error of the mean.

For the main effect of Condition, I followed up with custom hypothesis tests in ANOVA of all pairwise comparisons. Familiar (*M* = 0.22, *SE* = 0.02) showed marginally fewer false alarms than New Shapes (*M* = 0.27, *SE* = 0.02), *F*(1,196) = 3.82, *p* = .05 *partial-eta-squared* = 0.02. Similarly, Shuffled (*M* = 0.18, *SE* = 0.02) had fewer false alarms than New Shapes, *F*(1,196) = 10.98, *p* = .001, *partial-eta-squared* = 0.05. Familiar and Shuffled did not differ (*p* = .17). Learning in terms of reduced false alarm rates transferred from Familiar to Shuffled.

Sensitivity. Figure 27 showed the effects of Condition, Session Length, and Exposure Duration on sensitivity (d'). It appeared that Familiar and Shuffled tended to have higher sensitivity than New Shapes, but this appeared to depend on both Session Length and Exposure Duration. To test these apparent effects, I conducted an ANOVA of Condition by Session Length by Exposure Duration on sensitivity, which demonstrated a significant main effect of Condition *F*(2,196) = 4.05, *p* = .02, *partial-eta-squared* = 0.04. It also showed a main effect of Exposure Duration *F*(3,588) = 3.90, *p* = .009, *partial-eta-squared* = 0.02, and an interaction of Condition, Session Length, and Exposure Duration, *F*(12,588) = 1.82, *p* = .04, *partial-eta-squared* = 0.04. No other effects were found (all *p*'s > .13).

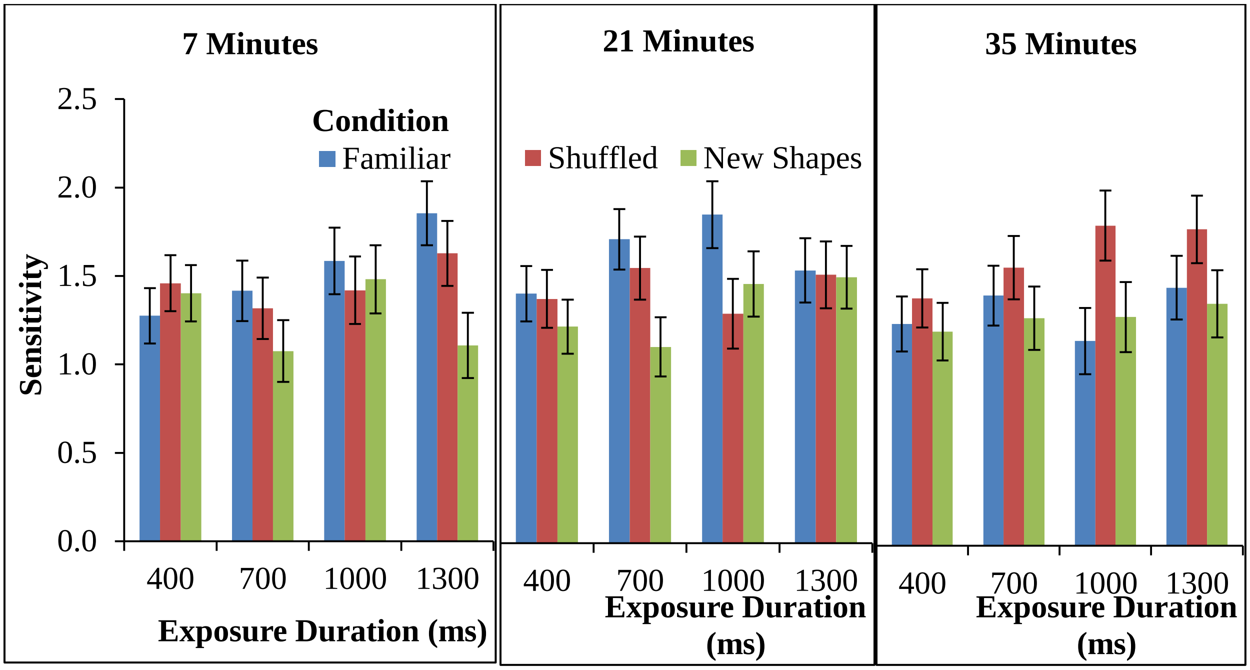

Figure 27. Condition by Session Length by Exposure Duration on sensitivity. Error bars indicate standard error of the mean.

In looking at Figure 27, it appeared that the three-way interaction of Condition, Session Length, and Exposure Duration was due to Familiar showing the highest sensitivity for 21 minutes and Shuffled showing the highest sensitivity for 35 minutes, and to learning for the 7-minute exposure duration being apparent only for the longest exposure duration. To test the interaction, I broke the data on Exposure Duration and tested simple interactions for Condition and Session Length at each exposure duration. For 400ms, 700ms, and 1300ms there were no simple interactions (all *p*'s > .44). However, at 1000ms there was a significant simple interaction of Condition and Session Length, *F*(4,196) = 2.90, *p* = .02, *partial-eta-squared* = 0.06. Custom hypothesis tests in ANOVA were used to test simple simple effects of Condition at each Session Length at 1000ms, and to follow up on any significant simple simple effects with pairwise tests of conditions. There was no simple simple effect of Condition at the 7-minute session length and at the 1000ms exposure duration (*p* = .86). At 21 minutes and 1000ms, there was a marginal simple simple effect of Condition, *F*(2,196) = 2.72, *p* = .07, *partial-eta-squared* = 0.03. Familiar (*M* = 1.91, *SE* = 0.19) was more sensitive than Shuffled (*M* = 1.32, *SE* = 0.19), *F*(1,196) = 4.67, *p* = .03, *partial-eta-squared* = 0.02. Familiar was also marginally more sensitive than New Shapes (*M* = 1.42, *SE* = 0.19), *F*(1,196) = 3.35, *p* = .07, *partial-eta-squared* = 0.02. Shuffled and New Shapes did not differ (*p* = .73). At 35 minutes and 1000ms, there was a significant simple simple effect of Condition, *F*(2,196) = 3.43, *p* = .03, *partial-eta-squared* = 0.03. Familiar (*M* = 1.10, *SE* = 0.19) was significantly less sensitive than Shuffled (*M* = 1.79, *SE* = 0.19), *F*(1,196) = 6.65, *p* = .01, *partial-eta-squared* = 0.03. Shuffled was also marginally more sensitive than New Shapes (*M* = 1.34, *SE* = 0.19), *F*(1,196) = 2.79, *p* = .096, *partial-eta-squared* = 0.01. Familiar and New Shapes did not differ (*p* = .38).

I also directly tested the apparent "flip" from Familiar showing the numerically highest performance across exposure durations for 21 minutes to Shuffled showing the numerically highest performance across exposure durations for 35 minutes by examining the interaction of

Condition and Session Length in an ANOVA of Condition by Session Length (21, 35) by Exposure Duration on sensitivity. There was no reliable interaction of Condition and Session Length across exposure duration (*p* = .23).

I also directly tested the apparent effect of the learning for the 7-minute session length being restricted to the longest exposure duration. I tested the simple interaction of Condition and Exposure Duration for 7 minutes of familiarization[22]. For 7 minutes of Familiarization, there was a simple interaction of Condition and Exposure Duration on sensitivity, *F*(6,198) = 2.44, *p* = .03, *partial-eta-squared* = 0.07. At 1300ms there was a simple simple effect of Condition, *F*(2,66) = 4.34, *p* = .02, *partial-eta-squared* = 0.12. I followed this up with additional custom hypothesis tests: I compared Familiar (*M* = 1.66, *SE* = 0.19) to Shuffled (*M* = 1.68, *SE* = 0.19) and found that they did not differ in sensitivity at 1300ms and 7 minutes (*p* = .95). Then I compared Familiar and Shuffled New Shapes (*M* = 0.95, *SE* = 0.20) and found that Familiar and Shuffled had significantly higher sensitivity than New Shapes at 7 minutes and 1300ms, *F*(1,66) = 8.67, *p* = .004, *partial-eta-squared* = 0.12. For 7 minutes of familiarization, there was no simple simple effect of Condition for 400ms (*p* = .20), 700ms (*p* = .21), or for 1000ms (*p* = .82). The three-way interaction was driven most clearly by the differences at 1000ms - the change from Familiar having the best performance for 21 minutes to Shuffled having the best at 35 minutes (and no learning at 7 minutes) - and by the learning only at 1300ms for 7 minutes.

Bias. Figure 28 showed effects of Condition, Session Length, and Exposure Duration on bias (the tendency to respond present or absent), as measured by criterion. Criterion appeared to differ by Condition and Exposure Duration. To test these apparent effects, I conducted an

[22] There was no simple interaction of Condition and Exposure Duration at 21minutes (*p* = .13) or 35 minutes (*p* = .31).

ANOVA of Condition by Session Length by Exposure Duration on criterion, and found only a main effect of Exposure Duration *F*(3,588) = 4.71, *p* = .003, *partial-eta-squared* = 0.02. No other effects were significant (all *p*'s > .11).

Custom hypothesis tests of all pairwise comparisons demonstrated that 400ms (*M* = -0.15, *SE* = 0.03) showed more bias to respond absent than 1300ms (*M* = -0.04, *SE* = 0.03), *F*(1,196) = 12.20, *p* = .001, *partial-eta-squared* = 0.06. No other pairwise comparisons were significant (all *p*'s > .01).

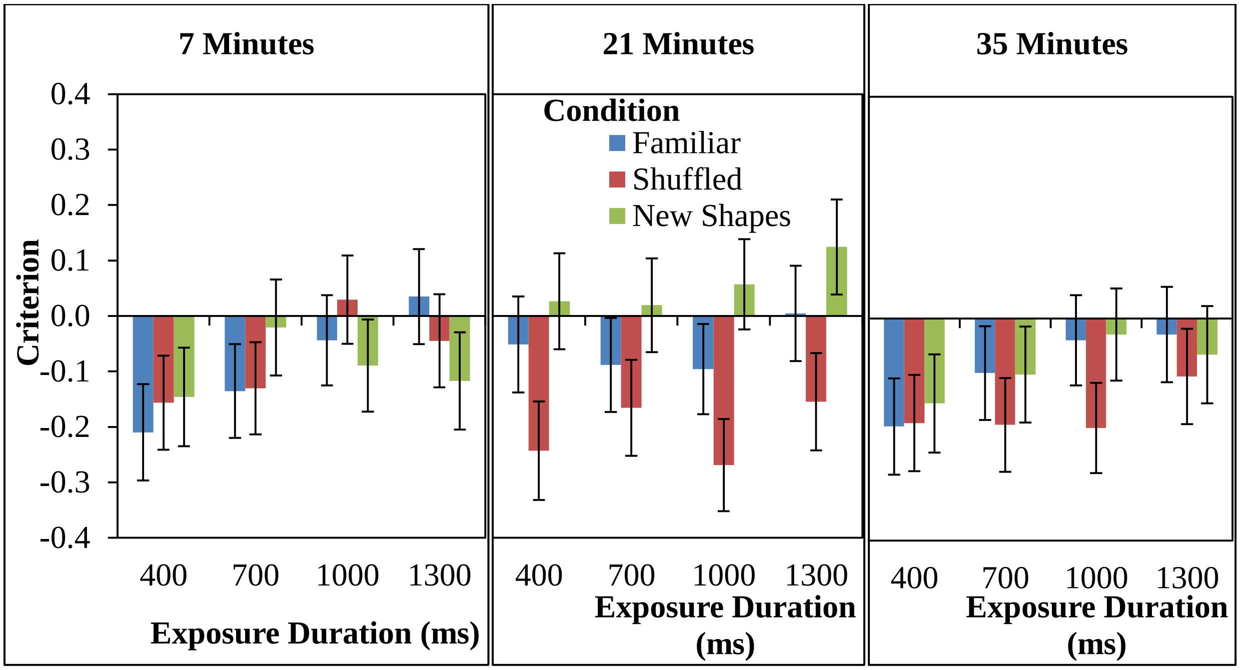

Figure 28. Condition by Session Length by Exposure Duration on criterion. Error bars indicate standard error of the mean.

Survey Data

Noticing the Pairs. Forty-one participants reported noticing the pairs, 39 likely noticed the pairs, 111 did not, and 14 did not respond to the question. Noticing the pairs was positively correlated with recognition accuracy, such that participants who reported Noticing the pairs were more accurate, *r*(190) = .34, *p* < .001. Noticing the pairs was also positively correlated with assessment accuracy, again with Noticing being associated with higher accuracy, *r*(190) = .28, *p* < .001.

An ANCOVA of Condition by Session Length on recognition accuracy, covarying out Noticing, was used to follow up on the correlation of recognition accuracy and Noticing. This analysis demonstrated a significant effect of the covariate, *F*(1,181) = 19.52, *p* < .001, *partial-eta-squared* = 0.10; and a main effect of Session Length, *F*(2,181) = 3.88, *p* = .02, *partial-eta-squared* = 0.04. No other effects were significant (all *p*'s > .59). I used custom hypothesis tests in ANCOVA on the adjusted marginal means[23] to investigate the main effect of Session Length. I first compared 21 minutes (*M* = 0.67, *SE* = 0.03) and 35 minutes (*M* = 0.68, *SE* = 0.03) and found that they did not differ (*p* = .69). Then I combined the long session lengths and compared them to 7 minutes (*M* = 0.58, *SE* = 0.03) and found that 21 minutes and 35 minutes together showed marginally higher recognition than 7 minutes, *F*(1,181) = 7.63, *p* = .006, *partial-eta-squared* = 0.04. Reducing the session length reduced recognition.

Strategy. Seventy-nine participants used the Linguistic Coding strategy and 126 did not. Independent-samples *t*-tests showed no advantage of Linguistic Coding for recognition accuracy (*p* = 38). However, participants who used the strategy (*M* = 0.78, *SE* = 0.01) were more accurate than those who did not (*M* = 0.73, *SE* = 0.01), *t*(203) = 3.34, *p* = .001, *Cohen's d* = 0.23.

Alertness and Sleep. Mean alertness was 3.03 (*SE* = 0.06). Alertness did not correlate with recognition accuracy (*p* = .08) or with assessment accuracy (*p* = .14). Mean number of

[23] Adjusted marginal means were calculated at the average of Noticing, M = .32

hours slept the prior evening was 6.31 hours (*SE* = 0.11). Sleep did not correlate with recognition accuracy (*p* = .75), or with assessment accuracy (*p* = .66).

Psychophysical Assessment: Effects of Noticing, Recognition, and Linguistic Coding

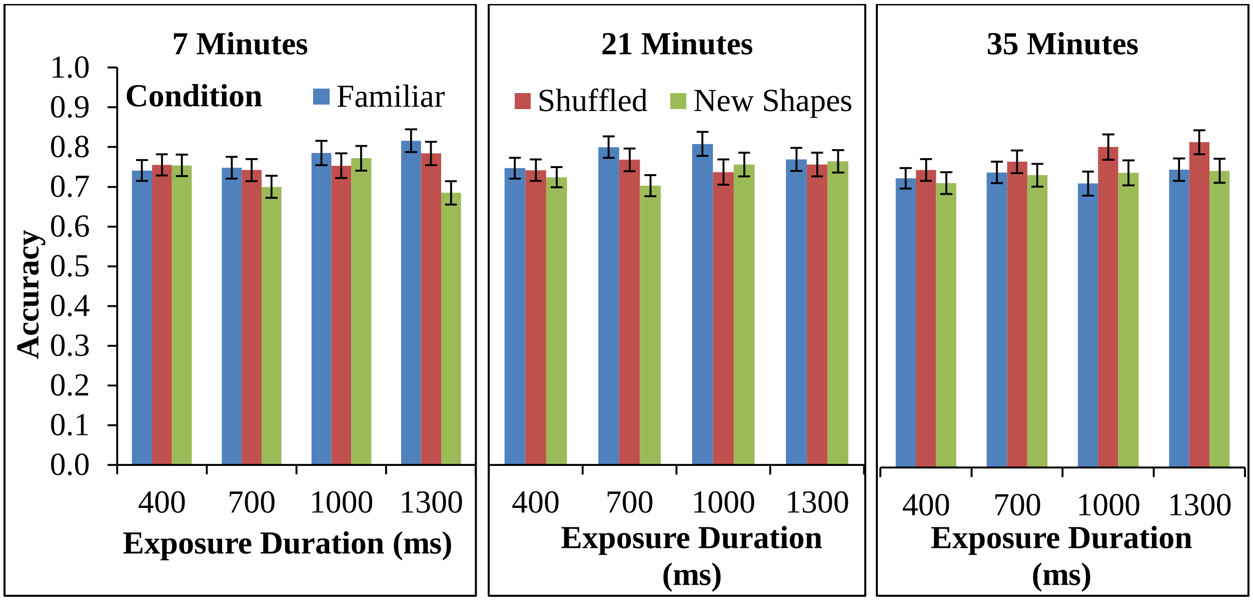

Figure 29. Condition by Session Length by Exposure Duration on accuracy (collapsed across Target Presence). Bar heights indicate adjusted marginal means and error bars indicate standard error of the mean.

Accuracy. Figure 29 shows the effects of Condition, Session Length, and Exposure Duration on accuracy, statistically controlling for the effects of Noticing, Recognition, and Linguistic Coding. It appears that Familiar showed the highest accuracy across exposure durations for 21 minutes, Shuffled showed the highest accuracy across exposure durations for 35 minutes, and that learning only occurred for the longest exposure duration for 7 minutes. To test these apparent effects, I conducted an ANCOVA of Condition by Session Length by Exposure

Duration by Target Presence on accuracy, covarying out Recognition performance, Noticing the pairs, and use of the Linguistic Coding strategy. There was a main effect of Target Presence, such that Present (*M* = 0.73, *SE* = 0.01) was less accurate than Absent (*M* = 0.78, *SE* = 0.01), *F*(1,179) = 4.88, *p* = .03, *partial-eta-squared* = 0.03. There were also significant effects of all three covariates: Noticing, *F*(1,179) = 4.37, *p* = .04, *partial-eta-squared* = 0.02; Recognition, *F*(1,179) = 9.71, *p* = .002, *partial-eta-squared* = 0.05; and Linguistic Coding, *F*(1,179) = 5.46, *p* = .02, *partial-eta-squared* = 0.03. There were interactions of Exposure Duration and Noticing, *F*(3,537) = 3.09, *p* = .03, *partial-eta-squared* = 0.02; Condition and Session Length and Exposure Duration, *F*(12,537) = 1.96, *p* = .03, *partial-eta-squared* = 0.04; and Exposure Duration and Target Presence, *F*(3,537) = 2.14, *p* = .095, *partial-eta-squared* = 0.01.

In looking at Figure 29, it appeared that the three-way interaction of Condition, Session Length, and Exposure Duration was due to Familiar showing the highest sensitivity for 21 minutes and Shuffled showing the highest sensitivity for 35 minutes, and to learning for the 7-minute exposure duration being apparent only for the longest exposure duration. To test the interaction, I broke the data on Exposure Duration and tested simple interactions for Condition and Session Length at each exposure duration. For significant simple interactions, I followed up with custom hypothesis tests in ANCOVA of simple simple effects of Condition, and additional custom tests for significant simple simple effects. There was a marginal simple interaction of Condition and Session Length for 1300ms, *F*(4,179) = 2.26, *p* = .07, *partial-eta-squared* = 0.05. There was a significant simple simple effect of Condition at 1300ms and 7 minutes, *F*(1,179) = 5.57, *p* = .005, *partial-eta-squared* = 0.06. I compared Familiar (*M* = 0.82, *SE* = 0.03) and Shuffled (*M* = 0.78, *SE* = 0.03) and found that they did not differ in accuracy (*p* = .61), but when I compared them together to New Shapes (*M* = 0.69, *SE* = 0.03), I found that Familiar and

Shuffled were more accurate than New Shapes, *F*(1,179) = 10.45, *p* = .001, *partial-eta-squared* = 0.06. No other simple simple effects of Condition were significant (all *p*'s > .15). No other simple interactions of Condition and Session Length were significant (all *p*'s > .10).

I also directly tested the apparent "flip" from Familiar showing the numerically highest performance across exposure durations for 21 minutes to Shuffled showing the numerically highest performance across exposure durations for 35 minutes by examining the interaction of Condition and Session Length in an ANOVA of Condition by Session Length (21, 35) by Exposure Duration on sensitivity. There was no reliable interaction of Condition and Session Length across exposure duration (*p* = .15).

I also directly tested the apparent effect of the learning for the 7-minute session length being restricted to the longest exposure duration. I tested the simple interaction of Condition and Exposure Duration for 7 minutes of familiarization[24]. For 7 minutes of Familiarization, there was a simple interaction of Condition and Exposure Duration in accuracy, *F*(6,174) = 2.97, *p* = .009, *partial-eta-squared* = 0.09. I followed up with custom hypothesis tests of simple simple effects of Condition, and additional tests for significant simple simple effects. There was a significant simple simple effect of Condition at 1300ms, *F*(2,58) = 5.07, *p* = .009, *partial-eta-squared* = 0.15. I compared Familiar (*M* = 0.82, *SE* = 0.03) and Shuffled (*M* = 0.76, *SE* = 0.03), and they did not differ (*p* = .47). Then I compared Familiar and Shuffled together to New Shapes (*M* = 0.69, *SE* = 0.03), *F* (1,58) = 3.52, *p* = .003, *partial-eta-squared* = 0.14. There were no other simple simple effects of Condition (all *p*'s > .28). The three-way interaction was driven most

[24] There was no simple interaction of Condition and Exposure Duration at 35 minutes (*p* = .56). There was a marginal simple interaction of Condition and Exposure Duration at 21minutes (*p* = .09), but there was no reliable simple simple effect of Condition at any exposure duration (all *p*'s > .10).

clearly by the differences at 1000ms - the change from Familiar having the best performance for 21 minutes to Shuffled having the best at 35 minutes (and no learning at 7 minutes) - and by the learning only at 1300ms for 7 minutes.

For the interaction of Exposure Duration and Target Presence, custom hypothesis tests in ANCOVA were used to test simple effects of Target Presence at each exposure duration. There was a marginal simple effect of Target Presence for 400ms such that participants were less accurate when the target was Present (*M* = 0.70, *SE* = 0.01) than when it was Absent (*M* = 0.78, *SE* = 0.01), *F*(1,179) = 8.16, *p* = .005, *partial-eta-squared* = 0.04. No other simple effects were significant (all *p*'s > .03). The simple effect of Target Presence at 400ms drove the interaction of Exposure Duration and Target Presence.

False Alarm Rate. Figure 30 showed effects of Condition, Session Length, and Exposure Duration on false alarm rate. Familiar and Shuffled appeared to have lower false alarm rates across exposure durations and session lengths, but this pattern did not appear to hold for all exposure durations for 35 minutes. To test these apparent effects, I conducted an ANCOVA of Condition by Session Length by Exposure Duration on false alarm rate, covarying out Recognition performance and Noticing of the pairs and use of the Linguistic Coding strategy, which revealed a marginal main effect of Condition, *F*(1,179) = 3.04, *p* = .05, *partial-eta-squared* = 0.03. It also revealed significant effects of two covariates: Noticing, *F*(1,179) = 4.10, *p* = .04, *partial-eta-squared* = 0.02; and Linguistic Coding, *F*(1,179) = 4.49, *p* = .04, *partial-eta-squared* = 0.02. There was no main effect of Session Length (*p* = .67), and no other effects were significant (all *p*'s > .20).

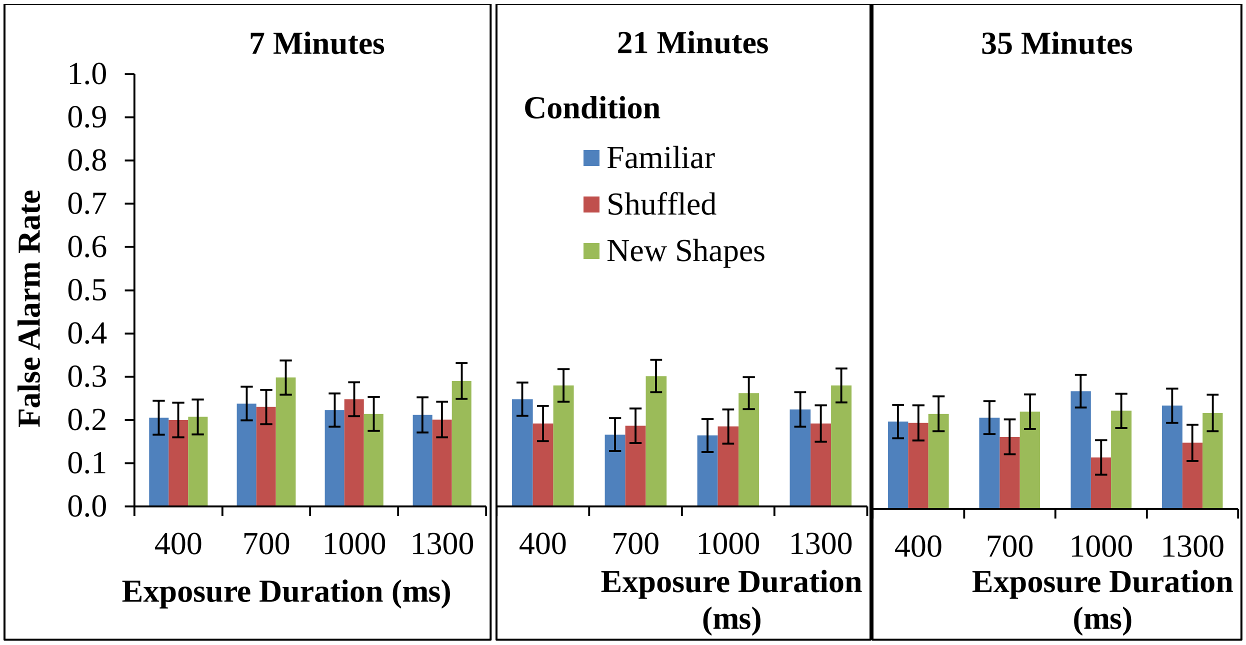

Figure 30. Condition by Session Length by Exposure Duration on false alarm rate. Bar heights indicate adjusted marginal means and error bars indicate standard error of the mean.

Custom hypothesis tests in ANCOVA on the adjusted marginal means of all pairwise comparisons were used to investigate the main effect of Condition. Shuffled (*M* = 0.19, *SE* = 0.02) showed fewer false alarms than New Shapes, *F*(1,179) = 6.05, *p* = .02, *partial-eta-squared* = 0.03. Familiar (*M* = 0.22, *SE* = 0.02) did not differ from Shuffled (*p* = .27) or New Shapes (*p* = .16). Learning, in terms of decreased false alarming, was cleared for Shuffled, a transfer condition.

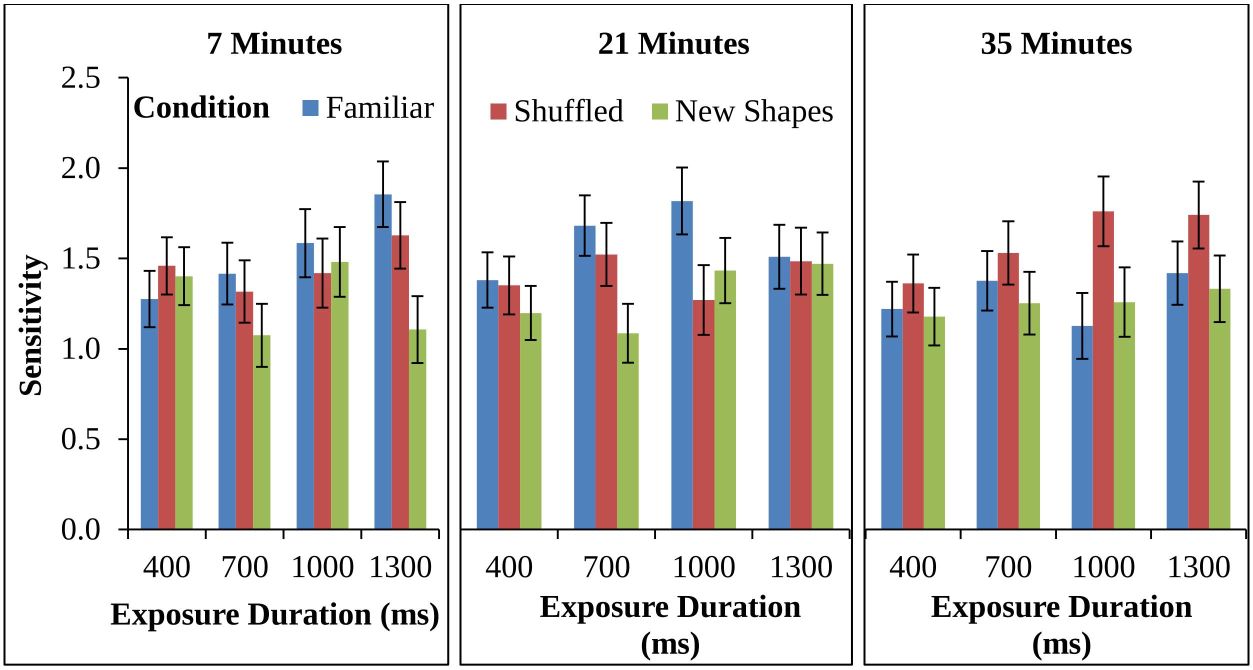

Figure 31. Condition by Session Length by Exposure Duration on sensitivity. Bar heights indicate adjusted marginal means and error bars indicate standard error of the mean.

Sensitivity. Figure 31 showed effects of Condition, Session Length, and Exposure Duration on sensitivity, statistically controlling for Recognition, Noticing, and Linguistic Coding. It appears that for 7 minutes, learning was only demonstrated for the longest exposure duration. For 21 minutes, Familiar had the highest sensitivity across exposure durations, but for 35 minutes, Shuffled had the highest sensitivity across exposure durations. To test these apparent effects, I conducted an ANCOVA of Condition by Session Length by Exposure Duration on sensitivity, covarying out Recognition performance and Noticing the pairs and use of the Linguistic Coding strategy, which showed significant effects of all the covariates: Recognition, *F*(1,179) = 11.58, *p* = .001, *partial-eta-squared* = 0.06; Noticing, *F*(1,179) = 5.25, *p* = .02, *partial-eta-squared* = 0.03; and Linguistic Coding, *F*(1,179) = 5.45, *p* = .02, *partial-eta-squared*

= 0.03. There were interactions of Condition and Session Length and Exposure Duration, *F*(12,537) = 1.99, *p* = .02, *partial-eta-squared* = 0.04, and Exposure Duration with Noticing, *F*(3,537) = 3.84, *p* = .01, *partial-eta-squared* = 0.02. There was no main effect of Condition (*p* = .10) or of Session Length (*p* = .73), and no other effects were significant (all *p*'s > .29).

In looking at Figure 31, it appeared that the three-way interaction of Condition, Session Length, and Exposure Duration was due to Familiar showing the highest sensitivity for 21 minutes and Shuffled showing the highest sensitivity for 35 minutes, and to learning for the 7-minute exposure duration being apparent only for the longest exposure duration. To test the interaction, I broke the data on Exposure Duration and tested simple interactions for Condition and Session Length at each exposure duration. Significant simple interactions were followed up with custom hypothesis tests in ANCOVA on the adjusted marginal means of simple simple effects of Condition for each session length. For 1000ms, there was a simple interaction of Condition and Session Length, *F*(1,179) = 2.64, *p* = .035, *partial-eta-squared* = 0.06. There was a simple simple effect of Condition at 1000ms and 35 minutes, *F*(2,179) = 3.12, *p* = .05, *partial-eta-squared* = 0.03. I tested all pairwise comparisons of conditions at 1000ms and 35 minutes. Familiar (*M* = 1.15, *SE* = 0.19) and showed lower sensitivity than Shuffled (*M* = 1.80, *SE* = 0.19), *F*(1,179) = 5.71, *p* = .02, *partial-eta-squared* = 0.03. New Shapes (*M* = 1.29, *SE* = 0.18) also showed lower sensitivity than Shuffled, *F*(1,179) = 3.43, *p* = .07, *partial-eta-squared* = 0.02. Familiar and New Shapes did not differ in sensitivity for 1000ms and 35 minutes (*p* = .62). No other simple simple effects of Condition were significant at 1000ms minutes (all *p*'s > .10). There were no simple interactions of Condition and Session Length for 400ms, 700ms, or 1300ms (all *p*'s > .18).

I also directly tested the apparent “flip” from Familiar showing the numerically highest performance across exposure durations for 21 minutes to Shuffled showing the numerically highest performance across exposure durations for 35 minutes by examining the interaction of Condition and Session Length in an ANOVA of Condition by Session Length (21, 35) by Exposure Duration on sensitivity. There was no reliable interaction of Condition and Session Length across exposure duration (*p* = .23).

I also directly tested the apparent effect of the learning for the 7-minute session length being restricted to the longest exposure duration. I tested the simple interaction of Condition and Exposure Duration for 7 minutes of familiarization[25]. For 7 minutes of Familiarization, there was a simple interaction of Condition and Exposure Duration on sensitivity, *F*(6,174) = 2.56, *p* = .02, *partial-eta-squared* = 0.08. I followed up with custom hypothesis tests of simple simple effects of Condition, and followed up significant simple simple effects with additional custom hypothesis tests. There was a significant simple simple effect of Condition at 1300ms, *F*(2,58) = 3.92, *p* = .03, *partial-eta-squared* = 0.12. I compared Familiar (*M* = 1.86, *SE* = 0.18) and Shuffled (*M* = 1.63, *SE* = 0.18), and found no difference (*p* = .41). Then I compared the average of Familiar and Shuffled to New Shapes (*M* = 1.11, *SE* = 0.19), and found that Familiar and Shuffled showed a higher level of sensitivity than New Shapes, *F*(1,58) = 7.09, *p* = .01, *partial-eta-squared* = 0.11. No other simple simple effects of Condition were reliable (all *p*’s > .19). The three-way interaction was driven most clearly by the differences at 1000ms - the change from

[25] There was no simple interaction of Condition and Exposure Duration at 35 minutes (*p* = .44). There was a marginal simple interaction of Condition and Exposure Duration at 21minutes, *F*(6,177) = 2.78, *p* = .04, *partial-eta-squared* = 0.07, but there were no reliable simple simple effects of Condition at 21 minutes (all *p*’s > .12).

Familiar having the best performance for 21 minutes to Shuffled having the best at 35 minutes (and no learning at 7 minutes) - and by the learning only at 1300ms for 7 minutes.

Discussion

Experiment 3 was designed to directly replicate and build on Experiment 1: 1) to replicate the finding that the familiarization in a well-known statistical learning (SL) paradigm (Fiser & Aslin, 2001) caused perceptual learning (PL) and 2) to expand it to shorter and longer familiarization session lengths. I replicated the significant SL found by the authors of the paradigm (Fiser & Aslin, 2001) as we did in Experiment 1, and critically, I replicated Experiment 1's findings of PL effects: transfer and improved sensitivity.

Comparing my Data to the Hypotheses

The Experiment 3 data do not support the hypothesis of separate kinds of learning, but also do not wholly align with the predictions of the hypothesis of a unified learning process. The hypothesis of separate kinds of learning predicts recognition, no differences in sensitivity across conditions, no transfer, and, possibly, an increased false alarm rate for Shuffled. The hypothesis of a unified learning process predicts a correlation between recognition accuracy and assessment accuracy, especially for conditions that show learning, and PL effects - improved sensitivity and transfer.

I replicated our Experiment 1 assessment and recognition results: I again found hallmarks of PL – transfer of learning and improved psychophysical sensitivity. Participants showed high accuracy and sensitivity for both Familiar and Shuffled conditions across session lengths and exposure durations even though in the Shuffled condition the shapes were shuffled from the (Familiar) pairings seen in the familiarization into new pairings in the search grids. Accuracy and sensitivity for these conditions were significantly higher than for New Shapes. Similarly, participants had low false alarm rates for Familiar and Shuffled and significantly higher false alarm rates for New Shapes across session lengths and exposure durations, replicating the pattern of assessment results in Experiment 1 of PL effects following an SL paradigm

But I also found an overall correlation of recognition and assessment accuracy, unlike in Experiment 1. Perhaps the higher power due to higher total sample size in Experiment 2 explains this difference, because Experiment 1 also had a positive correlation, but it was non-significant. Or perhaps there was some difference by chance between the samples of participants (both from the same population) that caused the different results. Either way, the differing correlation results between Experiments 1 and 3, in addition to the observed PL effects, give additional evidence that the relationship between SL and PL may be complex.

Session Length

Recognition. I predicted that recognition accuracy would increase with a longer session length and decrease with a shorter session length. I found a main effect of Session Length for recognition accuracy, such that accuracy was lower for the 7 minute session length than the 21 and 35 minute session lengths, which did not differ from each other. The fact that 7 minutes showed significantly lower recognition than the other session lengths matched my prediction, as well as Experiment 1 piloting results[26]. I did not expect that increasing the session length would not increase recognition accuracy. Perhaps a session length of at least 21 minutes represents some kind of ceiling for passive learning, at least without overlearning or longer spacing between familiarization blocks.

[26] Because our piloting results for 7 minutes of familiarization showed recognition near chance, we increased our session length to 21 minutes for Experiment 1 to have more SL and therefore a stronger test of whether SL and PL are a single learning process or separate processes.

The chi-square analysis revealed that longer session lengths improved recognition accuracy primarily by increasing the proportion of participants who passed the recognition test (answered more than 50% of the recognition trials correctly). With a 7 minute session length, more than half of participants failed the recognition test. But with a longer session length, more than two-thirds of participants passed the recognition test.

Psychophysical Assessment. There were no main effects of Session Length for any assessment analyses. Main effects of Condition for accuracy, false alarm rate, and sensitivity held across session lengths and exposure durations, so PL effects were observed for all session lengths. It was possible that reducing the session length could have qualitatively changed the kinds of effects seen from PL effects to possible SL effects or a mix of SL and PL effects, but reducing the session length did not change the pattern of results: in the three-way interactions of Condition, Session Length, and Exposure Duration, at the 7-minute session length learning was only observed at the longest exposure duration. Participants needed a longer exposure to assessment arrays to demonstrate their learning, but they did show qualitatively the same condition effect observed with the other session lengths (including Experiment 1). This is evidence of weaker but qualitatively similar learning. This was true in the original ANOVA analyses and in the ANCOVA analyses, which statistically controlled for the effects of Recognition accuracy (as well as Noticing the pairs and use of the Linguistic Coding strategy).

Session Length significantly interacted with Condition and Exposure Duration for accuracy and sensitivity. In addition to observed learning being restricted to the longest exposure duration for 7 minutes of familiarization (as discussed above) at 1000ms, for 21 minutes Familiar showed the highest performance, but for 35 minutes the transfer condition Shuffled did the best.

Shuffled having higher sensitivity than Familiar is not what either PL or SL would predict, but it is evidence of transfer, a PL effect.

Correlations. There was a significant correlations of recognition (SL) and assessment accuracy (PL) overall, and there were two significant correlations when the data were divided by session length and condition. One was for Shuffled at a 7-minute session length and the other was for Familiar at a 21-minute session length. These appear to be related to the results of the assessment three-way interaction: the former correlation was related to learning for 7 minutes for the longest exposure duration, and the latter was related to Familiar showing the highest performance for 21 minutes. The correlations suggest that these PL effects were related to SL.

Recognition, Noticing, and Linguistic Coding

Several measures correlated with psychophysical assessment accuracy in Experiment 3: Recognition test accuracy, Noticing of the pairs, and use of the Linguistic Coding strategy. Higher Recognition predicted higher assessment accuracy, as did Noticing the pairs, and using the Linguistic Coding strategy. In ANCOVA analyses of psychophysical assessment data using these measures as covariates, there was only a main effect of Condition for false alarms. Condition was only involved in three-way interactions with Session Length and Exposure Duration for accuracy and sensitivity, not also in main effects, unlike in the main assessment analyses. Otherwise, the patterns of ANCOVA results were similar to the ANOVA results, showing transfer and improved sensitivity, evidence of PL.

Summary

In sum, I replicated and extended the findings of Experiment 1. I replicated the finding of PL effects following a SL paradigm: I found transfer of learning from Familiar to Shuffled in both psychophysical assessment accuracy and sensitivity, and improved sensitivity (relative to

New Shapes), in both the main analyses and in the ANCOVA analyses. I showed that increasing session length beyond 21 minutes did not impact performance in recognition or on the psychophysical assessment. Reducing the session length did not change the overall pattern of assessment results. However, reducing the session length reduced the proportion of participants who passed the recognition test, and reduced the exposure durations in the assessment at which Condition effects were observed to only the longest – 1300ms. The Learning was present, but weaker with a shorter session length. Unlike in Experiment 1, recognition (SL) and assessment accuracy (PL) significantly correlated, and this seemed to be related to assessment three-way interaction results. Together, all the results again give evidence that the relationship between SL and PL may be nuanced.

Next Steps

I qualitatively replicated the findings of Experiment 1, but it would be helpful to verify that the pattern holds quantitatively as well, specifically at 21 minutes, the replicated session length. I weakened learning and reduced performance by shortening the session length, so it would be helpful to compare performance at the session length of 7 minutes to Baseline performance to see if the weak learning was above Baseline. I did not successfully increase performance by increasing session length, but comparing results of the longest session length to results of Experiment 2, the PL training experiment, could further illuminate if performance with a long session length approximates results following PL training. It would also be interesting to quantitatively compare performance across experiments and Baseline.

CHAPTER 8: MULTI-EXPERIMENT ANALYSES

Comparing Across Experiments on the Psychophysical Assessment

In addition to analyses for each experiment individually, I also analyzed the experiments together and against baseline psychophysical assessment performance. I used the same assessment for all three experiments as well as for the Baseline sample, so assessment performance could be compared and analyzed together. By doing so, I was able to statistically explore additional questions about my data that could only be answered by comparing across experiments, and to the baseline group. Additionally, Experiments 1 and 3 both used the same recognition test, so they could be compared on recognition as well.

How does Exp. 1 compare to Baseline?

Experiment 1 replicated a well-known visual statistical learning (SL) paradigm (Fiser & Aslin, 2001), and added a psychophysical assessment to test for perceptual learning (PL) effects following the SL familiarization. Having found improved accuracy, decreased false alarming, and increased sensitivity in the Familiar condition as well as in the near-transfer Shuffled condition, it was important to compare these findings to a baseline group that only participated in the assessment. The comparison of Experiment 1 assessment results to the baseline group can be found in Chapter 3, on page 40.

How does Exp. 3 at 21 minutes compare to Exp. 1?

Experiment 3 directly replicated Experiment 1 and extended it by varying the length of the familiarization. Thus, it was important to compare the results of the replicated 21 minute session length in Experiment 3 to the results of Experiment 1, on both recognition and the psychophysical assessment.

Recognition Accuracy. An independent-samples *t*-test comparing Experiment 1 to Experiment 3 at the 21 minutes session length revealed no difference in recognition accuracy (*p* = .45).

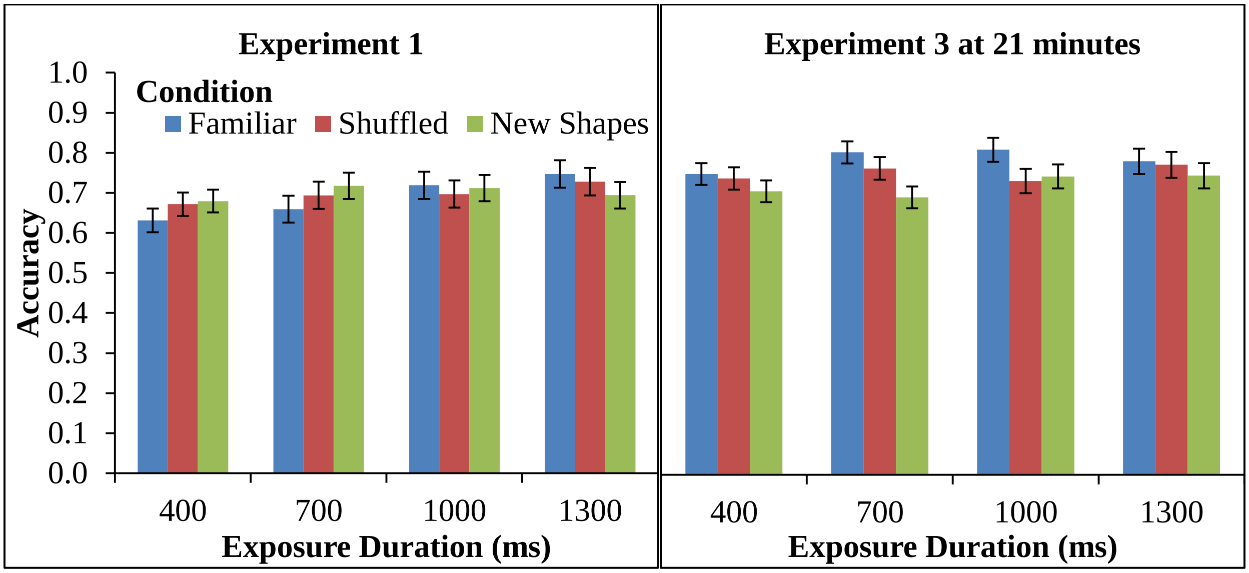

Figure 32. Experiment Version by Condition by Exposure Duration on accuracy (collapsed across Target Presence). Error bars indicate standard error of the mean.

Psychophysical Assessment Accuracy. Figure 32 showed the effects of Experiment Version, Condition, and Exposure Duration on accuracy. It appeared that results were similar across experiments, except that Shuffled tended to have higher accuracy in Experiment 1 and Familiar tended to have higher accuracy in Experiment 3 at 21 minutes. To test these apparent effects, I conducted an ANOVA of Experiment Version by Condition by Exposure Duration by Target Presence on accuracy. This analysis revealed two marginal interactions involving Experiment Version: a marginal interaction of Experiment Version with Condition and Target Presence, *F*(2,152) = 2.43, *p* = .092, *partial-eta-squared* = 0.03; and a marginal interaction of

Experiment Version with Exposure Duration and Target Presence, *F*(3,456) = 2.54, *p* = .06, *partial-eta-squared* = 0.02. There was no main effect of Experiment Version (*p* = .18) or interaction of Experiment Version and Condition (*p* = .24). There was a main effect of Condition, *F*(2,152) = 5.56, *p* = .005, *partial-eta-squared* = 0.07 and a main effect of Exposure Duration, *F*(3,456) = 4.04, *p* = .03, *partial-eta-squared* = 0.03. There were three additional interactions: a marginal interaction of Condition by Exposure Duration by Target Presence, *F*(6,456) = 1.83, *p* = .09, *partial-eta-squared* = 0.02; an interaction of Condition by Target Presence, *F*(2,152) = 3.31, *p* = .04, *partial-eta-squared* = 0.04; and an interaction of Exposure Duration by Target Presence, *F*(3,456) = 6.02, *p* < .001, *partial-eta-squared* = 0.04. No other effects were significant (all *p*'s > .16).

For the marginal interaction of Experiment Version with Condition and Target Presence, the data were split on Target Presence and then simple interactions of Experiment Version and Condition were tested. For significant simple interactions, simple simple effects of Experiment for each condition were tested via custom hypothesis tests in ANOVA. There was no simple interaction of Experiment Version and Condition for Absent (*p* = .96). For Present, there was a significant simple interaction of Experiment Version and Condition, *F*(2,152) = 3.82, *p* = .02, *partial-eta-squared* = 0.05. There was a significant simple simple effect of Experiment Version for Shuffled, such that Experiment 1 (*M* = 0.78, *SE* = 0.03) showed higher accuracy than Experiment 3 (*M* = 0.69, *SE* = 0.03), *F*(1,152) = 4.30, *p* = .04, *partial-eta-squared* = 0.03. There was no simple simple effect of Condition for Familiar (*p* = .28) or New Shapes (*p* = .14). The marginal three-way interaction of Experiment Version, Condition, and Target Presence was driven by higher accuracy in Experiment 1 than Experiment 3 for Shuffled when the target was Present, but otherwise the experiments did not differ.

For the marginal interaction of Experiment Version with Exposure Duration and Target Presence, I divided the data on Target Presence to test simple interactions for Experiment Version and Exposure Duration, but found no reliable simple interactions (both *p*'s > .27). Then I divided the data on Exposure Duration to examine simple interactions of Experiment Version and Target Presence. For 1000ms, there was a marginal simple interaction of Experiment Version and Target Presence, *F*(1,151) = 3.80, *p* = .05, *partial-eta-squared* = 0.03. Custom hypothesis tests in ANOVA of simple simple effects of Experiment Version at 1000ms revealed a marginal simple simple effect of Experiment Version when the target was Absent, such that participants in Experiment 1 (*M* = 0.74, *SE* = 0.02) showed lower accuracy than those in Experiment 3 (*M* = 0.80, *SE* = 0.03), *F*(1,152) = 3.78, *p* = .05, *partial-eta-squared* = 0.02. There was no simple simple effect when the target was Present at 1000ms (*p* = .48). There were no simple interactions of Experiment Version and Target Presence for 400ms (*p* = .50), 700ms (*p* = .26), or 1300ms (*p* = .31). The marginal three-way interaction of Experiment Version, Exposure Duration, and Target Presence was driven by lower accuracy in Experiment 1 than Experiment 3 when the target was Absent at the 1000ms exposure duration, but otherwise the experiments did not differ.

Custom hypothesis tests in ANOVA were used to investigate the main effects. For Condition, I compared Familiar (*M* = 0.77, *SE* = 0.02) and Shuffled (*M* = 0.77, *SE* = 0.02), which did not differ in accuracy (*p* = .91), but when I compared them to New Shapes (*M* = 0.71, *SE* = 0.02), Familiar and Shuffled combined showed higher accuracy than New Shapes, *F*(1,152) = 11.11, *p* = .001, *partial-eta-squared* = 0.07. Learning transferred from Familiar to Shuffled in both experiments.

For Exposure Duration, I compared 400ms (*M* = 0.73, *SE* = 0.01) and 700ms (*M* = 0.75, *SE* = 0.01), which did not differ in accuracy (*p* = .10). Then I compared 1000ms (*M* = 0.76, *SE* = 0.01) and 1300ms (*M* = 0.76, *SE* = 0.01), which also did not differ (*p* = .77). Then I compared 1000ms and 1300ms to 400ms and 700ms, and found that the longer exposure durations together showed marginally higher accuracy than the shorter exposure durations combined, *F*(1,152) = 8.35, *p* = .004, *partial-eta-squared* = 0.05.

For the marginal interaction of Condition with Exposure Duration and Target Presence, the data were split on Target Presence to investigate simple interactions of Condition and Exposure Duration. There was a simple interaction of Condition and Exposure Duration when the target was Present, *F*(6,456) = 2.42, *p* = .03, *partial-eta-squared* = 0.03. I followed up this simple interaction with a custom hypothesis tests of simple simple effects of Condition at each exposure duration, but found no reliable simple simple effects (all *p*'s > .13). Then I tested simple simple effects of Exposure Duration at each condition. There was a simple simple effect of Exposure Duration for New Shapes, *F*(3,150) = 9.68, *p* < .001, *partial-eta-squared* = 0.16. For Present and New Shapes, 1000ms (*M* = 0.76, *SE* = 0.03) and 1300ms (*M* = 0.78, *SE* = 0.03) did not differ in accuracy (*p* = .39). 400ms (*M* = 0.65, *SE* = 0.03) showed lower accuracy than the two longest exposure durations combined, *F*(1,50) = 26.09, *p* < .001, *partial-eta-squared* = 0.34. 700ms (*M* = 0.70, *SE* = 0.03) showed marginally lower accuracy than the longest exposure durations, *F*(1,50) = 8.83, *p* = .005, *partial-eta-squared* = 0.15. 400ms and 700ms did not differ (*p* = .08). There were no simple simple effects of Exposure Duration for Familiar (*p* = .14) or Shuffled (*p* = .03). There was no simple interaction for Absent (*p* = .73).

Custom hypothesis tests (in the original ANOVA) were used to examine the interaction of Condition and Target Presence via tests of simple effects of Condition at each level of Target

Presence, and significant simple effects were followed up with additional custom hypothesis tests. There was a simple effect of Condition when the target was Absent, *F*(2,152) = 8.16, *p* < .001, *partial-eta-squared* = 0.10. For Absent, I compared Familiar (*M* = 0.79, *SE* = 0.02) and Shuffled (*M* = 0.80, *SE* = 0.02), which did not differ (*p* = .57). Then I compared them to New Shapes (*M* = 0.69, *SE* = 0.02), and found that Familiar and Shuffled together they showed higher accuracy than New Shapes, *F*(1,152) = 15.96, *p* < .001, *partial-eta-squared* = 0.10. There was no simple effect of Condition when the target was Present (*p* = .50). The interaction of Condition and Target Presence was driven by the Condition effect for Absent (and there was no effect for Present).

For the interaction of Exposure Duration and Target Presence, the data were split on Target Presence to examine simple effects of Exposure Duration. There was a simple effect of Exposure Duration when the target was Present, *F*(3,456) = 9.86, *p* < .001, *partial-eta-squared* = 0.06. I conducted all pairwise comparisons to follow up on the significant simple effect. 400ms (*M* = 0.69, *SE* = 0.02) was less accurate than 1000ms (*M* = 0.75, *SE* = 0.02), *F*(1,152) = 14.02, *p* < .001, *partial-eta-squared* = 0.08. 400ms was also less accurate than 1300ms (*M* = 0.77, *SE* = 0.02), *F*(1,152) = 26.42, *p* < .001, *partial-eta-squared* = 0.15. 700ms (*M* = 0.73, *SE* = 0.0s) was marginally less accurate than 1300ms, *F*(1,152) = 7.53, *p* = .007, *partial-eta-squared* = 0.05. No other pairwise comparisons for Present were significant (all *p*'s > .01). There was no simple effect of Exposure Duration for Absent (*p* = .73). The interaction of Exposure Duration and Target Presence was driven by lower accuracy for shorter than longer exposure durations when the target was Present and no effect of Exposure Duration when the target was Absent.

False Alarm Rate. Figure 33 showed the effects of Experiment Version, Condition, and Exposure Duration on false alarm rate. It appeared that effects were similar across experiments, but perhaps stronger in Experiment 3. To test the apparent effects, I conducted an ANOVA of Experiment Version by Condition by Exposure Duration on false alarm rate, which revealed a main effect of Condition, *F*(2,152) = 8.16, *p* < .001, *partial-eta-squared* = 0.10, and a marginal main effect of Experiment Version, such that participants in Experiment 1 (*M* = 0.26, *SE* = 0.02) false alarmed more than participants in Experiment 3 (*M* = 0.22, *SE* = 0.02), *F*(1,152) = 2.80, *p* = .097, *partial-eta-squared* =0.02. There was no interaction of Experiment Version and Condition (*p* = .96), and no other effects (all *p*'s > .31). I used custom hypothesis tests in ANOVA for Condition. I first compared Familiar (*M* = 0.21, *SE* = 0.02) and Shuffled (*M* = 0.20, *SE* = 0.02), which did not differ in false alarm rate (*p* = .57). Then I compared them to New Shapes (*M* = 0.31, *SE* = 0.02), and found that Familiar and Shuffled showed significantly fewer false alarms than New Shapes, *F*(1,152) = 15.96, *p* < .001, *partial-eta-squared* = 0.10. The learning transferred from Familiar to Shuffled across both experiments.

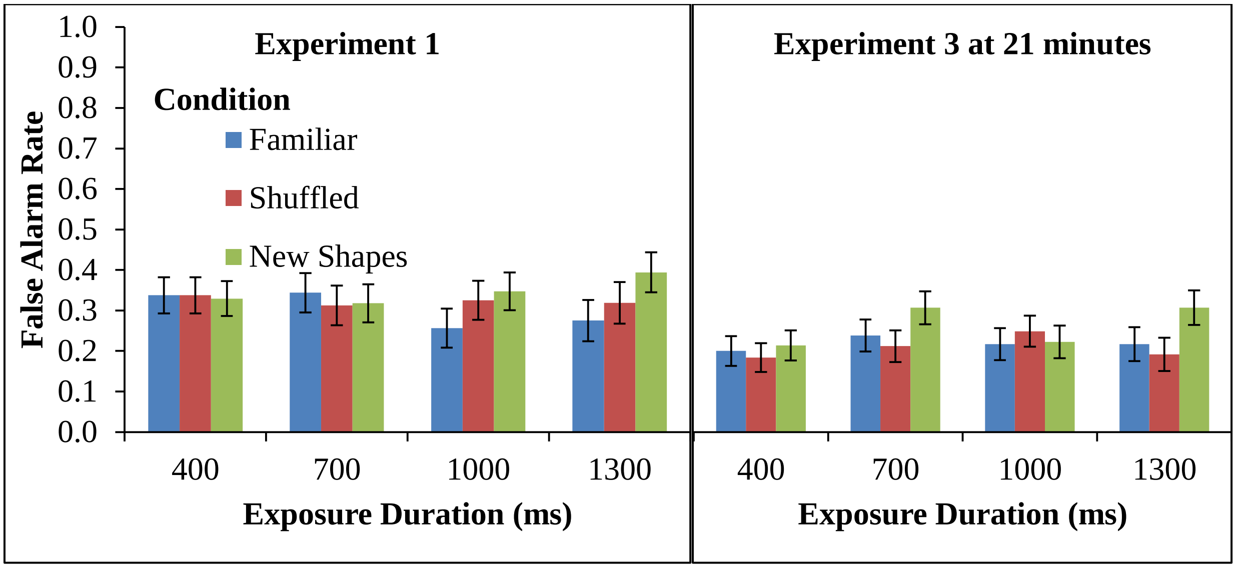

Figure 33. Experiment Version by Condition by Exposure Duration on false alarms. Error bars indicate standard error of the mean.

Sensitivity. Figure 34 showed effects of Experiment Version, Condition, and Exposure Duration on sensitivity (d'). It appeared that effects were overall similar across experiments, except that sensitivity was highest for Shuffled in Experiment 1 but was highest for Familiar in Experiment 3 at 21 minutes. To test these apparent effects, I conducted an ANOVA of Experiment Version by Condition by Exposure Duration on sensitivity, which revealed a main effect of Condition, *F*(2,151) = 5.88, *p* = .003, *partial-eta-squared* = 0.07, and a main effect of Exposure Duration, *F*(3,456) = 4.32, *p* = .004, *partial-eta-squared* = 0.03. There was no main effect of Experiment Version (*p* = .11) or interaction of Experiment Version and Condition (*p* = .31), so the apparent effect of different conditions showing higher accuracy in each experiment was not supported. No other effects were significant (all *p*'s > .27).

Custom hypothesis tests in ANOVA were used to investigate the main effects. For Condition, I compared Familiar (*M* = 1.55, *SE* = 0.10) and Shuffled (*M* = 1.53, *SE* = 0.10), which did not show different sensitivity (p = .89). When I compared them to New Shapes (*M* = 1.12, *SE* = 0.10), Familiar and Shuffled showed higher sensitivity than New Shapes, *F*(1,152) = 11.75, *p* = .001, *partial-eta-squared* = 0.07. The learning transferred from Familiar to Shuffled across both experiments.

For Exposure Duration, I conducted all pairwise comparisons. 400ms (*M* = 1.27, *SE* = 0.07) showed less sensitivity than 1300ms (*M* = 1.49, *SE* = 0.08), *F*(1,152) = 11.19, *p* = .001, *partial-eta-squared* = 0.07. 400ms showed marginally less sensitivity than 1000ms (*M* = 1.46, *SE* = 0.07), *F*(1,152) = 8.76, *p* = .004, *partial-eta-squared* = 0.06. No other pairwise comparisons were significant (all *p*'s > .07).

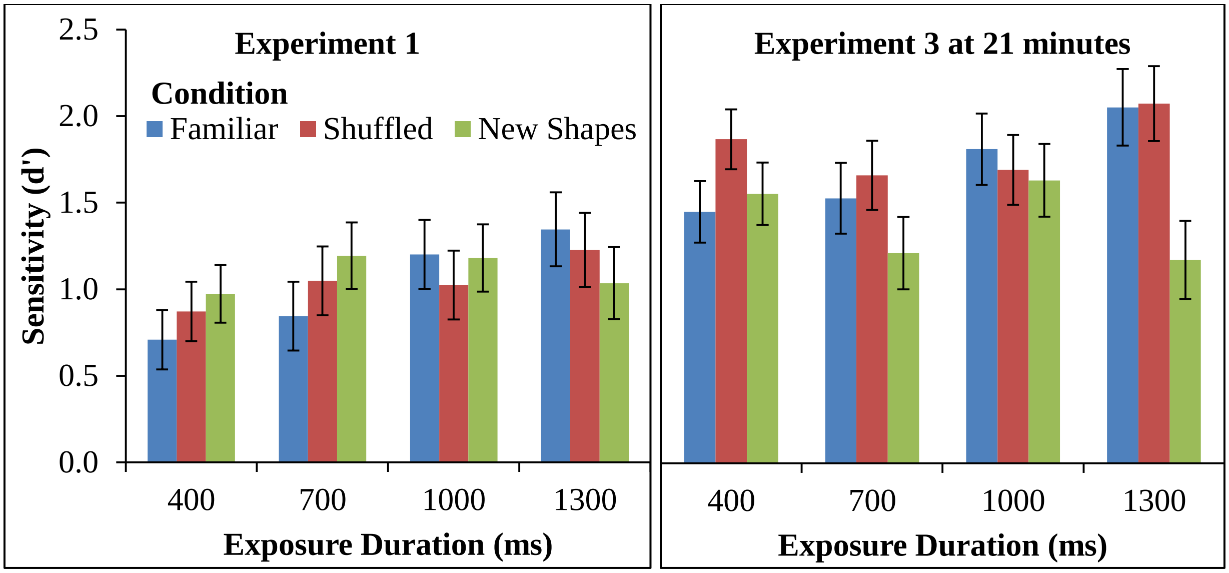

Figure 34. Experiment Version by Condition by Exposure Duration on sensitivity. Error bars indicate standard error of the mean.

How does Exp. 3 at 7 minutes compare to Baseline?

Experiment 3 replicated Experiment 1 and extended it by varying the length of the familiarization. The 7 minutes of familiarization in Experiment 3 were analyzed with the Baseline group to determine if there was any learning relative to Baseline with this brief familiarization session length.

Accuracy Figure 35 showed effects of Experiment Version, Condition, and Exposure Duration on accuracy. It appeared that Experiment 3 at 7 minutes showed higher accuracy than Baseline, but also that Experiment 3 only showed a strong difference of conditions for the longest exposure duration. I tested these apparent effects via an ANOVA of Experiment Version by Condition by Exposure Duration by Target Presence on accuracy, which revealed a main effect of Experiment Version, such that accuracy was higher in Experiment 3 with a 7 minute session length (*M* = 0.74, *SE* = 0.01) than at Baseline (*M* = 0.70, *SE* = 0.02), *F*(1,112) = 5.14, *p* =

.03, *partial-eta-squared* = 0.04. It also revealed a main effect of Exposure Duration, *F*(3, 336) = 4.05, *p* = .008, *partial-eta-squared* = 0.04 and three interactions: Experiment Version and Target Presence, *F*(1,112) = 5.20, *p* = .03, *partial-eta-squared* = 0.04; Condition and Exposure Duration, *F*(6,336) = 3.43, *p* = .003, *partial-eta-squared* = 0.06; and Exposure Duration and Target Presence, *F*(3,336) = 5.17, *p* = .002, *partial-eta-squared* = 0.04. No other effects were significant (all *p*'s > .12).

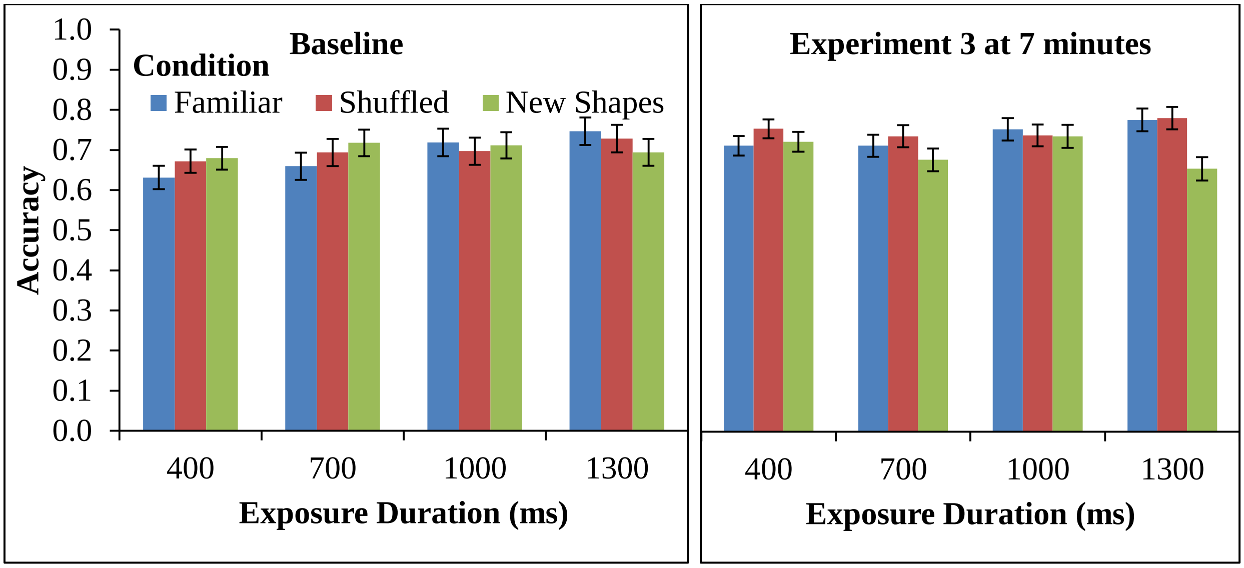

Figure 35. Experiment Version by Condition by Exposure Duration on accuracy (collapsed across Target Presence). Error bars indicate standard error of the mean.

The main effect of Exposure Duration was investigated via custom hypothesis tests in ANOVA. I compared 400ms (*M* = 0.70, *SE* = 0.01) and 700ms (*M* = 0.70, *SE* = 0.01), which did not differ in false alarm rate (*p* = .74). Similarly, I compared 1000ms (*M* = 0.73, *SE* = 0.01) and 1300ms (*M* = 0.74, *SE* = 0.01), which also did not differ (*p* = .73). However, when I compared the shorter exposure durations to the longer ones, 400ms and 700ms showed significantly lower accuracy than 1000ms and 13000ms, *F*(1,112) = 11.40, *p* = .001, *partial-eta-squared* = 0.09.

For the interaction of Experiment Version with Target Presence and the interaction of Condition with Exposure Duration, custom hypothesis tests in ANOVA were used to explore these effects. For Experiment Version and Target Presence, simple effects of Experiment Version were examined. There was a significant simple effect of Experiment Version when the target was Absent, such that participants were more accurate in Experiment 3 after only a 7 minute session length (*M* = 0.77, *SE* = 0.02) than at Baseline (*M* = 0.68, *SE* = 0.02), *F*(1,112) = 10.72, *p* = .001, *partial-eta-squared* = 0.09. There was no simple effect of Experiment Version for Present (*p* = .91). The interaction of Experiment Version and Target Presence was driven by higher Experiment 3 at 7 minutes showing higher accuracy than Baseline only for target-absent trials.

For Condition with Exposure Duration, simple effects of Condition for each exposure duration were tested. There was a simple effect of Condition at 1300ms, *F*(2,112) = 4.83, *p* = .01, *partial-eta-squared* = 0.08. At 1300ms, I compared Familiar (*M* = 0.77, *SE* = 0.02) and Shuffled (*M* = 0.76, *SE* = 0.02), which did not differ in accuracy (*p* = .82), but when I compared them to New Shapes (*M* = 0.68, *SE* = 0.02), Familiar and Shuffled together showed higher accuracy than New Shapes, *F*(1,112) = 9.62, *p* = .002, *partial-eta-squared* = 0.08. There was no simple effect of Condition for 400ms (*p* = .29), 700ms (*p* = .64), or 1000ms (*p* = .83). Learning was observed only at the longest exposure duration.

For the interaction of Exposure Duration and Target Presence, the data were split on Target Presence and simple effects of Exposure Duration were evaluated. There was a simple effect of Exposure Duration for Present, *F*(3,336) = 9.20, *p* < .001, *partial-eta-squared* = 0.08. I followed up with all pairwise comparisons. 400ms (*M* = 0.67, *SE* = 0.02) showed lower accuracy than 1300ms (*M* = 0.76, *SE* = 0.02), *F*(1,112) = 23.92, *p* < .001, *partial-eta-squared* = 0.18. 700ms (*M* = 0.70, *SE* = 0.02) also showed lower accuracy than 1300, *F*(1,112) = 11.70, *p* = .001, *partial-eta-squared* = 0.10. 400ms showed marginally lower accuracy than 1000ms (*M* = 0.73, *SE* = 0.02), *F*(1,112) = 10.35, *p* = .002, *partial-eta-squared* = 0.09. No other pairwise comparisons of exposure durations for Present were significant (all *p*'s > .06). There was no simple effect of Exposure Duration for Absent (*p* = .59). The interaction of Exposure Duration and Target Presence was driven by lower accuracy for shorter exposure durations and higher accuracy for longer exposure durations only when the target was Present.

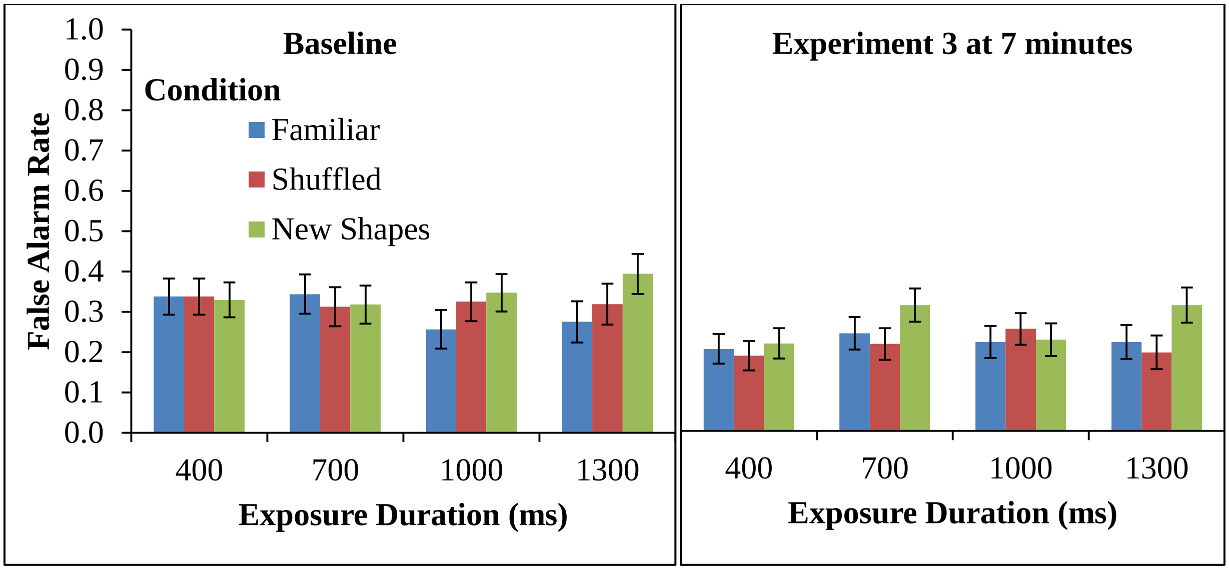

Figure 36. Experiment Version by Condition by Exposure Duration on false alarm rate. Error bars indicate standard error of the mean.

False Alarm Rate. Figure 36 showed effects of Experiment Version, Condition, and Exposure Duration on false alarm rate. It appeared that Experiment 3 at 7 minutes had lower false alarm rates across conditions and exposure durations compared to Baseline, but especially for Familiar and Shuffled at 700ms and 1300ms. To test these apparent effects, I conducted an ANOVA of Experiment Version and Condition and Exposure Duration on false alarm rate, which showed only a main effect of Experiment Version, such that Experiment 3 at 7 minutes of familiarization (*M* = 0.24, *SE* = 0.02) showed a lower false alarm rate than Baseline (*M* = 0.33, *SE* = 0.03), *F*(1,112) = 10.72, *p* = .001, *partial-eta-squared* = 0.09. No other effects were significant (all *p*'s > .15), so the apparent three-way interaction of Experiment Version, Condition, and Exposure Duration was not significant.

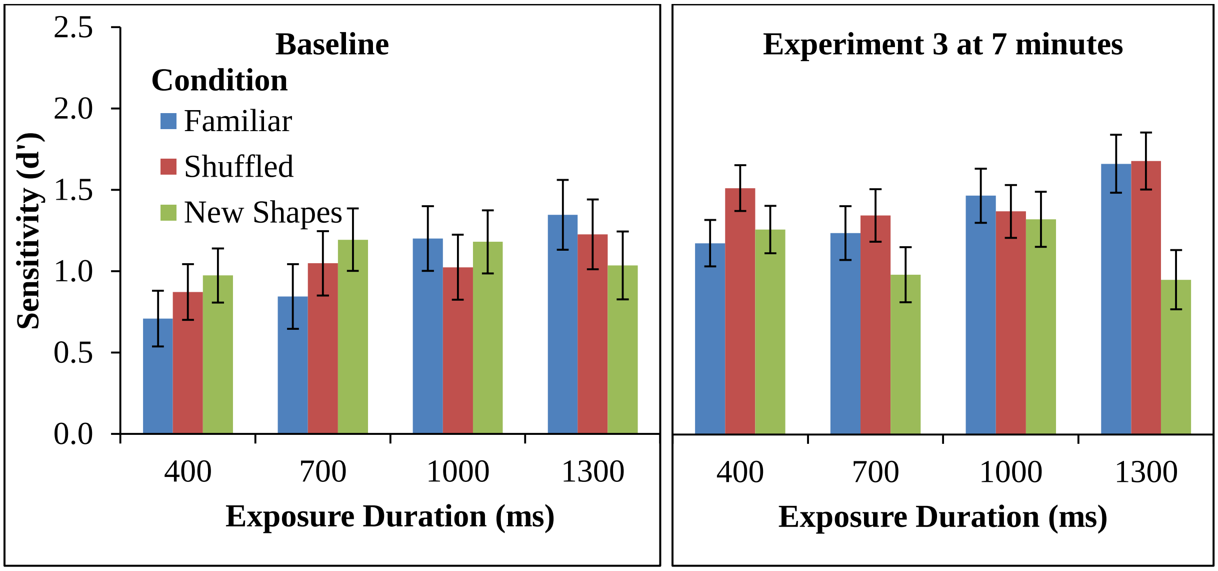

Figure 37. Experiment Version by Condition by Exposure Duration on sensitivity. Error bars indicate standard error of the mean.

Sensitivity. Figure 37 showed effects of Experiment Version, Condition, and Exposure Duration on sensitivity (d'). It appeared that Experiment 3 at 7 minutes showed higher sensitivity than Baseline across conditions and exposure durations, but especially for Shuffled at 400ms and

700ms and both Familiar and Shuffled at 1300ms. To test these apparent effects, I conducted an ANOVA of Experiment Version with Condition and Exposure Duration on sensitivity. There was a main effect of Experiment Version, such that Experiment 3 with a 7 minute session length (*M* = 1.33, *SE* = 0.07) showed higher sensitivity than Baseline (*M* = 1.06, *SE* = 0.09), *F*(1,112) = 6.21, *p* = .01, *partial-eta-squared* = 0.05. There was also a main effect of Exposure Duration, *F*(3,336) = 4.22, *p* = .006, *partial-eta-squared* = 0.04, and an interaction of Condition and Exposure Duration, *F*(6,336) = 3.04, *p* = .007, *partial-eta-squared* = 0.05. No other effects were significant (all *p*'s > .24), so the apparent three-way interaction of Experiment Version, Condition, and Exposure Duration was not supported by the data.

The main effect of Exposure Duration was studied via custom hypothesis tests in ANOVA. I compared 400ms (*M* = 1.08, *SE* = 0.06) and 700ms (*M* = 1.11, *SE* = 0.07), which did not differ in sensitivity (*p* = .74). Similarly, I compared 1000ms (*M* = 1.26, *SE* = 0.08) and 1300ms (*M* = 1.32, *SE* = 0.08), which also did not differ in sensitivity (*p* = .50). When I compared the shorter exposure durations to the longer exposure durations, 400ms and 700ms showed less sensitivity than 1000ms and 1300ms, *F*(1,112) = 11.74, *p* = .001, *partial-eta-squared* = 0.10.

Custom hypothesis tests in ANOVA were used to investigate the interaction of Condition and Exposure Duration. I tested simple effects of Condition, and followed up significant simple effects with additional tests. There was a simple effect of Condition at 1300ms, *F*(1,112) = 4.14, *p* = .02, *partial-eta-squared* = 0.07. I compared Familiar (*M* = 1.51, *SE* = 0.14) and Shuffled (*M* = 1.45, *SE* = 0.14), which did not differ in sensitivity (*p* = .80). When I compared Familiar and Shuffled to New Shapes (*M* = 0.99, *SE* = 0.14), Familia and Shuffled showed higher sensitivity than New Shapes, *F*(1,112) = 8.22, *p* = .005, *partial-eta-squared* = 0.07. There was no simple effect of Condition for 400ms (*p* = .27), 700ms (*p* = .68), or 1000ms (*p* = .75). The pattern of Conditions only held for the longest exposure duration.

How does Exp. 3 at 35 minutes compare to Exp. 2?

Experiment 3 replicated Experiment 1 and extended it by varying the length of the familiarization. Extending the familiarization to 35 minutes could have made results of this condition in Experiment 3 more similar to Experiment 2 results, so this was tested.

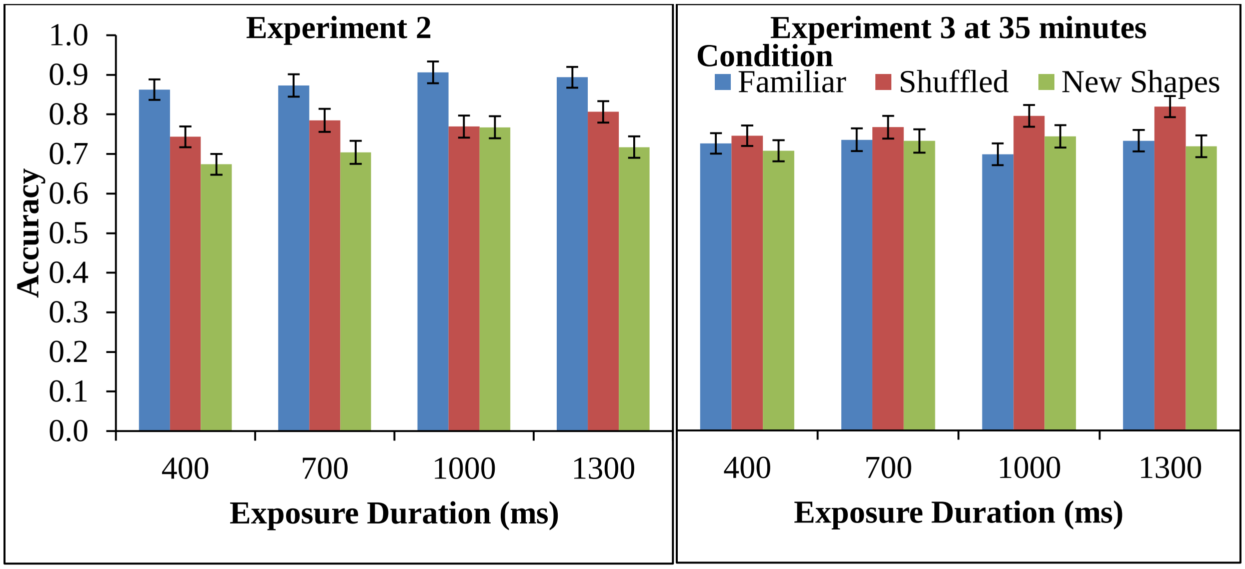

Figure 38. Experiment Version by Condition by Exposure Duration on accuracy (collapsed across Target Presence). Error bars indicate standard error of the mean.

Accuracy. Figure 38 showed effects of Experiment Version, Condition, and Exposure Duration on accuracy. It appeared that Experiment 2 showed higher accuracy across durations and across conditions, but especially for the Familiar condition. To test these apparent effects, I conducted an ANOVA of Experiment Version by Condition by Exposure Duration by Target

Presence on accuracy. There was a main effect of Experiment Version, such that Experiment 2 (*M* = 0.79, *SE* = 0.01) showed higher accuracy than Experiment 3 at 35 minutes (*M* = 0.75, *SE* = 0.01), *F*(1,132) = 5.47, *p* = .02, *partial-eta-squared* = 0.04. There was also a main effect of Target Presence, such that participants were less accurate when the target was Present (*M* = 0.75, *SE* = 0.01) than when it was Absent (*M* = 0.79, *SE* = 0.01), *F*(1,132) = 6.50, *p* = .01, *partial-eta-squared* = 0.05. There were also main effects of Condition, *F*(2,132) = 7.45, *p* = .001, *partial-eta-squared* = 0.10, and Exposure Duration, *F*(3,396) = 5.13, *p* = .002, *partial-eta-squared* = 0.04. There were two interactions: Experiment Version by Condition, *F*(2,132) = 9.93, *p* < .001, *partial-eta-squared* = 0.13, and Experiment Version by Target Presence, *F*(2,132) = 4.14, *p* = .04, *partial-eta-squared* = 0.03. No other effects were significant (all *p*'s > .13).

Custom hypothesis tests in ANOVA were used to investigate significant effects. For Condition, I first compared Familiar (*M* = 0.81, *SE* = 0.02) and Shuffled (*M* = 0.78, *SE* = 0.02), which did not differ (*p* = .27). Then I compared Familiar and Shuffled combined to New Shapes (*M* = 0.72, *SE* = 0.02), and found that the combined conditions showed higher accuracy than New Shapes, *F*(1,132) = 13.63, *p* < .001, *partial-eta-squared* = 0.09. Learning transferred from Familiar to Shuffled.

For Exposure Duration, I conducted all pairwise comparisons, and found that 400ms (*M* = 0.75, *SE* = 0.01) was less accurate than 1000ms (*M* = 0.78, *SE* = 0.01), *F*(1,132) = 12.14, *p* = .001, *partial-eta-squared* = 0.08. 400ms was also less accurate than 1300ms (*M* = 0.79, *SE* = 0.01), *F*(1,132) = 15.66, *p* < .001, *partial-eta-squared* = 0.11. No other pairwise comparisons were significant (all *p*'s > .05).

For the interaction of Experiment Version and Condition, simple effects of Experiment Version were examined at each condition via custom hypothesis tests in ANOVA. For Familiar, there was a simple effect of Experiment Version, such that Experiment 2 (*M* = 0.88, *SE* = 0.02) showed higher accuracy than Experiment 3 at 35 minutes (*M* = 0.73, *SE* = 0.02), *F*(1,132) = 25.23, *p* < .001, *partial-eta-squared* = 0.16. There was no simple effect of Experiment Version for Shuffled (*p* = .69) or New Shapes (*p* = .61). The interaction of Condition and Experiment Version was driven by simple effect of Experiment Version for Familiar.

For the interaction of Experiment Version and Target Presence, simple effects of Experiment version were examined at each level of Target Presence using custom hypothesis tests in ANOVA. There was a significant simple effect of Experiment Version for Present, such that Experiment 2 (*M* = 0.79, *SE* = 0.02) showed higher accuracy than Experiment 3 at 35 minutes (*M* = 0.71, *SE* = 0.02), *F*(1,132) = 9.91, *p* = .002, *partial-eta-squared* = 0.07. There was no simple effect for Absent (*p* = .13). The interaction of Experiment Version and Target Presence was driven by the simple effect of Experiment Version for Present.

False Alarms Rate. Figure 39 showed effects of Experiment Version, Condition, and Exposure Duration on false alarm rate. It appeared that overall false alarm rates were similar (though perhaps lower in Experiment 2), but Familiar had the lowest false alarm rate in Experiment 2 whereas Shuffled had the lowest false alarm rate in Experiment 3 at 35 minutes. To test these apparent effects, I conducted an ANOVA of Experiment Version by Condition by Exposure Duration on false alarm rate. There was a main effect of Condition, *F*(2,132) = 3.34, *p* = .04, *partial-eta-squared* = 0.05, and an interaction of Experiment Version with Condition, *F*(2,132) = 4.13, *p* = .02, *partial-eta-squared* = 0.06. No other effects were significant (all *p*'s > .19).

Custom hypothesis tests in ANOVA were used to evaluate significant effects. For the main effect of Condition, I compared Familiar (*M* = 0.18, *SE* = 0.02) and Shuffled (*M* = 0.19, *SE*

= 0.02), which did not differ in false alarm rate (*p* = .91). When I compared Familiar and Shuffled together to New Shapes (*M* = 0.25, *SE* = 0.02), I found that Familiar and Shuffled showed a lower false alarm rate than New Shapes, *F*(1,132) = 6.667, *p* = .01, *partial-eta-squared* = 0.05. Learning transferred from Familiar to Shuffled in terms of decreased false alarming.

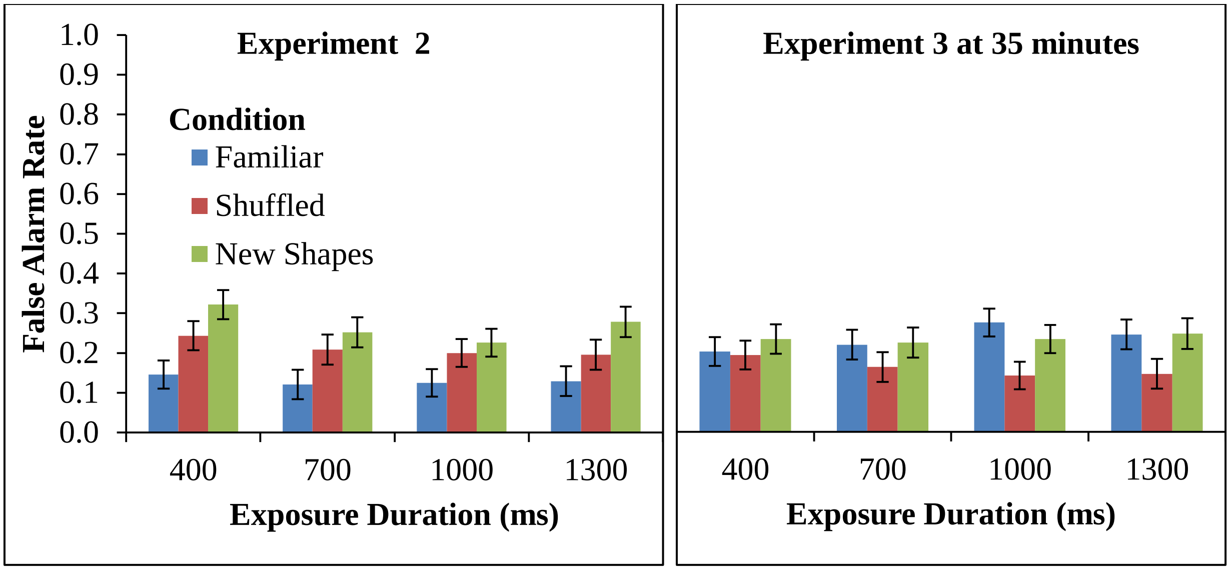

Figure 39. Experiment Version by Condition by Exposure Duration on false alarm rate. Error bars indicate standard error of the mean.

For the interaction of Experiment Version with Condition, simple effects of Experiment Version were tested for each level of Condition. For Familiar, there was a simple effect of Experiment Version, such that Experiment 2 (*M* = 0.13, *SE* = 0.03) showed a lower false alarm rate than Experiment 3 with 35 minutes of familiarization (*M* = 0.24, *SE* = 0.03), *F*(1,132) = 6.55, *p* = .01, *partial-eta-squared* = 0.05. There was no simple effect of Experiment Version for Shuffled (*p* = .25) or New Shapes (*p* = .46). The interaction of Experiment Version and Condition was driven by the simple effect of Experiment Version for Familiar.

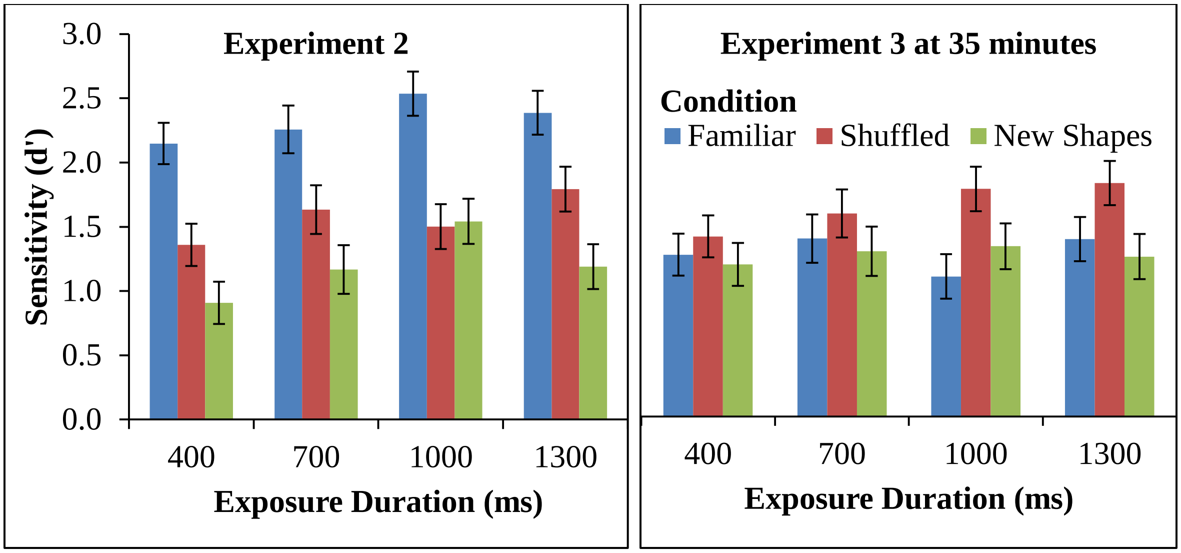

Figure 40. Experiment Version by Condition by Exposure Duration on sensitivity. Error bars indicate standard error of the mean.

Sensitivity. Figure 40 showed effects of Experiment Version, Condition, and Exposure Duration on sensitivity. It appeared that sensitivity was higher overall for Experiment 2, especially for the Familiar condition. To test these apparent effects, I conducted an ANOVA of Experiment Version by Condition by Exposure Duration on sensitivity (d'), which revealed a main effect of Experiment Version, such that Experiment 2 (*M* = 1.70, *SE* = 0.08) showed higher sensitivity than Experiment 3 (*M* = 1.41, *SE* = 0.08), *F*(1,132) = 6.60, *p* = .01, *partial-eta-squared* = 0.05. It also revealed main effects of Condition, *F*(2,132) = 8.57, *p* < .001, *partial-eta-squared* = 0.12, and Exposure Duration, *F*(3,396) = 5.84, *p* = .001, *partial-eta-squared* = 0.04, and an interaction of Experiment Version and Condition, *F*(2,132) = 10.63, *p* < .001, *partial-eta-squared* = 0.14. No other effects were significant (all *p*'s > .14).

All significant effects were investigated via custom hypothesis tests in ANOVA. For the main effect of Condition, I first compared Familiar (*M* = 1.81, *SE* = 0.10) and Shuffled (*M* =

1.62, *SE* = 0.10), which did not differ in sensitivity (*p* = .16). Then I compared Familiar and Shuffled combined to New Shapes (*M* = 1.24, *SE* = 0.10), and found that Familiar and Shuffled combined showed higher sensitivity than New Shapes, *F*(1,132) = 15.10, *p* < .001, *partial-eta-squared* = 0.10. Learning, as measured by increased sensitivity, transferred from Familiar to Shuffled.

For Exposure Duration, I compared 1000ms (*M* = 1.63, *SE* = 0.07) and 1300ms (*M* = 1.64, *SE* = 0.07), which did not differ in sensitivity (*p* = .91). When I compared 1000ms and 1300ms together to 400ms (*M* = 1.38, *SE* = 0.07), I found that the longer durations showed higher sensitivity than 400ms, *F*(1,132) = 24.17, *p* < .001, *partial-eta-squared* = 0.16. 700ms (*M* = 1.59, *SE* = 0.08) did not differ in sensitivity from 400ms (*p* = .02) or from the longer exposure durations combined (*p* = .21).

For the interaction of Experiment Version and Condition, simple effects of Experiment Version for each condition were tested. There was a simple effect of Experiment Version for Familiar, such that Experiment 2 (*M* = 2.33, *SE* = 0.14) showed higher sensitivity than Experiment 3 with a 35 minute session length (*M* = 1.29, *SE* = 0.14), *F*(1,132) = 27.92, *p* < .001, *partial-eta-squared* = 0.18. There was no simple effect of Experiment Version for Shuffled (*p* = .67) or New Shapes (*p* = .73). The interaction of Experiment Version and Condition was driven by the simple effect of Experiment Version for Familiar.

How does learning time relate to performance across experiments?

In Experiment 3, participants with a longer session length performed better in recognition accuracy. There was no main effect of Session Length for psychophysical assessment accuracy, but Session Length was involved in significant interactions. In Experiment 2, PL training time did not correlate with assessment performance. It would be interesting to investigate if amount of learning time (session length for Experiments 1 and 3, PL training time for Experiment 2) mattered across all three experiments. A Pearson correlation revealed no association between learning time and psychophysical assessment accuracy across experiments (*p* = .54).

How do all of the experiments compare to each other and Baseline?

Experiment 2 employed a PL intervention based on the chosen SL paradigm used in Experiments 1 and 3. The PL intervention caused strong PL effects. These were similar to the effects observed in Experiments 1 and 3, but stronger for Familiar than Shuffled, unlike Experiments 1 and 3. To test the degree to which transfer occurred in Experiment 2 and to compare transfer across all of the experiments against each other and Baseline, I analyzed Experiments 1, 2, and 3 (collapsed across Session Length) together with Baseline.

Accuracy. Figure 41 showed effects of Experiment Version, Condition, and Exposure Duration on accuracy, statistically controlling for Recognition, Noticing, and Linguistic Coding, and it appeared that Baseline had the lowest accuracy across exposure durations and Experiment 2 had the highest, particularly for Familiar. Shuffled appeared to have the highest accuracy for Experiments 1 and 3. To test these apparent effects, I conducted an ANOVA of Experiment Version by Condition by Exposure Duration by Target Presence on accuracy, which revealed main effects of Experiment Version, *F*(3,402) = 8.13, *p* < .001, *partial-eta-squared* = 0.06; Condition, *F*(2,402) = 10.38, *p* < .001, *partial-eta-squared* = 0.06; and Exposure Duration, *F*(3,1206) = 13.06, *p* < .001, *partial-eta-squared* = 0.03. It also revealed interaction effects of Experiment Version and Condition, *F*(6,402) = 3.98, *p* = .001, *partial-eta-squared* = 0.06; Experiment Version and Target Presence, *F*(3,402) = 3.57, *p* = .02, *partial-eta-squared* = 0.03; Condition and Exposure Duration, *F*(6,1206) = 2.37, *p* = .03, *partial-eta-squared* = 0.01; and

Exposure Duration and Target Presence, *F*(3,1206) = 11.00, *p* < .001, *partial-eta-squared* = 0.03. No other effects were significant (all *p*'s > .26).

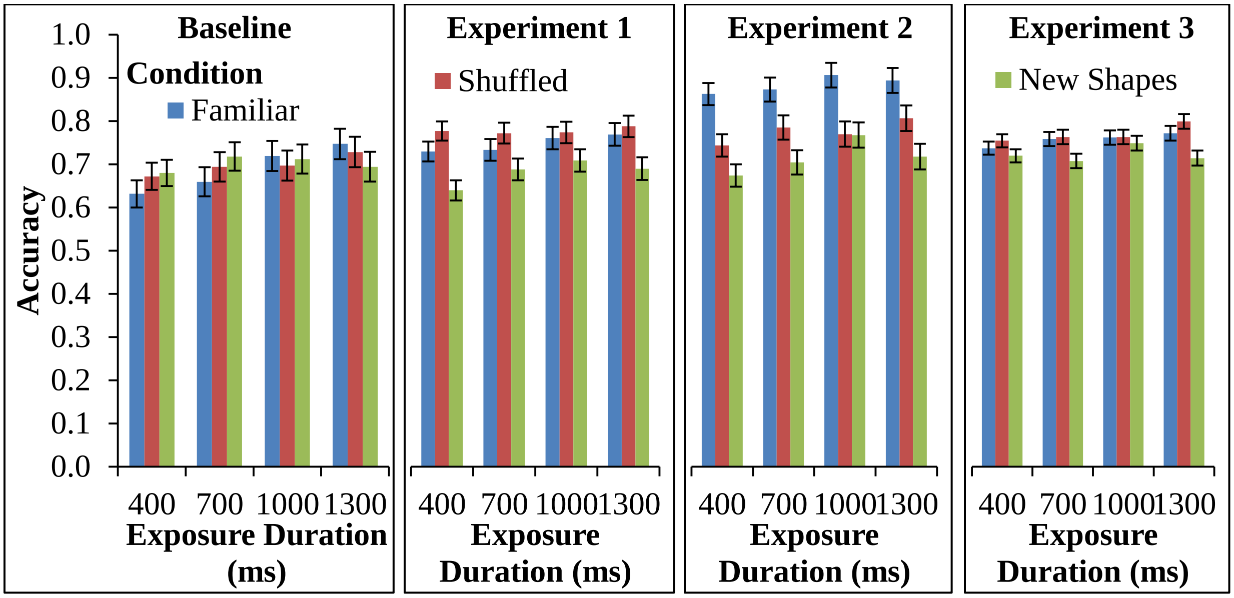

Figure 41. Experiment Version by Condition by Exposure Duration on accuracy (collapsed across Target Presence). Error bars indicate standard error of the mean.

The main effects were evaluated via custom hypothesis tests in ANOVA. For the main effect of Experiment Version, I first compared Experiment 1 (*M* = 0.74, *SE* = 0.01) and Experiment 3 (*M* = 0.75, *SE* = 0.01), which did not differ (*p* = .30). Then I compared Experiments 1 and 3 to Baseline (*M* = 0.70, *SE* = 0.02), and found that Experiments 1 and 3 together showed higher accuracy than Baseline, *F*(1,402) = 7.56, *p* = .006, *partial-eta-squared* = 0.02. Then I compared Experiments 1 and 3 to Experiment 2 (*M* = 0.79, *SE* = 0.01), and found that Experiments 1 and 3 showed lower accuracy than Experiment 2, *F*(1,402) = 11.44, *p* = .001, *partial-eta-squared* = 0.03. I also compared Experiment 2 to Baseline, and found that

Experiment 2 also had higher accuracy than Baseline, *F*(1,402) = 22.96, *p* < .001, *partial-eta-squared* = 0.06. All experiments showed learning relative to Baseline and experiment 2 showed the most learning.

For Condition, I compared Familiar (*M* = 0.77, *SE* = 0.01) and Shuffled (*M* = 0.76, *SE* = 0.01), which did not differ in accuracy (*p* = .34). Then I compared Familiar and Shuffled to New Shapes (*M* = 0.71, *SE* = 0.01), and found that Familiar and Shuffled were more accurate than New Shapes, *F*(1,402) = 19.87, *p* < .001, *partial-eta-squared* = 0.05. For Exposure Duration, I conducted all pairwise comparisons. 400ms (*M* = 0.72, *SE* = 0.01) was less accurate than 1000ms (*M* = 0.76, *SE* = 0.01), *F*(1,402) = 29.19, *p* < .001, *partial-eta-squared* = 0.07. 400ms was also less accurate than 1300ms (*M* = 0.76, *SE* = 0.01), *F*(1,402) = 34.74, *p* < .001, *partial-eta-squared* = 0.08. 700ms (*M* = 0.74, *SE* = 0.01) was marginally less accurate than 1300ms, *F*(1,402) = 7.51 *p* = .006, *partial-eta-squared* = 0.02. No other comparisons were significant (all *p*’s > .01).

Custom hypothesis tests in ANOVA of simple effects of Experiment Version were used for the interaction of Experiment Version and Condition. For Familiar, there was a simple effect of Experiment Version, *F*(3,402) = 12.71, *p* < .001, *partial-eta-squared* = 0.09. I first compared Experiment 1 (*M* = 0.75, *SE* = 0.02) and Experiment 3 (*M* = 0.76, *SE* = 0.01), which did not differ in accuracy (*p* = .70). Then I compared Experiments 1 and 3 to Baseline. Combined, Experiments 1 and 3 showed higher accuracy than Baseline (*M* = 0.69, *SE* = 0.03), *F*(1,402) = 4.64, *p* = .03, *partial-eta-squared* = 0.01. Then I compared Experiment 2 (*M* = 0.88, *SE* = 0.02) to Experiments 1 and 3, and found that Experiment 2 showed higher accuracy than Experiments 1 and 3, *F*(1,402) = 27.67, *p* < .001, *partial-eta-squared* = 0.06. Finally, I compared Experiment 2 to Baseline and showed that Experiment 2 also had higher accuracy than Baseline, *F*(1,402) = 31.45, *p* < .001, *partial-eta-squared* = 0.07. There was also a marginal simple effect of

Experiment Version for Shuffled, *F*(3,402) = 2.36, *p* = .067 *partial-eta-squared* = 0.02. Again, I first compared Experiment 1 (*M* = 0.78, *SE* = 0.03) and Experiment 3 (*M* = 0.77, *SE* = 0.01), which did not differ in accuracy (*p* = .75). Then I compared Experiments 1 and 3 to Experiment 2 (*M* = 0.78, *SE* = 0.02), which did not differ in accuracy (*p* = .92). Finally, I compared all three experiments to Baseline (*M* = 0.70, *SE* = 0.03), and found that all three Experiments (1, 2, and 3) combined showed higher accuracy than Baseline, *F*(1,402) = 7.03, *p* = .008, *partial-eta-squared* = 0.02. There was no simple effect of Experiment Version for New Shapes (*p* = .38). The interaction of Experiment Version and Condition was driven by the simple effect of Experiment Version for Familiar - Experiment 2 showing higher accuracy than Experiments 1 and 3 which had higher accuracy than Baseline - and the simple effect of Experiment Version for Shuffled - all experiments not differing in accuracy, but together having higher accuracy than Baseline.

For Experiment Version and Target Presence, custom hypothesis tests in ANOVA of simple effects of Experiment Version for each level of target present were examined, and significant simple effects were followed up with all pairwise comparisons. There was a simple effect of Experiment Version when the target was Present, *F*(3,402) = 3.74, *p* = .01, *partial-eta-squared* = 0.03. Experiment 2 (*M* = 0.79, *SE* = 0.02) showed higher accuracy than Experiment 1 (*M* = 0.73, *SE* = 0.02), *F*(1,402) = 5.46, *p* = .02, *partial-eta-squared* = 0.01. Experiment 2 also showed higher accuracy than Experiment 3 (*M* = 0.72, *SE* = 0.01), *F*(1,402) = 10.20, *p* = .002, *partial-eta-squared* = 0.03. Experiment 2 also showed higher accuracy than Baseline (*M* = 0.72, *SE* = 0.02), *F*(1,402) = 6.77, *p* = .01, *partial-eta-squared* = 0.02. No other pairwise comparisons were significant for Present (all *p*'s > .52). There was also a simple effect of Experiment Version for Absent, *F*(3,402) = 8.18, *p* < .001, *partial-eta-squared* = 0.06. Experiment 1 (*M* = 0.74, *SE* = 0.02) showed lower accuracy than Experiment 3 (*M* = 0.78, *SE* = 0.01), *F*(1,402) = 4.23, *p* = .04, *partial-eta-squared* = 0.01. Experiment 1 also showed lower accuracy than Experiment 2 (*M* = 0.80, *SE* = 0.02), *F*(1,402) = 5.92, *p* = .02, *partial-eta-squared* = 0.02. Experiment 1 showed higher accuracy than Baseline (*M* = 0.72, *SE* = 0.02), *F*(1,402) = 5.67, *p* = .02, *partial-eta-squared* = 0.01.Experiment 2 showed higher accuracy than Baseline (*M* = 0.68, *SE* = 0.02), *F*(1,402) = 18.95, *p* < .001, *partial-eta-squared* = 0.05. Experiment 3 also showed higher accuracy than Baseline, *F*(1,402) = 18.45, *p* < .001, *partial-eta-squared* = 0.04. Experiment 2 and Experiment 3 did not differ in accuracy for Absent, (*p* = .36). The interaction of Experiment Version and Target Presence was driven by different patterns of accuracies for Present and Absent.

For the interaction of Condition and Exposure Duration, custom hypothesis tests in ANOVA were used to test simple effects of Condition. There was a simple effect of Condition for 400ms, *F*(2,402) = 8.14, *p* < .001, *partial-eta-squared* = 0.04. I compared Familiar (*M* = 0.74, *SE* = 0.01) and Shuffled (*M* = 0.74, *SE* = 0.01), which did not differ in accuracy (*p* = .84). Then I compared Familiar and Shuffled combined to New Shapes (*M* = 0.68, *SE* = 0.01), and found that the combined conditions were more accurate than New Shapes, *F*(1,402) = 16.25, *p* < .001, *partial-eta-squared* = 0.04. There was also a simple effect of Condition for 700ms, *F*(2,402) = 4.82, *p* = .009, *partial-eta-squared* = 0.02. Similarly, I compared Familiar (*M* = 0.76, *SE* = 0.01) and Shuffled (*M* = 0.75, *SE* = 0.01), which did not differ in accuracy (*p* = .90). When I compared Familiar and Shuffled combined to New Shapes (*M* = 0.70, *SE* = 0.01), I found that the combination was more accurate than New Shapes, *F*(1,402) = 9.62, *p* = .002, *partial-eta-squared* = 0.02. Again, there was a simple effect of Condition for 1300ms, *F*(2,402) = 12.65, *p* < .001, *partial-eta-squared* = 0.06. I compared Familiar (*M* = 0.80, *SE* = 0.01) and Shuffled (*M* = 0.78, *SE* = 0.01), which again did not differ in accuracy (*p* = .45). When I compared Familiar and

Shuffled combined to New Shapes (*M* = 0.70, *SE* = 0.01), I found that the combined conditions were more accurate than New Shapes, *F*(1,402) = 24.73, *p* < .001, *partial-eta-squared* = 0.06. Finally, there was a simple effect of Condition at 1000ms too, *F*(2,402) = 3.91, *p* = .02, *partial-eta-squared* = 0.02. However, in contrast to the other simple effects, at 1000ms I compared Shuffled (*M* = 0.75, *SE* = 0.01) and New Shapes (*M* = 0.73, *SE* = 0.01), which did not differ in accuracy (*p* = .39). When I compared Shuffled and New Shapes to Familiar (*M* = 0.79, *SE* = 0.01), I found that Familiar was more accurate than Shuffled and New Shapes combined, *F*(1,402) = 7.07, *p* = .008, *partial-eta-squared* = 0.02. The interaction of Condition and Exposure Duration was driven by the different patterns of accuracy by condition for 1000ms than for the other exposure durations.

The interaction of Exposure Duration and Target Presence was explored via custom hypothesis tests in ANOVA. There was a marginal simple effect of Target Presence at 400ms, such that participants were more accurate when the target was Absent (*M* = 0.74, *SE* = 0.01) than when the target was Present (*M* = 0.69, *SE* = 0.01), *F*(1,402) = 9.67, *p* = .002, *partial-eta-squared* = 0.02. There were no other simple effects of Target Presence (all *p*’s >.01). The interaction of Exposure Duration and Target Presence was driven by the marginal simple effect of Target Presence at 400ms.

False Alarm Rate. Figure 42 showed effects of Experiment Version, Condition, and Exposure Duration on false alarm rate. Baseline appeared to have the highest false alarm rate across conditions and exposure durations, and Experiment 2 the lowest, especially for Familiar. Shuffled appeared to have the lowest false alarm rate across exposure durations for Experiments 1 and 3. To test these apparent effects, I conducted an ANOVA of Experiment Version by Condition by Exposure Duration on false alarm rate revealed main effects of Experiment

Version, *F*(3,402) = 8.18, *p* < .001, *partial-eta-squared* = 0.06; and Condition, *F*(2,402) = 8.93, *p* < .001, *partial-eta-squared* = 0.04. No other effects were significant (all *p*'s > .18). Custom hypothesis tests were used to follow up on the significant effects.

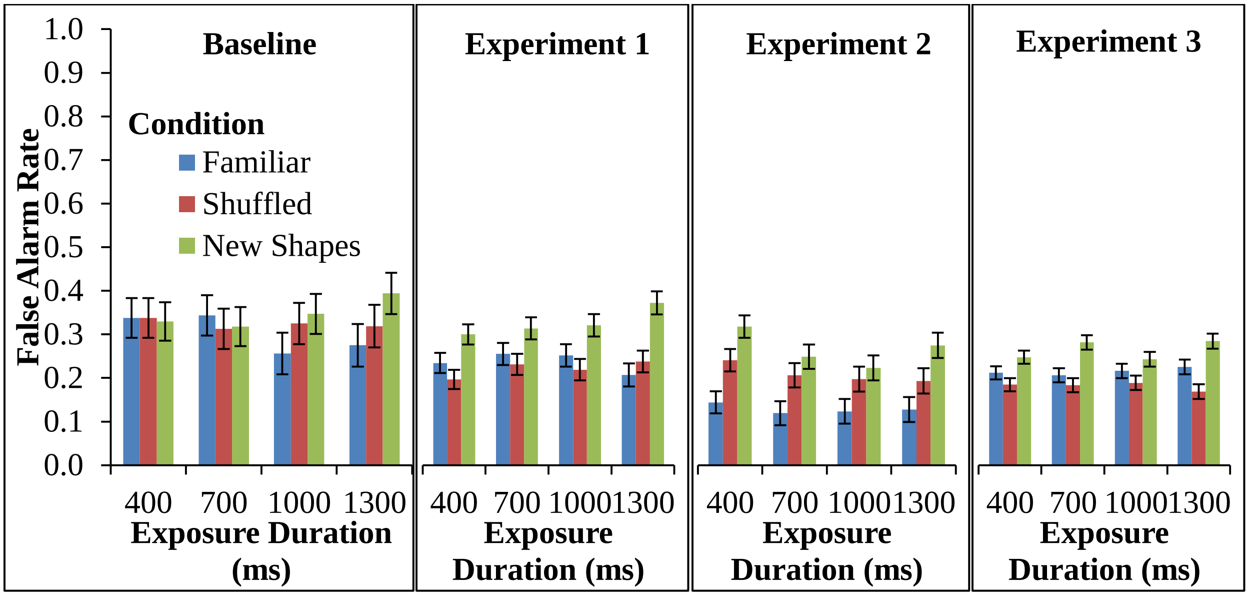

Figure 42. Experiment Version by Condition by Exposure Duration on accuracy (collapsed across Target Presence). Error bars indicate standard error of the mean.

All pairwise comparisons were made to investigate Experiment Version. Baseline (*M* = 0.33, *SE* = 0.02) had a higher false alarm rate than Experiment 1 (*M* = 0.26, *SE* = 0.02), *F*(1,402) = 5.67, *p* = .02, *partial-eta-squared* = 0.01. Baseline also had a higher false alarm rate than Experiment 2 (*M* = 0.20, *SE* = 0.02), *F*(1,402) = 18.95, *p* < .001, *partial-eta-squared* = 0.05. Experiment 3 (*M* = 0.22, *SE* = 0.01) also had a lower false alarm rate than Baseline, *F*(1,402) = 18.45, *p* < .001, *partial-eta-squared* = 0.04. Experiment 1 had more false alarms than Experiment 2, *F*(1,402) = 5.92, *p* = .02, *partial-eta-squared* = 0.02. Experiment 1 also had more false alarms than Experiment 3, *F*(1,402) = 4.23, *p* = .04, *partial-eta-squared* = 0.01.

Experiments 2 and 3 did not differ in false alarms (*p* = .36). All three experiments decreased their false alarm rates relative to Baseline.

For Condition, Familiar (*M* = 0.22, *SE* = 0.02) and Shuffled (*M* = 0.24, *SE* = 0.01) did not differ in false alarm rate (*p* = .53). New Shapes (*M* = 0.30, *SE* = 0.01) had more false alarms than the other two conditions combined, *F*(1,402) = 17.48, *p* < .001, *partial-eta-squared* = 0.04. There was transfer from Familiar to Shuffled across experiments.

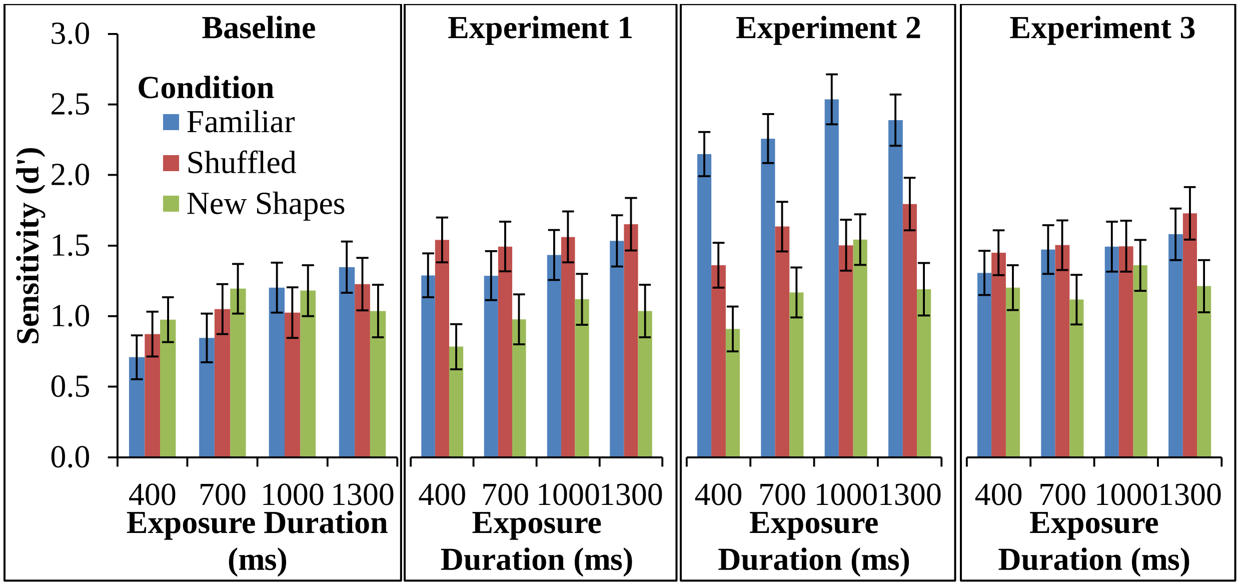

Figure 43. Experiment Version by Condition by Exposure Duration on sensitivity. Error bars indicate standard error of the mean.

Sensitivity. Figure 43 showed effects of Experiment Version, Condition, and Exposure Duration on sensitivity. It appeared that Experiment 2 had the highest sensitivity across exposure durations and conditions, but especially for Familiar. Shuffled appeared to have the highest sensitivity in Experiments 1 and 3. Baseline appeared to have the lowest sensitivity across exposure durations and conditions. To test these apparent effects, I conducted an ANOVA of

Experiment Version by Condition by Exposure Duration on sensitivity, which revealed main effects of Experiment Version, *F*(3,402) = 9.72, *p* < .001, *partial-eta-squared* = 0.07; Condition, *F*(2,402) = 11.38, *p* < .001, *partial-eta-squared* = 0.05; and Exposure Duration, *F*(3,1206) = 13.38, *p* < .001, *partial-eta-squared* = 0.03. It also revealed interaction effects of Experiment Version with Condition, *F*(6,402) = 4.64, *p* < .001, *partial-eta-squared* = 0.07; and Condition with Exposure Duration, *F*(6,1206) = 2.43, *p* = .02, *partial-eta-squared* = 0.01. No other effects were significant (all *p*'s > .69).

Custom hypothesis tests in ANOVA were used to follow up on main effects. For Experiment Version, I first compared Experiment 1 (*M* = 1.31, *SE* = 0.07) and Experiment 3 (*M* = 1.41, *SE* = 0.05), which did not differ in sensitivity (*p* = .23). Then I compared Baseline (*M* = 1.06, *SE* = 0.10) to Experiments 1 and 3, and found that Baseline showed lower sensitivity than Experiments 1 and 3, *F*(1,402) = 8.50, *p* = .004, *partial-eta-squared* = 0.02. I compared Experiment 2 (*M* = 1.70, *SE* = 0.08) to Experiments 1 and 3, and found that Experiment 2 showed higher sensitivity than Experiments 1 and 3, *F*(1,402) = 14.46, *p* < .001, *partial-eta-squared* = 0.04. I also compared Experiment 2 to Baseline, and found that Experiment 2 also showed higher sensitivity than Baseline, *F*(1,402) = 27.10, *p* < .001, *partial-eta-squared* = 0.06. All experiments showed learning relative to Baseline, but Experiment 2 showed more learning than the other two experiments.

For Condition, I first compared Familiar (*M* = 1.55, *SE* = 0.07) and Shuffled (*M* = 1.43, *SE* = 0.07), which did not differ in sensitivity (*p* = .19). Then I compared Familiar and Shuffled together to New Shapes (*M* = 1.13, *SE* = 0.07), and found that together they showed higher sensitivity than New Shapes, *F*(1,402) = 21.07, *p* < .001, *partial-eta-squared* = 0.05. There was improved sensitivity and transfer of learning from Familiar to Shuffled across experiments.

For Exposure Duration, I first compared 1000ms (*M* = 1.45, *SE* = 0.05) and 1300ms (*M* = 1.48, *SE* = 0.05), which did not differ in sensitivity (*p* = .63). Then I compared 400ms (*M* = 1.21, *SE* = 0.04) to 1000ms and 1300ms, and found that 400ms showed less sensitivity than the longer exposure durations combined, *F*(1,402) = 45.63, *p* < .001, *partial-eta-squared* = 0.10. Then I compared 700ms (*M* = 1.33, *SE* = 0.05) to 1000ms and 1300ms, and found that 700ms showed marginally lower sensitivity than the longer exposure durations combined, *F*(1,402) = 9.12, *p* = .003, *partial-eta-squared* = 0.02. I also compared 400ms and 700ms, which did not differ (*p* = .01).

For the interaction of Experiment Version and Condition, custom hypothesis tests were used to investigate simple effects of Experiment Version for each condition, and additional custom hypothesis tests were used for significant simple effects. There was a significant simple effect of Experiment Version for Familiar, *F*(3,402) = 15.27, *p* < .001, *partial-eta-squared* = 0.10. I compared Experiment 1 (*M* = 1.39, *SE* = 0.12) and Experiment 3 (*M* = 1.46, *SE* = 0.08), which did not differ in sensitivity (*p* = .60). Then I compared Experiments 1 and 3 together to Baseline (*M* = 1.03, *SE* = 0.17), and found that together they showed higher sensitivity than Baseline, *F*(1,402) = 4.78, *p* = .03, *partial-eta-squared* = 0.01. I compared Experiment 2 (*M* = 2.33, *SE* = 0.14) to the other two experiments and found that Experiment 2 showed higher sensitivity than Experiments 1 and 3, *F*(1,402) = 34.36, *p* < .001, *partial-eta-squared* = 0.08. I also compared Experiment 2 to Baseline and found that Experiment 2 also showed higher sensitivity than Baseline, *F*(1,402) = 36.81, *p* < .001, *partial-eta-squared* = 0.08. There was also a simple effect of Experiment Version for Shuffled, *F*(3,402) = 2.77, *p* = .04, *partial-eta-squared* = 0.02. I first compared Experiment 1 (*M* = 1.56, *SE* = 0.12) and Experiment 3 (*M* = 1.54, *SE* = 0.08), which did not differ in sensitivity (*p* = .90). Next, I compared Experiment 2 (*M* = 1.57, *SE*

= 0.14) to the other two experiments, and found that Experiment 2 did not differ from Experiments 1 and 3 in sensitivity (*p* = .90). Finally, I compared Baseline (*M* = 1.04, *SE* = 0.17) to all three experiments combined, and found that Baseline showed lower sensitivity than Experiments 1, 2, and 3 combined, *F*(1,402) = 8.25, *p* = .004, *partial-eta-squared* = 0.02. There was no simple effect for New Shapes (*p* = .40). The interaction of Experiment Version and Condition was driven by learning in improved sensitivity in all experiments for Familiar and Shuffled relative to Baseline, and the significantly higher sensitivity for Experiment 2 for Familiar (but not Shuffled).

For the interaction of Condition and Exposure Duration, custom hypothesis tests examined simple effects of Condition for each exposure duration, and additional custom tests were used to follow up significant simple effects. For 400ms, there was a simple effect of Condition, *F*(2,402) = 8.24, *p* < .001, *partial-eta-squared* = 0.04. I compared Familiar (*M* = 1.36, *SE* = 0.08) and Shuffled (*M* = 1.31, *SE* = 0.07), which did not differ in sensitivity (*p* = .59). Then I compared Familiar and Shuffled together to New Shapes (*M* = 0.97, *SE* = 0.07), and found that together they showed higher sensitivity than New Shapes, *F*(1,402) = 16.20, *p* < .001, *partial-eta-squared* = 0.04. For 700ms, there was also a simple effect of Condition, *F*(2,402) = 5.34, *p* = .005, *partial-eta-squared* = 0.03. Similarly, I first compared Familiar (*M* = 1.47, *SE* = 0.08) and Shuffled (*M* = 1.42, *SE* = 0.08), which did not differ in sensitivity (*p* = .70), and then compared Familiar and shuffled together to New Shapes (*M* = 1.11, *SE* = 0.08), and found that together they showed higher sensitivity than New Shapes, *F*(1,402) = 10.53, *p* = .001, *partial-eta-squared* = 0.03. For 1300ms, there was also simple effect of Condition, *F*(2,402) = 13.21, *p* < .001, *partial-eta-squared* = 0.06. Again, I compared Familiar (*M* = 1.71, *SE* = 0.09) and Shuffled (*M* = 1.60, *SE* = 0.09), which did not differ in sensitivity (*p* = .36). When I compared Familiar and Shuffled together to New Shapes (*M* = 1.12, *SE* = 0.09), I found that together they showed higher sensitivity than New Shapes, *F*(1,402) = 25.61, *p* < .001, *partial-eta-squared* = 0.06. For 1000ms, there was another simple effect of Condition, *F*(2,402) = 5.03, *p* = .007, *partial-eta-squared* = 0.02. In contrast to the other simple effects, I compared Shuffled (*M* = 1.40, *SE* = 0.08) and New Shapes (*M* = 1.30, *SE* = 0.08), which did not differ in sensitivity (*p* = .43). Then I compared Shuffled and New Shapes together to Familiar (*M* = 1.67, *SE* = 0.09), and found that together they showed lower sensitivity than Familiar, *F*(1,402) = 9.42, *p* = .002, *partial-eta-squared* = 0.02. The interaction of Condition and Exposure Duration was driven by the different pattern of results for 1000ms compared to the other exposure durations: 1000ms did not show transfer to Shuffled on sensitivity, but the other exposure durations did.

How do all of the experiments compare on survey measures?

Experiments 1, 2, and 3 used the same survey (with the exception of adjustment to the sleep scale) as well as the same psychophysical assessment. Given that noticing the pairs was associated with improved performance in Experiments 2 and 3, and that the linguistic coding strategy was associated with higher assessment accuracy in Experiment 3, I examined the relationships of these measures across experiments.

Correlations Noticing the pairs correlated with increased accuracy across experiments, *r*(343) = 0.31, *p* < .001, as did use of the linguistic coding strategy, *r*(353) = 0.19, *p* < .001. I followed up on these correlations with ANCOVA analyses of assessment data.

Accuracy. Figure 44 showed effects of Experiment Version, Condition, and Exposure Duration on accuracy. It appeared that Experiment 2 had higher accuracy for Familiar than the other experiments, but otherwise performance across experiments was similar with higher accuracy in Shuffled than New Shapes. To test these apparent effects, I conducted an ANCOVA of Experiment Version by Condition by Exposure Duration by Target Presence on accuracy, covarying out Noticing and Linguistic Coding. This revealed a main effect of Condition, $F(2,324) = 15.07$, $p < .001$, *partial-eta-squared* = 0.09, and a marginal main effect of Exposure Duration, $F(3,972) = 2.59$, $p = .05$, *partial-eta-squared* = 0.01. There were several interactions: Experiment Version and Condition, $F(4,324) = 5.26$, $p < .001$, *partial-eta-squared* = 0.06; Condition and Exposure Duration, $F(6,972) = 2.11$, $p = .05$, *partial-eta-squared* = 0.01; Exposure Duration and Target Presence, $F(3,972) = 3.57$, $p = .01$, *partial-eta-squared* = 0.01; a marginal interaction of Experiment Version and Target Presence, $F(2,324) = 2.50$, $p = .08$, *partial-eta-squared* = 0.02; Target Presence and Noticing, $F(1,324) = 4.82$, $p = .03$, *partial-eta-squared* = 0.02; and a marginal interaction of Exposure Duration and Noticing, $F(1,972) = 2.59$, $p = .05$, *partial-eta-squared* = 0.01. There were also significant effects of both covariates: Noticing, $F(1,324) = 18.60$, $p < .001$, *partial-eta-squared* = 0.05; and Linguistic Coding, $F(1,324) = 7.17$, $p = .008$, *partial-eta-squared* = 0.02. No other effects were significant (all *p*'s > .10).

Custom hypothesis tests in ANCOVA on adjusted marginal means[27] were used to explore significant effects. For the main effect of Condition, I compared Familiar ($M = 0.79$, $SE = 0.01$) and Shuffled ($M = 0.77$, $SE = 0.01$), which showed similar accuracy ($p = .18$). Then I compared Familiar and Shuffled together to New Shapes ($M = 0.71$, $SE = 0.01$), and found that Familiar and Shuffled showed higher accuracy than New Shapes, $F(1,330) = 28.16$, $p < .001$, *partial-eta-squared* = 0.08. Learning transferred from Familiar to Shuffled.

[27] Marginal means were calculated at the mean of each covariate: Recognition $M = 0.64$, Noticing $M = 0.32$, Linguistic Coding $M = 0.40$.

For Exposure Duration, I compared 1000ms (*M* = 0.77, *SE* = 0.01) and 1300ms (*M* = 0.77, *SE* = 0.01), which did not differ (*p* = .45). Then I compared the longest exposure durations to 400ms (*M* = 0.74, *SE* = 0.01), and found that 1000ms and 1300ms together showed marginally higher accuracy than 400ms, *F*(1,324) = 9.53, *p* = .002, *partial-eta-squared* = 0.03. No other comparisons were significant (all *p*'s > .04).

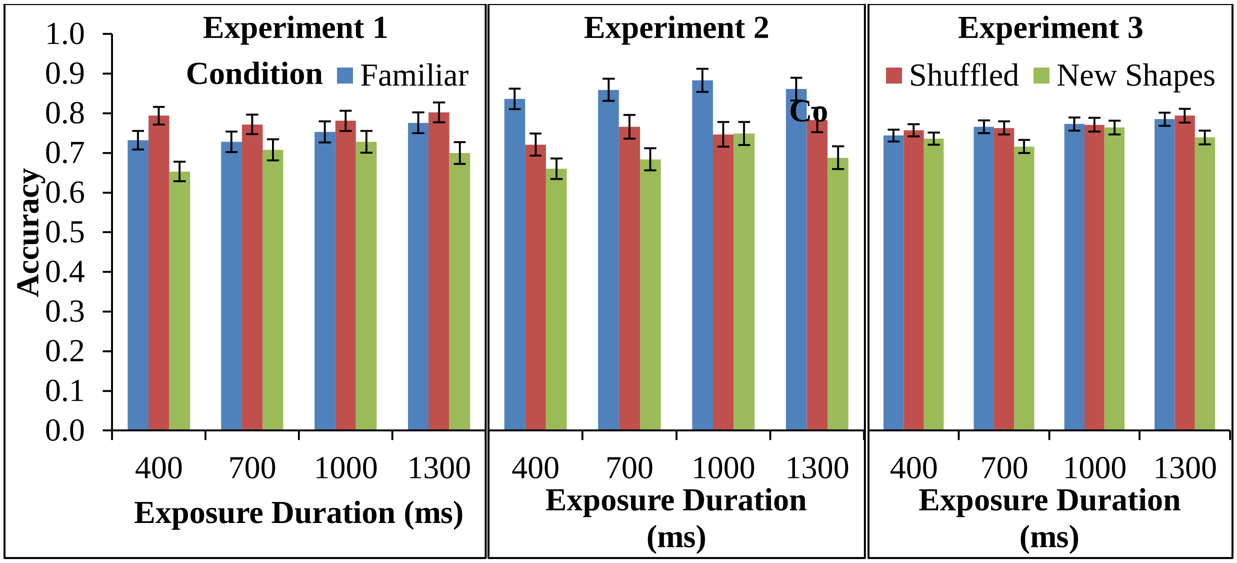

Figure 44. Experiment Version by Condition by Exposure Duration on accuracy (collapsed across Target Presence), statistically controlling for Recognition, Noticing, and Linguistic Coding. Bar heights are adjusted marginal means and error bars indicate standard error of the mean.

Custom hypothesis tests in ANCOVA were used for the interaction of Experiment Version and Condition, to test simple effects of Condition and follow up on significant simple effects. There was a simple effect of Condition for Familiar, *F*(2,324) = 8.42, *p* < .001, *partial-eta-squared* = 0.05. I compared Experiments 1 (*M* = 0.75, *SE* = 0.02) and 3 (*M* = 0.77, *SE* =

0.01), which did not differ (*p* = .41). Then I compared the SL familiarization experiments to Experiment 2 (*M* = 0.86, *SE* = 0.02), and found that Experiments 1 and 3 together showed higher accuracy than Experiment 2, *F*(1,324) = 16.83, *p* < .001, *partial-eta-squared* = 0.05. There were no other simple effects of Experiment Version (all *p*'s > .10). The interaction of Experiment Version and Condition was driven by Experiment 2 having higher accuracy than the other experiments for Familiar.

Custom hypothesis tests in ANCOVA were also used for the interaction of Experiment Version and Target Presence, but revealed no simple effects of Experiment Version (all *p*'s > .15) or simple effects of Target Presence (all *p*'s > .15).

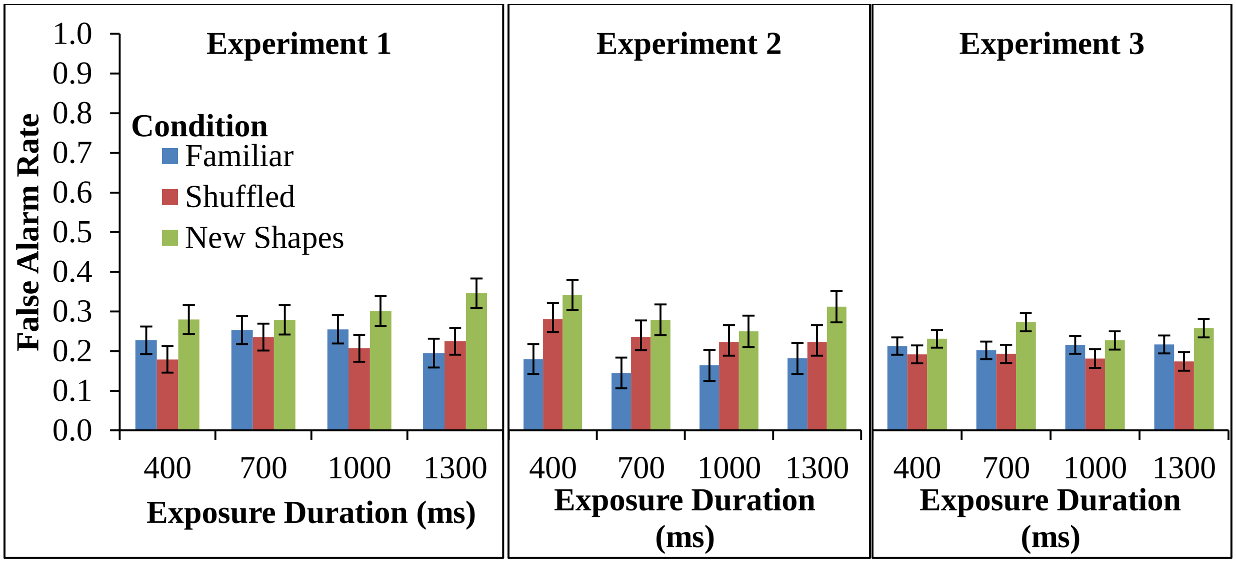

Figure 45. Experiment Version by Condition by Exposure Duration on false alarm rate, statistically controlling for Recognition, Noticing, and Linguistic Coding. Bar heights are adjusted marginal means and error bars indicate standard error of the mean.

False Alarm Rate. Figure 45 showed effects of Experiment Version, Condition, and Exposure Duration on false alarm rate, when statistically controlling for Noticing and Linguistic Coding. It appeared that false alarm rates were lower for Familiar and Shuffled than New Shapes across experiments, and lowest for Familiar in Experiment 2, but lowest for Shuffled in Experiments 1 and 3. To test these apparent effects, I conducted an ANCOVA of Condition by Exposure Duration on false alarm rate, covarying out Noticing and Linguistic Coding. There was a main effect of Condition, $F(2,324) = 7.91$, $p < .001$, *partial-eta-squared* = 0.05. There were also effects of the covariates Noticing, $F(1,324) = 21.49$, $p < .001$, *partial-eta-squared* = 0.06, and Linguistic Coding, $F(1,324) = 5.04$, $p = .03$, *partial-eta-squared* = 0.02, and a marginal interaction of Exposure Duration and Noticing, $F\ (1,972) = 2.52$, $p = .06$, *partial-eta-squared* = 0.01. No other effects were significant (all *p*’s > .18).

Custom hypothesis tests in ANCOVA were used to investigate the main effect of Condition. I compared Familiar ($M = 0.20$, $SE = 0.02$) and Shuffled ($M = 0.21$, $SE = 0.02$), which did not differ in false alarm rate ($p = .69$). Then I compared Familiar and Shuffled together to New Shapes ($M = 0.28$, $SE = 0.02$), and found that together they showed a lower false alarm rate than New Shapes, $F(1,324) = 15.62$, $p < .001$, *partial-eta-squared* = 0.05. Across experiments, learning transferred from Familiar to Shuffled as measured by decreased false alarm rates.

Sensitivity. Figure 46 showed effects of Experiment Version, Condition, and Exposure Duration on sensitivity, when effects of Noticing and Linguistic Coding were statistically controlled. It appeared that Experiment 2 had higher accuracy in Familiar than other conditions and than other experiments. It also appeared that Familiar and Shuffled showed higher sensitivity than New Shapes across experiments and exposure durations. To test the apparent effects, I conducted an ANCOVA of Condition by Exposure Duration on sensitivity, covarying out

Noticing and Linguistic Coding. There was a main effect of Condition, *F*(2,324) = 16.44, *p* < .001, *partial-eta-squared* = 0.09, and a main effect of Exposure Duration, *F*(3,972) = 3.08, *p* = .03, *partial-eta-squared* = 0.01. Both covariates had significant effects: Noticing, *F*(1,324) = 21.36, *p* < .001, *partial-eta-squared* = 0.06; and Linguistic Coding, *F*(1,324) = 6.45, *p* = .01, *partial-eta-squared* = 0.02. There was a significant interaction of Experiment Version and Condition, *F*(1,324) = 5.82, *p* < .001, *partial-eta-squared* = 0.07; a marginal interaction of Condition and Exposure Duration, *F*(6,972) = 1.90, *p* = .08, *partial-eta-squared* = 0.01; and a significant interaction of Exposure Duration and Noticing, *F*(3,972) = 3.09, *p* = .03, *partial-eta-squared* = 0.01. No other effects were significant (all *p*'s > .19).

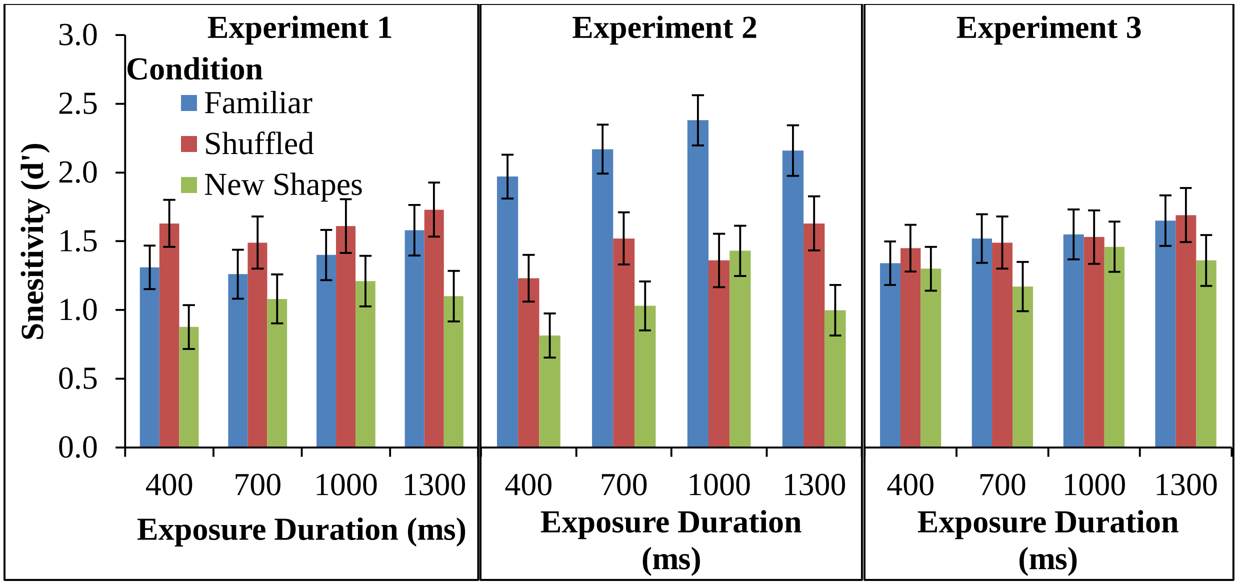

Figure 46. Experiment Version by Condition by Exposure Duration on sensitivity, statistically controlling for Recognition, Noticing, and Linguistic Coding. Bar heights are adjusted marginal means and error bars indicate standard error of the mean.

Custom hypothesis tests in ANCOVA were used to explore the main effects. For Condition, I conducted all pairwise comparisons. Familiar (*M* = 1.70, *SE* = 0.07) was marginally more sensitive than Shuffled (*M* = 1.54, *SE* = 0.07), *F*(1,324) = 2.75, *p* = .01, *partial-eta-squared* = 0.01. Familiar was significantly more sensitive than New Shapes (*M* = 1.16, *SE* = 0.07), *F*(1,324) = 31.38, *p* < .001, *partial-eta-squared* = 0.09. Shuffled significantly more sensitive than New Shapes, *F*(1,324) = 15.32, *p* < .001, *partial-eta-squared* = 0.05. The highest sensitivity was seen in Familiar, but learning also transferred to Shuffled.

For Exposure Duration, I compared 1000ms (*M* = 1.55, *SE* = 0.05) and 1300ms (*M* = 1.55, *SE* = 0.05), which did not differ (*p* = .54). I compared the longest durations combined to 400ms (*M* = 1.33, *SE* = 0.05), and found that together 1000ms and 1300ms were marginally more sensitive than 400ms, *F*(1,324) = 9.68, *p* = .002, *partial-eta-squared* = 0.03. No other comparisons were significant (all *p*'s > .03).

Custom hypothesis tests were used for the interaction of Experiment Version and Condition. I tested simple effects of Experiment Version for each condition, and followed up significant simple effects with additional tests. There was a simple effect of Experiment Version for Familiar, *F*(2,324) = 10.42, *p* < .001, *partial-eta-squared* = 0.06. I first compared Experiment 1 (*M* = 1.39, *SE* = 0.13) and Experiment 3 (*M* = 1.52, *SE* = 0.08), which did not differ in sensitivity (*p* = .38). Then I compared Experiments 1 and 3 combined to Experiment 2 (*M* = 2.18, *SE* = 0.14), and found that Experiments 1 and 3 together had lower sensitivity than Experiment 2, *F*(1,324) = 20.83, *p* < .001, *partial-eta-squared* = 0.06. There were no other simple effects of Experiment Version (all *p*'s > .12). The interaction of Experiment Version and Condition was driven by the higher sensitivity for Familiar in Experiment 2.

Custom hypothesis tests in ANCOVA were also used for the interaction of Condition and Exposure Duration. I tested for simple effects of Condition at each exposure duration, and for significant simple effects, I followed up with additional tests. For 400ms, there was a simple effect of Condition, *F*(2,324) = 13.56, *p* < .001, *partial-eta-squared* = 0.08. I compared Familiar (*M* = 1.54, *SE* = 0.08) and Shuffled (*M* = 1.44, *SE* = 0.08), which did not differ in sensitivity (*p* = .36), but when I tested them together against New Shapes (*M* = 1.00, *SE* = 0.08), I found that Familiar and Shuffled were more sensitive than New Shapes , *F*(1,324) = 26.16, *p* < .001, *partial-eta-squared* = 0.07. For 700ms, there was a similar simple effect of Condition, *F*(2,324) = 10.68, *p* < .001, *partial-eta-squared* = 0.06. I compared Familiar (*M* = 1.66, *SE* = 0.09) and Shuffled (*M* = 1.51, *SE* = 0.09), which did not differ in sensitivity (*p* = .23), but when I tested them together against New Shapes (*M* = 1.10, *SE* = 0.09), Familiar and Shuffled together were more sensitive than New Shapes , *F*(1,324) = 19.79, *p* < .001, *partial-eta-squared* = 0.06. Again, there was a similar simple effect of Condition for 1300ms, *F*(2,324) = 14.35, *p* < .001, *partial-eta-squared* = 0.08. I again compared Familiar (*M* = 1.80, *SE* = 0.09) and Shuffled (*M* = 1.69, *SE* = 0.09), which did not differ in sensitivity (*p* = .78), but when I tested them together against New Shapes (*M* = 1.16, *SE* = 0.09), I found that together Familiar and Shuffled were more sensitive than New Shapes, *F*(1,324) = 27.80, *p* < .001, *partial-eta-squared* = 0.08. In contrast, there was a different marginal simple effect of Condition at 1000ms, *F*(2,324) = 5.42, *p* = .005, *partial-eta-squared* = 0.03. I conducted all pairwise comparisons. Familiar (*M* = 1.78, *SE* = 0.09) showed higher sensitivity than Shuffled (*M* = 1.51, *SE* = 0.09), *F*(1,324) = 4.66, *p* = .03, *partial-eta-squared* = 0.01. Familiar also showed higher sensitivity than New Shapes (*M* = 1.37, *SE* = 0.09), *F*(1,324) = 10.39, *p* = .001, *partial-eta-squared* = 0.03. Shuffled and New Shapes did not differ (*p* = .29). The interaction of Condition and Exposure Duration was driven by the different simple effect at 1000ms: there was no transfer to Shuffled, unlike the other exposure durations.

Discussion

All of the experiments and the Baseline group participated in the same psychophysical assessment, so it was possible to quantitatively compare learning across Experiments 1, 2, and 3. These analyses explored questions about the relationships between findings across experiments, such as the relationship of the data of the direct replication session length of 21 minutes in Experiment 3 to the data of Experiment 1, or the relationships of all experiments to each other and to baseline performance. By statistically testing such questions, we gained answers that quantified such relationships and gave additional insights into the relationship of statistical learning (SL) and perceptual learning (PL).

Successful Direct Replication of Experiment 1

Experiment 3 at the 21-minute session length successfully replicated Experiment 1. Familiar and Shuffled showed higher accuracy and sensitivity and lower false alarm rates than New Shapes across both experiments. There was only a main effect of Experiment Version for false alarm rate such that Experiment 1 had more false alarms than Experiment 3, likely due to the higher accuracy for Absent trials in Experiment 3. For accuracy, Experiment 3 was also higher than Experiment 1 for Shuffled when the target was Present. These small differences could have been due to either sampling variability, or to the absence of gridlines for the third block of familiarization in Experiment 3, or both. There was no effect of Experiment Version for sensitivity, an analysis in which any difference in response biases were removed. The main pattern of results held across experiments. Experiment 3 at the 21-minute session length successfully replicated Experiment 1 in accuracy, false alarm rates, and sensitivity.

Effects of Short and Long Session Lengths

Weaker PL with a Shorter Session Length. PL effects were observed across session lengths in Experiment 3. Analyzing the 7-minute session length together with baseline performance demonstrated that PL effects for 7 minutes were quantitatively stronger than baseline performance. This was seen in main effects of Experiment Version such that experimental participants showed higher accuracy and sensitivity and low false alarm rates than Baseline group participants. Because overall performance in Experiment 3 was better than at Baseline, we quantitatively demonstrated learning with a short session length.

There were no main effects of Condition like those in Experiment 1 or across session lengths in Experiment 3. However, Condition and Exposure Duration interacted for accuracy and sensitivity: the pattern of results of transfer from Familiar to Shuffled (but not New Shapes) was only observed in the longest exposure duration for the 7-minute session length. Because only at 1300ms did we observe the same pattern of results as with other session lengths, learning with the 7-minute session length was weaker but qualitatively similar to other session lengths.

Increasing Session Length is Different than PL Training. Experiment 2 showed superior performance to Experiment 3 with a 35-minute session length. Experiment 2 had higher accuracy and sensitivity across conditions than Experiment 3. However, Experiment 2 was only more accurate, more sensitive, and less likely to false alarm for the Familiar condition. Experiments 2 and 3 (at the longest session length) did not differ for Shuffled and New Shapes. It appears that a statistical learning familiarization is sufficient to cause high performance in rejecting learned shape pairs (in Shuffled), but SL familiarization is not sufficient to cause high performance on identifying learned pairs present in the search grid (Familiar).

Comparing across Experiments

All experiments showed transfer from Familiar to Shuffled, improved sensitivity, and decreased false alarming relative to baseline performance, so all showed PL effects. Experiment 2 showed higher accuracy and higher sensitivity than Experiments 1 and 3 (across session lengths), and than Baseline performance. Experiments 1 and 3 did not differ from each other but together showed higher accuracy and sensitivity than Baseline. Experiment 2 only showed better performance than the other experiments in the learned condition (Familiar), so it appears that PL training primarily improves performance on trained stimuli. Performance did not differ across experiments for Shuffled though combined the experiments showed better performance than Baseline. No learning was seen in any experiments for New Shapes.

Noticing the pairs and use of the linguistic coding strategy were correlated with psychophysical assessment accuracy across experiments. Noticing the pairs and use of the linguistic coding strategy each predicted higher accuracy. ANCOVA analyses statistically controlling for these effects revealed the same patterns of results as seen in the main multi-experiment analyses: when noticing and linguistic coding were controlled for, participants still showed transfer from Familiar to Shuffled but not New Shapes in increased accuracy and sensitivity and decreased false alarming, and Experiment 2 still showed higher accuracy and sensitivity than the other experiments for Familiar.

Summary

In sum, most data support the hypothesis of co-occurrence of SL and PL, or a unified learning process. Analyses confirmed that Experiment 3 successfully replicated Experiment 1, both at the replicated session length and across session lengths. Analyses also indicated superior performance following perceptual learning (PL) training in Experiment 2, and that this advantage was limited to the Familiar condition, and these effects held when noticing the pairs and linguistic coding were factored out.Time spent on learning mattered less than the quality of learning. The lack of an SL effect as tested in Exp. 2, as well as the specificity of learning shown in the Familiar relative to Shuffled conditions, detracts somewhat from the hypothesis of a unified learning process. Possible interpretations of the totality of the results are considered in Chapter 9.

CHAPTER 9: GENERAL DISCUSSION

Perceptual learning (PL) and statistical learning (SL) are conceptually distinct. PL is about perceptual pickup of patterns and structure in current stimulation, so it predicts improvement in sensitivity and transfer. SL is about tracking co-occurrences of items in the world. In encounters with new stimulation, SL might predict that those precise elements in their precise relationship will continue appearing, so SL appears as familiarity, and might possibly appear as false alarming to elements in their relationships even when they do not actually appear as predicted. The relationship between SL and PL in humans is unknown. It is possible that they are distinct kinds of learning (hypothesis of different kinds of learning), or that they are part of a single, unified learning process (hypothesis of a unified learning process), or that the relationship might be more complex.

If SL and PL are different learning processes, then paradigms inducing SL might show no evidence of PL, and vice versa. If PL and SL are a unified learning process, then exposure to stimuli with statistically reliable relationships would cause PL effects. If the relationship is more complex, it could, perhaps, be asymmetric. Many of our findings support the hypothesis of a unified learning process – that PL and SL are, in fact, a unified learning process in humans. All three experiments showed significant transfer of learning from familiarized (or trained, in Experiment 2) pairs (Familiar) to new pairs made of the same shapes but in new relationships (Shuffled), improved sensitivity, and reduced false alarm rates relative to Baseline group performance, even though two of them (Experiments 1 and 3) were SL paradigms and only Experiment 2 was a PL paradigm. However, assessment performance was only best in Familiar for Experiment 2. Also, familiarity (SL) did not show strong or uniform correlations with assessment accuracy across experiments, so it appears that the relationship between SL and PL is complex. We will further discuss the relationship of SL and PL by reviewing the findings of the experiments, discussing the match between our data and several different possible relationships of SL and PL, dealing with objections to our interpretation, comparing our findings to Bayesian theories, discussing what participants learned, connecting our interpretation to the SL literature, and finishing with limitations, future directions, and conclusions.

Experiments and Findings

Statistical learning and perceptual learning have been studied in different paradigms using different methodologies, so our approach was to start with replicating a well-known SL paradigm and test afterward for PL effects (Experiments 1 and 3). Separately, we used a paradigm expected to yield PL effects with the same stimuli, and followed this manipulation by the same posttest PL assessment as in the other experiments (Experiment 2). We did so by designing a psychophysical assessment of learning that could discriminate between SL and PL, by testing for transfer (a PL effect) and improved sensitivity (also a PL effect), and secondarily by probing for bias (a possible SL effect). The hypothesis of two kinds of learning predicted different patterns of results from the SL familiarization and from the PL training; in contrast, the hypothesis of a unified learning process predicted similar results across all experiments.

Experiment 1. The goals of Experiment 1 were to replicate Fiser and Aslin (2001)'s SL paradigm and to develop an assessment of learning that would discriminate between PL effects and possible SL effects. We were successful in both replicating the paradigm and in developing the assessment. Participants showed familiarity for learned pairs, learning of the shapes in the spatial pairings seen during familiarization (Familiar), and even transfer of learning to new pairs made of the same shapes shuffled into new relationships (Shuffled). Performance on the Familiar and Shuffled conditions did not differ for accuracy, sensitivity, or false alarm rate, but was significantly better (more accurate, more sensitive, less likely to false alarm) than when new shapes were paired in the same way (New Shapes) on all of these dependent measures.

Different performance for Familiar and Shuffled than New Shapes could have indicated that the New Shapes were harder, causing decreased performance. I ruled out this explanation by comparing participants' performance to performance of other participants on the assessment without familiarization (Baseline group). This analysis demonstrated that performance in the New Shapes condition did not differ from Baseline, but that Familiar and Shuffled showed significantly better performance. Given the novelty and importance of the findings and conclusions of Experiment 1 – of PL effects following an SL paradigm – it was important to directly replicate the experiment (Experiment 3) and to seek converging evidence for the conclusions (Experiment 2).

Experiment 2: Targets and Psychophysical Analyses. In Experiment 2, we sought converging evidence for the hypothesis of SL and PL occurring together (the hypothesis of a unified learning process) by designing a PL training paradigm based on much prior research PL research (including Schneider & Shiffrin, 1977) using the stimuli from Experiment 1 (Fiser & Aslin, 2001). We trained participants to search for shape pairs shown in advance of presentations of arrays: the target pair was presented, followed by a search grid containing the target (when present) and additional shapes for a total of six shapes per grid, followed by feedback. We used this paradigm because tasks involving recurrent search for particular stimuli typically leads to PL effects of faster and more accurate extraction of these stimuli (e.g., Schneider & Shiffrin, 1977; Karni & Sagi, 1993). Participants' accuracy for each of their six shape pairs was tracked, and as their accuracy reached a threshold for a given search grid on-screen duration, the search grid duration was reduced, until participants reached mastery criteria for all six pairs.

After training, participants completed the same psychophysical assessment used in Experiment 1. Following PL training in Experiment 2, participants' psychophysical data showed PL effects of improved sensitivity for extracting information from briefly presented grids, particularly in extracting Familiar pairs, but also showing some transfer to Shuffled. Importantly, this overall pattern of results was qualitatively similar to the pattern observed following SL in Experiment 1, and much more similar than the hypothesis of different kinds of learning would predict.

Experiment 3. Experiment 3 directly replicated and expanded on Experiment 1 by exploring whether increasing and decreasing the familiarization session length could strengthen or weaken the observed effects. I used three session lengths: 7 minutes, as in the original study (Fiser & Aslin, 2001); 21 minutes, replicating the familiarization session length used in Experiment 1; and 35 minutes. Participants showed higher recognition for the 21 and 35 minutes session lengths (which did not differ) than for 7 minutes, though they showed higher recognition than chance across session lengths and at all session lengths.

Across session lengths, I replicated the pattern of results of Experiment 1: transfer from Familiar to shuffled and improved sensitivity on the psychophysical assessment. There were no main effects of Session Length for the assessment. However, Session Length was involved in interactions because learning was weaker, though qualitatively similar, for the 7 minute session length than the other session lengths. Increasing the session length did not increase SL as measured by recognition accuracy or increase performance on the assessment, but decreasing session length did decrease both recognition accuracy and assessment performance. Covarying out the effects of recognition, noticing the pair, and use of the linguistic coding strategy weakened the observed effects, but there was still transfer to Shuffled and improved sensitivity, so the results still supported the hypothesis of a unified learning process.

Multi-Experiment Analyses. Quantitatively comparing across experiments confirmed the qualitative similarity of results across experiments. Experiment 2 showed higher accuracy and sensitivity and fewer false alarms than the other experiments only for the Familiar condition. Experiments 1, 2, and 3 (across session lengths) showed equally strong transfer for Shuffled condition, with and without controlling for noticing the pairs and for use of the linguistic coding strategy. The experiments showed significantly better performance than the Baseline group for Familiar and Shuffled, but not for New Shapes. Comparing Experiment 1 and Experiment 3 at the 21-minute session length showed that the direct replication was successful. Learning for the shortest 7-minute session length in Experiment 3 was weaker than for longer session lengths, but it was still above Baseline, particularly for the longest exposure duration. Extending the session length to 35 minutes in Experiment 3 did not increase performance for Familiar to a level similar to that seen in Experiment 2, though performance for Shuffled and New Shapes did not differ for Experiments 2 and 3.

The method of learning (active training versus passive familiarization) appeared to matter more than just the time spent learning, as there was no correlation of time spent learning and assessment accuracy across experiments. Increasing session length from 7 minutes to 21 improved recognition and assessment performance in Experiment 3, but increasing again to 35 minutes did not yield any additional gains. In contrast, Participants in Experiment 2 who needed fewer learning trials (and thus, somewhat less time) were performed better on the assessment than those who needed more PL learning trials, so learning was inversely related to the amount of time spent on learning.

Experiment 2 participants spent an average of 24.65 minutes (*SE* = 0.42) on their learning. Participants in Experiments 1 and 3 showed less accuracy and sensitivity and more false alarms than participants in Experiment 2. All participants in Experiment 1 and participants in Experiment 3 in the 21- and 35-minute session lengths spent a similar or longer amount of time learning than the Experiment 2 average. Active learning (PL training) in Experiment 2 appeared to be more effective than passive familiarization in Experiments 1 and 3.

Experiment 2: SL and One-Trial Recognition. Experiment 2 investigated the relationship of PL and SL in a second way (the first being through the psychophysical assessment, as discussed above). In Experiment 2, on half of the PL training trials, two of the non-target search grid shapes composed a consistent pair, the SL pair. Participants did not recognize the SL pair above chance – they did not show learning of the SL pair – even though they saw the SL pair more frequently than the target pairs (including when the target pairs were shown before the search grids). Participants in Experiment 2 on average saw this pair 113.52 times, which was more times than participants in Experiment 3 with the 7-minute session length saw each pair on average (in half of the 144 grids, or 72 times), so it was not the case that they did not have enough exposures (on average) to learn the SL pair.

Merely having a statistically reliable relationship in the stimuli was not sufficient to cause learning that could be measured on the recognition test. Participants with more trials with the SL pair did require more trials, but not more time to complete training. It is possible that such participants experienced response time facilitation from the presence of the SL pair. This would perhaps count as a global SL effect, but not as SL as defined in this dissertation. Experiment 2 participants showed strong learning of all the trained pairs, and participants in Experiments 1 and 3 and in prior research (e.g. Fiser & Aslin, 2001) have shown learning of statistically reliable but untrained[28] shape pairs, so the fact that participants showed no learning via recognition of the SL pair does not support the hypothesis of two kinds of learning.

What the Relationship Between SL and PL Could Be

To better understand our data and how they relate to the hypotheses of different kinds of learning and of a unified learning process, we have identified five possible relationships between SL and PL. We have structured each possibility by first introducing the hypothesis, then what pattern of data would support that possibility, then discussed how our data fit with that possibility. We have identified the results that most clearly could indicate whether or not our data support each possibility: the correlation between recognition accuracy (SL) and psychophysical assessment accuracy (PL), how increasing or decreasing SL via the experimental design (Experiment 3) impacts the psychophysical assessment performance (PL), and how increasing PL via the experimental design (Experiment 2) impacts recognition performance (SL).

First, the hypothesis of different kinds of learning does not inherently imply that there need be any relationship between SL and PL, in terms of conditions of occurrence. If there were no relationship, then recognition accuracy and assessment accuracy would likely have been uncorrelated. Increasing or decreasing SL through the experiment design would have had no impact on PL (and there would be no PL), and increasing PL through the design would not have influence SL (and there would be no SL). Our data did show correlations and evidence of both SL and PL in Experiments 1 and 3. Only the data of Experiment 2 were consistent with this possibility, so our data across experiments did not support this possibility specifically, or the hypothesis of different kinds of learning generally.

[28] Pairs were untrained in that no task other than to "pay attention" was given.

Second, the strongest interpretation of the hypothesis of a unified learning process would be that SL and PL are the exact same learning process. Under this possibility, recognition accuracy and assessment accuracy would have been perfectly positively correlated. Increasing or decreasing SL would have always increased or decreased PL, and increasing PL would have increased SL. We did not find a correlation in every experiment, and increasing PL in Experiment 2 did not increase SL, though decreasing SL did decrease PL in Experiment 3. Our data fit this possibility better than the first, but the data also did not support the strongest interpretation of the hypothesis of a unified learning process.

A third possibility, related to the second (though perhaps less plausible) is that SL and PL could be related in that they are negatively correlated. This could either be an unusual version of the hypothesis of a unified learning process, or a special case of the hypothesis of different kinds of learning that recognizes that different learning processes could overlap in conditions that facilitate them. Recognition accuracy and assessment accuracy would have been negatively correlated. Increasing or decreasing SL would have decreased or increased PL. Increasing PL would have decreased SL. The correlations we did find were positive and results of Experiment 3 contradict that changing SL should change PL in the opposite direction. However, the results of Experiment 2 were consistent with the idea that increasing PL would decrease SL. Overall, our results did not support this possibility either.

The fourth possibility we considered was whether SL and PL might be weakly related. This could be a weaker version of the hypothesis of a unified learning process, or this could be a version of the hypothesis of different kinds of learning that recognizes that different learning processes could overlap in conditions that facilitate them. Under this possibility, recognition accuracy and assessment accuracy would weakly correlate or sometimes correlate and sometimes not. Increasing or decreasing SL might or might not influence PL, and increasing PL might or might not influence SL. Given the weakness of relationship and ambiguity of the pattern of results of this possibility, more potential results would have fit with it. That we found a correlation in some experiments but not all fits with this possibility. However, our data do show clear, strong, and reliable relationships between SL and PL - when one is changed, it influences the other. This was the best fit so far for our data, but still did not fully explain our results across experiments.

Finally, we considered whether the relationship between SL and PL might be more complex, in which SL and PL do not influence each other symmetrically. Specifically, we considered whether SL and PL might be a unified learning process, but one in which PL is stronger than or even dominant over SL. Or, whether SL and PL might be different learning processes of different strengths that overlap in their facilitating conditions. Under this final possibility, SL and PL would likely have been correlated in conditions that favor SL, but not conditions favoring PL. When SL was increased or decreased, PL would have increased or decreased, but increasing PL would have decreased or even have prevented SL. Across all three experiments, our data were consistent with this possibility: in Experiments 1 and 2 we found no correlation between recognition accuracy and assessment accuracy, but we did find a correlation in Experiment 3. When we decreased SL in Experiment 3, PL was also decreased[29]. When we increased PL in Experiment 2, we found no evidence of SL. Our data are consistent with this final possibility: the Fiser and Aslin (2001) visual SL paradigm lead to both SL and PL, but PL training with the stimuli did not induce SL. There is a clear asymmetry in this visual task.

[29] (We were unsuccessful in our attempt to increase SL in Experiment 3, so we were not able to tell if increasing SL increased PL, but similar levels of SL produced similar levels of PL.)

The fact that we found PL following SL but not the reverse suggests that while PL and SL might be a unified learning process in humans, the relationship between SL and PL is more nuanced than previously understood, or that even we would have predicted. It is possible that PL is a stronger or even more dominant learning process because it is more general than SL. SL operates on the statistically reliable relationships between elements, but PL operates on any pattern or structure that can distinguish different categories or identify new instances of the same category. Additionally, suppressing irrelevant information is a PL effect, so the SL pair in Experiment 2 might have been suppressed, which would explain why we did not show learning of it. Future research should further address the apparent nuance of the relationship between SL and PL, in other visual paradigms and in other modalities.

But Wasn't it All SL (Except Exp. 2)? And Other Objections and Alternative Explanations

Some readers might object to our definition of SL and/or to our interpretation of the psychophysical assessment as capturing PL because it captured encoding changes. To those who object to our definition: this systematic empirical investigation of the relationship between SL and PL was only possible because we clearly distinguished possible (and plausible, given the literature) concepts SL and PL. If I had defined SL in such a way as to include encoding changes, transfer, and/or influences on perception, SL and PL would no longer be distinguishable and empirical comparison would be impossible. Registration of statistical co-occurrences and improvements in the pickup of information are two conceptually separate ideas (and both are valuable in human learning), so it makes sense to inquire about their relationship. The variability in the literature of applications of the label "statistical learning" does not detract from this basic question. If preferred, one could refer to registration of statistical regularities as Process A and improvements in information encoding as Process B, to avoid entanglement with the inconsistencies in the use of these terms in other work.

Still, the potential objection that the assessment showed SL deserves more in-depth discussion. Remember that the psychophysical assessment required participants to identify whether whole pairs had been present or absent in the previously presented array. One could hypothesize a memory-based mechanism that would allow for high performance on this task in both the Familiar and Shuffled conditions: participants saw the array and decided whether or not its contents were familiar to them, then saw the target pair and made the same decision. If their decisions matched (Familiar: if the array seemed familiar and the target was familiar, then this was a match; Shuffled: if the array was unfamiliar and the target was unfamiliar, then this was also a match), they responded that the target was present, and if their decisions did not match, then they responded absent.

This is a familiarity-based explanation for improved encoding: participants in SL paradigms (Experiments 1 and 3) in the Familiar and Shuffled conditions showed higher accuracy and sensitivity and a lower false alarm rate than the Baseline group across exposure durations (except only with the longest exposure duration following a 7 minute session length in Experiment 3). Participants showed better ability to detect the stimulus - better use of the signal - after recording co-occurrences in memory than the ability shown by the Baseline group that had no learning. Encoding changes should not occur under my definition of SL but only under PL, but perhaps the reader has adopted a broader definition.

Regardless, if this hypothesis explained our data, then recognition accuracy should have strongly and positively correlated with psychophysical assessment performance, but such correlations were weak (Experiment 3) if even present (no correlation in Experiment 1).

Additionally, this hypothesis might expect better performance in Familiar than Shuffled because Familiar trials had more familiar pairings, but this was not the case in Experiment 3 for the 35 minute session length - Shuffled was numerically best across exposure durations and showed significantly higher accuracy and sensitivity than the other conditions at 1000ms. Also, all displays in the New Shapes condition fit the description of unfamiliar for both the target arrays and the probe pair. If this strategy were heavily used by participants in our psychophysical task, then one might have expected the bias of responding “match” to dominate the novel shape results. Bias, as evaluated using SDT methods, tended to be negative for all conditions (“absent” responses predominated), and it was generally negative for New Shapes for all familiarization time conditions in Exp. 3. Likewise, if an explanation of feeling of familiarity were guiding our results, false alarms for the New Shapes should have far exceeded false alarms in the other two conditions, which did not occur; notably such an effect might have been expected to be strongest for the 35 min condition in Exp. 3, which shows no hint of such an effect.) For definitional and data-based reasons, we reject this alternative interpretation of our data.

Another possible objection to our interpretation of our data is that performance in the Shuffled condition did not constitute transfer because the targets on target-absent trials were familiar. I acknowledge the justice of this objection. However, the learning phase in all three experiments did not require an explicit judgement of ruling out anything, so even if the reader disagrees that there was *transfer to new stimuli arrangements*, he or she must agree that there was at least *transfer to new tasks* in the psychophysical assessment - and not only in the Shuffled condition. Transfer to new tasks is another argument against SL (under my definition) and for PL. Objecting to transfer does not negate the improvement in sensitivity, which is more definitive evidence for PL following SL familiarization.

SL, PL, and Bayesian Decision Making

In signal detection terms, PL is improvement in sensitivity, and we hypothesized that SL might be a change in bias. Researchers from Helmholtz (1864) through today's Bayesian vision scientists believe that one's past influences what one sees today, and these are priors in the Bayesian thinking. Optimal decision making uses both priors and the current signal. We hypothesized that the SL familiarization might cause learning of new priors which would have been expressed as a bias for patterns seen before in the psychophysical assessment, via false alarming to familiarized pairs that appeared as targets on target-absent trials in Shuffled. Instead of an increased false alarm rate in Shuffled, we found decreased false alarming in both Familiar and Shuffled relative to New Shapes and relative to the Baseline group. Participants did not show illusory recognition in false alarming based on their prior experience; instead, their learning showed as improved encoding. This is a clear advance, in that SL did not impede encoding of the signal, but instead participants showed improved use of the signal (PL) following exposure to co-occurrences of shapes.

What was Learned: Shapes, Pairs, or Both?

Statistical learning predicted learning of the shapes and their co-occurrence relationships. Perceptual learning predicted learning of reliable structure: shapes and/or pairs. Learning pairs would have been most efficient for encoding arrays and responding to target pairs in the psychophysical assessment, but quickly encoding shapes would have been sufficient because there were separate sets of targets for target-present trials and target-absent trials, and different shapes were paired in each set. Survey results suggest that at least some participants learned shapes and some learned pairs, and some learned both. Participants who reported naming the shapes - a linguistic coding strategy - showed higher assessment accuracy across experiments than those who did not report using this strategy. Naming may have allowed participants to more quickly encode and/or better learn the shapes. Participants who reported noticing the pairs or some consequence of noticing the pairs also showed higher assessment accuracy across experiments than those who did not report noticing anything about the pair structure. Because both learning process could explain learning of shapes and learning of pairs, noticing the pairs and use of the linguistic coding can speak to what individual participants learned, but not what process by which they learned.

Explaining Findings in the SL Literature

The human vision research community disagrees about the relationship between SL and PL, with some believing that SL and PL are separate learning processes and others while others believe SL and PL are a unified learning process. Part of the confusion and disagreement may stem from the way that recent research has challenged assumptions of SL. If the relationship between SL and PL is complex and PL occurs incidentally with SL as indicated by our data, then the possibility of PL having also occurred in other researchers' SL experiments would help explain recent research demonstrating that SL assumptions under my definition (and other definitions like mine) do not hold.

For example, PL is known to be task-dependent in that perception is optimized for the task, and such changes are characterized by discovery and fluency effects. PL discovery effects include the suppression of irrelevant information (Kellman, 2002; Kellman & Garrigan, 2009). This explains why Turk-Browne and colleagues (2005) found no learning of stimuli in the unattended color: the unattended color was not task-relevant.

The influence of Gestalt grouping principles in SL can be explained by the hypothesis of one kind of learning. Several Gestalt principles, including connectedness (Baker, Olson, &

Behrmann, 2004), similarity (Glicksohn & Cohen, 2011), and common fate (Fiser, Scholl, & Aslin, 2007), have been shown to influence what can be learned in SL paradigms. SL assumes that only the statistics of elements drive learning. In contrast, PL assumes that the learning is perceptual, so any kind of perceptual information could be involved in learning.

Concluding that the hypothesis of a unified learning process is correct also explains SL chunking research. Chunking has been observed behaviorally in SL studies with stimuli similar to those used in this dissertation (e.g. Fiser & Aslin, 2005; Lu & Lee, 2013). Chunking models have also been found to fit SL paradigm data better than transition probability (TP) models (e.g. Orbán et al., 2008; Slone & Johnson, 2018). In chunking of stimuli, relationships between elements within a chunk are ignored because the elements are treated as a single unit. At least some effects of chunking comprise PL, in that speed of perceptual encoding can be dramatically improved (e.g., Goldstone, 2000; Kellman & Garrigan, 2008).

The hypothesis of a unified learning process also accounts for transfer. In this dissertation, participants in an SL paradigm (Fiser & Aslin, 2001) transferred learning from learned pairs (Familiar) to the same elements shuffled into new pairs (Shuffled). In prior research (e.g. Otsuka et al., 2013; Turk-Browne & Scholl, 2009), participants have shown transfer from forward sequences of elements to backwards and vice versa and from sequences to simultaneously presented elements and vice versa, as well as transfer from line drawings to words. Transfer is a hallmark of PL.

Explicit and/or Implicit. Finally, the hypothesis of a unified learning process addresses findings that SL can be explicit as well as implicit. Recent research has investigated the assumption that SL is implicit. Kim and colleagues (2009) compared a matching test to a reaction time behavioral measure and concluded that SL is implicit because the behavioral measure better captured learning. Bertels and colleagues (2012) concluded that SL is at least partially explicit for at least some participants by carefully analyzing a completion task and confidence rating in combination with a response time-based behavioral task. No strong claim has been made that PL is always implicit or explicit. At least some PL clearly occurs without conscious awareness (Mettler & Kellman, 2006; Watanabe, Nañez, & Sasaki, 2001), and it is commonplace for those with PL expertise to be unable to describe exactly what information they are using (Gibson, 1969; Kellman & Garrgan, 2008). Models of basic visual PL that emphasize selective reweighting of analyzers in early visual cortex surely involve processes that occur outside of awareness (Petrov, Dosher & Lu, 2005). On the other hand, some PL, especially in relatively simple stimulus situations, may be explicit in that observers can report the information that facilitates categorization or determines some classification.

In this dissertation, many participants directly reported noticing the pair structure, and many also reported on consequences of the pair structure. These participants' knowledge of the pairs was (at least partially) explicit. In Experiment 3 which used an SL paradigm (Fiser & Aslin, 2001), noticing the pairs correlated with higher recognition accuracy and with higher assessment accuracy. Noticing was also positively correlated with assessment accuracy across experiments. Explicit knowledge was associated with higher performance on both the recognition test (testing for SL), and the psychophysical assessment test (testing primarily for PL) on which participants demonstrated PL.

Both sets of authors (Bertels et al., 2012; Kim et al., 2009) agree that recognition tests are ambiguous as to whether they capture implicit or explicit data, and that other kinds of assessments could more clearly capture explicit (e.g. matching test, completion test, confidence ratings) or implicit (i.e. various behavioral tasks) aspects of SL. In this dissertation, the behavioral psychophysical assessment captured learning more frequently and with more nuance than the recognition test. The survey also captured aspects of learning missed by the recognition test, so future research should favor behavioral and survey measures over recognition to capture more complete and richer data – both implicit and explicit.

Global SL Might Be Different Than PL. In prior research, SL has been observed as response time facilitation, when SL in the context of the study was learning of a global statistic (e.g. Cosman & Vecera, 2014). Defining SL as any learnable statistic, including global statistics, broadens the definition of SL from being primarily about element correlations, and broadens it beyond this dissertation's definition of SL. The SL pair in Experiment 2 of the dissertation did not show learning on the recognition test – the typical test of SL.

Limitations and Future Directions

Psychophysical Assessment. The psychophysical assessment was designed to discriminate SL and PL through testing for PL effects, particularly transfer and improved sensitivity. Participants' performance in the Shuffled condition across dependent measures has been interpreted as transfer. However, success in the Shuffled assessment required not only correctly identifying shuffled pairs that were present in the grid on all target-present trials, but also rejecting familiar pairs (which were always absent) on all target-absent trials, so learning of familiar pairs could also have explained performance on Shuffled (as explained in more detail above). Future research could test for transfer following SL familiarization in ways that do not confound transfer to another task with possible transfer to other stimuli or with performance on learned stimuli.

SL Pair in Experiment 2. There are several possible explanations for the difference between Experiment 2 findings of no learning of the SL pair and the findings of Experiments 1 and 3 and prior research (e.g. Fiser & Aslin, 2001; Kim et al., 2009) of learning of statistically reliable relationships. One possible explanation is that attention to the target pairs blocked learning of the SL pair. This explanation derives from the hypothesis of a unified learning process: because the targets were the focus of attention and learning (and the SL pair was not), task-relevant stimuli only were learned.

Another possible explanation is that experimental design and analyses did not allow for detecting learning that might have been present: response time facilitation. Because we did not intend to compare participants on percentage of SL trials completed and only had relevant data because of a programming error, the high-percentage group was underpowered. Additionally, accuracy on PL learning trials was at ceiling, so this could also have prevented differences from being measured. Future researchers could manipulate the percentage at which global statistics occur to gain clarity on whether response time facilitation due to global SL can be measured in our or another experimental design.

Another limitation is that we focused on SL (as we and prior research have defined) as recording co-occurences of elements. We did not address global SL because it was not part of SL as we defined it, so our data cannot speak to whether or not global SL is part of a unified learning process with SL (under our definition) and PL. Future research should address how learning of a global statistic relates to SL and to PL, especially in the context of the hypothesis of a unified learning process.

Other. We studied the relationship of visual SL and PL in a particular visual SL paradigm testing for transfer in a particular way with university undergraduates, so it would be useful to replicate our findings with other tests, paradigms, populations, and modalities. Recent SL research have shown chunking (e.g. Orbán et al., 2008) and transfer (e.g. Turk-Browne &

Scholl, 2009) in visual SL, so these paradigms could be adapted to investigate SL and PL together by testing chunks and transfer psychophysically. SL and PL have both been studied with children and infants as well as adults, and children's memories (relevant to SL) work differently before they can speak, so it would be interesting to see if SL and PL are a unified learning process for very young learners as well. Because PL is known to depend on task (and thus depends on modality) whereas SL is assumed to be modality-independent and both have been studied in various modalities as well as in multi-modal studies, it would be informative to investigate whether SL and PL are a unified learning process in other modalities as well as cross-modally.

Finally, we found evidence that PL and SL are a unified learning process in (adult) human vision, in such as way that SL causes PL but not the reverse. As mentioned previously, future research should further address this nuance. Additionally, other authors have suggested that SL might be related to implicit learning (e.g. Perruchet & Pacton, 2006) and associative learning (Fiser, 2009), and that implicit learning and associative learning might be linked to PL (e.g. implicit: Fahle & Poggio, 2002; Jiang & chun, 2001; Reber, 1967; associative: Hall, 1991; Law & Gold, 2009; McLaren & Mackintosh, 2000). It would be useful for future researchers to also empirically study the interrelationships (or lack thereof) of all of these kinds of learning, and perhaps others, to better understand which learning processes are unified with or distinct from SL and PL, and to what extent.

Conclusion and Broader Impacts

Statistical learning and perceptual learning both seek to explain learning from experience, and how regularities in the world allow humans to do so. Across three experiments, participants showed transfer and improved sensitivity, hallmarks of PL. Experiments 1 and 3 showed familiarity, but Experiment 2 showed no SL. Familiarity correlated with psychophysical performance in Experiment 3 but not Experiment 1. So, several results are consistent with SL and PL being part of a unified learning process, or at least occurring under overlapping conditions, but there may be an asymmetry: PL may occur more under conditions designed for SL, but SL may be less likely to occur during focused PL tasks. This hypothesis of a complex, asymmetric relationship can explain recent findings in visual SL and help unify two largely distinct literatures and research communities. This will allow researchers and those interested in structuring learning for themselves or others to better understand both SL and PL, and to glean insights from both literatures, including that 1) active (PL) training produces stronger learning in a similar amount of time as passive learning, 2) learning is flexible, so that our perception can improve for any and all regularities in domains as diverse as music, aviation, mathematics, baking, and medicine, and 3) learning is for transfer, to prepare to successfully encounter both new instances of learned categories and instances of new, related categories in the same or similar domains.

Appendix A: Survey Questions

“What did you notice during the experiment?”

“Did you have any strategies? If you had any strategies, please describe them.”

“Did you notice any patterns? If you noticed patterns, please describe them.”

“Was anything unclear or confusing? If anything was unclear or confusing, please describe it.”

“Do you have any other feedback about the experiment? If you do, please write it here.”

Appendix B: Experiment 2 Results (Including Those with 75% SL Pair Trials)

Results

Perceptual Learning Condition

Participants needed 24.69 minutes (*SE* = 0.38) on average to complete an average of 246.13 trials (*SE* = 7.26)[30]. Mastery level was, on average, 24.05 levels (*SE* = 0.12). Training time and number of trials were marginally positively correlated, *r*(77) = .21, *p* = .06. 40 participants advanced from one trial to the next by clicking the mouse; 38 by pressing the spacebar. Participants who clicked had more trials (*M* = 264.30, *SE* = 13.26) than those who used the spacebar (*M* = 227.00, *SE* = 20.53), *t*(76) = 2.67, *p* = .009. Click advancement also required more time (*M* = 25.82 minutes, *SE* = 3.35) than spacebar advancement (*M* = 23.49, *SE* = 2.99), *t*(76) = 3.23, *p* = .002.

PL Familiarity Test

Participants' accuracy did not differ from chance (*M* = 0.43, *SE* = 0.06) on recognition of the statistical learning pair (see Figure 47) in the familiarity test, *t*(76) = -1.23, *p*=.21. 44 (of 77) participants did not indicate that the statistical learning pair seemed more familiar to them than the foil pair in the single 2AFC trial.

[30] Independent-samples *t*-tests showed that the 8 participants with a higher percentage of SL pair trials did not differ from the other participants in recognition accuracy (*p* = .28), average assessment accuracy (*p* = .23), or time to complete training (*p* = .79). However, these 8 participants (*M* = 413.25, *SE* = 22.70) required significantly more trials to complete assessments than those with 50% SL pair trials (*M* = 227.03, *SE* = 2.84), *t*(76) = 8.14, *p* < .001.

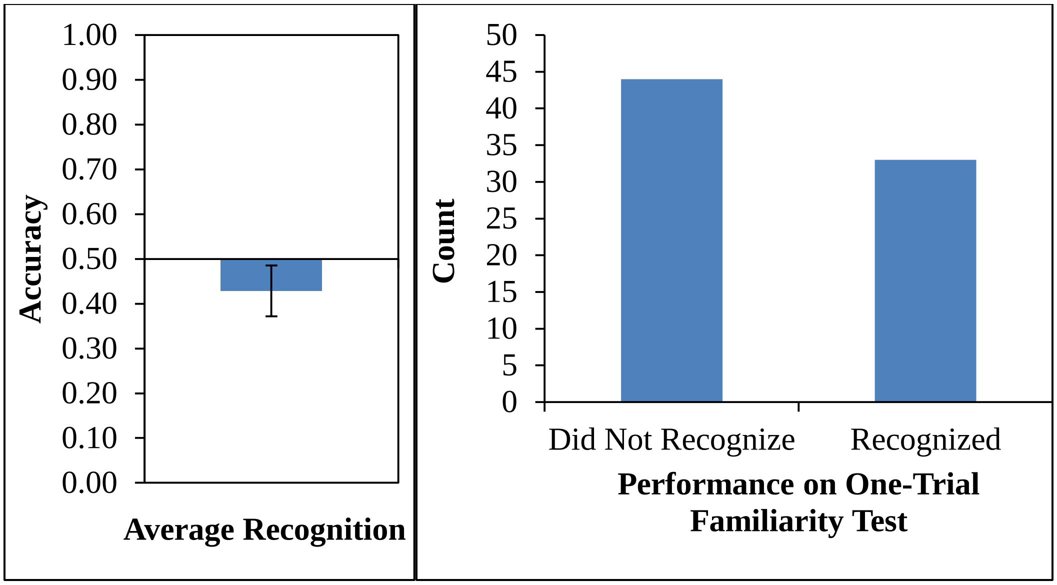

Figure 47. Average accuracy on the one-trail recognition test (left) and histogram of accuracy (right). Average recognition did not differ from chance (.5). Error bar (left) shows standard error of the mean.

An ANOVA of Condition on recognition accuracy showed that conditions did not differ in recognition (*p* = .47). Participants in all conditions recognized the statistical learning pair equally poorly. This is unsurprising because all participants received the same recognition test (relative to their familiarization shape set), but it does indicate that there were not significant differences across participants in different conditions by chance.

However, participants who did recognize the statistical learning pair might have performed differently from those who did not on the assessment. Perhaps those who recognized the SL pair learned targets less well from looking at distractors, or, conversely, perhaps they noticed the SL pair and were able to eliminate it more quickly from their searches and more efficiently find targets than other participants. To investigate this possibility, I conducted a Pearson's correlation of average psychophysical assessment accuracy and familiarity test accuracy. I found no significant correlation of familiarity and assessment accuracy, *r*(76) = 0.06, *p* = .58.

An ANOVA of Condition on recognition accuracy showed that conditions did not differ in recognition (*p* = .47). Participants in all condition recognized the statistical learning pair equally poorly. This is unsurprising because all participants received the same recognition test (relative to their familiarization shape set), but it does indicate that there were not significant differences across participants in different conditions by chance.

However, participants who did recognize the statistical learning pair might have performed differently from those who did not on the assessment. Perhaps those who recognized the SL pair learned targets less well from looking at distractors, or, conversely, perhaps they noticed the SL pair and were able to eliminate it more quickly from their searches and more efficiently find targets than other participants. To investigate this possibility, I conducted a Pearson's correlation of average psychophysical assessment accuracy and familiarity test accuracy. I found no significant correlation of familiarity and assessment accuracy, *r*(76) = 0.06, *p* = .58.

PL Training and Familiarity. I conducted analyses to investigate whether the PL training influenced performance on the familiarity test. Familiarity test accuracy did not correlate with the number of PL training trials completed, *r*(76) = -0.22, *p* = .06. Participants that completed more trials did worse on the familiarity test. PL training time did not correlate with familiarity test accuracy (*p* = .23). Those who advanced from one trial to the next by clicking the mouse (*n*=39, *M* = 0.33, *SE* = 0.08) were marginally less accurate than those that advanced via pressing the spacebar (*n*=38, *M* = 0.53, *SE* = 0.08) on familiarity, *t*(75) = -1.72, *p* = .09.

Psychophysical Assessment: Main Analyses

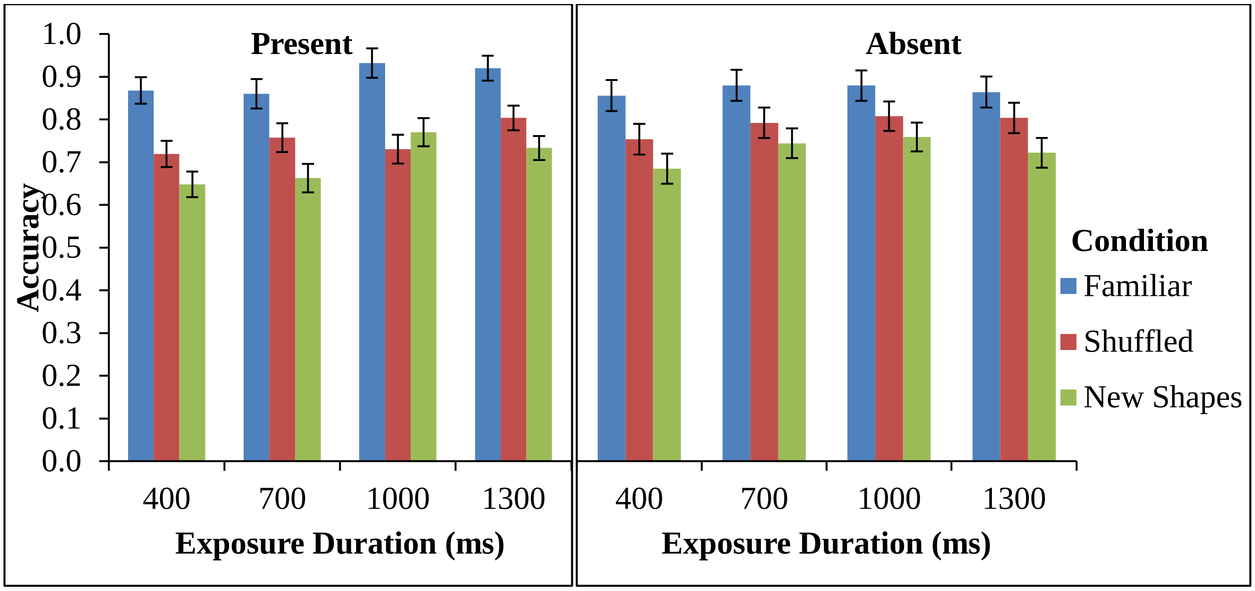

Figure 48. Condition by Exposure Duration by Target Presence on accuracy. Error bars indicate standard error of the mean.

Accuracy. Figure 48 showed the effects of Condition, Exposure Duration, and Target Presence on accuracy. It appeared that Familiar showed the highest accuracy across exposure durations and levels of Target Presence, and that Shuffled showed the next highest accuracy, also across exposure durations and levels of Target Presence. To test the apparent effects, I conducted an ANOVA of Condition by Exposure Duration by Target Presence on accuracy, which revealed a main effect of Condition, $F(2, 75) = 18.92$, $p < .001$, *partial-eta-squared* = 0.33, a main effect of Exposure Duration, $F(3, 225) = 7.15$, $p < .001$, *partial-eta-squared* = 0.09, a marginal interaction of Exposure Duration and TargetPresence, $F(3,225) = 2.21$, $p = .09$, *partial-eta-squared* = 0.03, and no other effects (all *p*'s > .28).

Custom hypothesis tests were used to investigate the significant main effects. For Condition, I conducted all pairwise comparisons. The Familiar condition (*n* = 25[31], *M* = 0.88, *SE* = 0.02) was significantly more accurate than the Shuffled condition (*n* = 26, *M* = 0.77, *SE* = 0.02), *F*(1,75) = 16.09, *p* < .001, *partial-eta-squared* = 0.18. Similarly, Familiar was significantly more accurate than New Shapes (*n* = 27, *M* = 0.72, *SE* = 0.02), *F*(1,75) = 36.75, *p* < .001, *partial-eta-squared* = 0.33. Shuffled was significantly more accurate than New Shapes, *F*(1,75) = 4.14, *p* = .05, *partial-eta-squared* = 0.05. Familiar showed the highest accuracy, but learning also transferred to Shuffled.

For Exposure Duration, I compared 1000ms (*M* = 0.81, *SE* = 0.01) and 1300ms (*M* = 0.81, *SE* = 0.01), which did not differ (*p* = .70). I compared 400ms (*M* = 0.76, *SE* = 0.01) to the longer exposure durations, and 400ms was significantly less accurate than 1000ms and 1300ms combined, *F*(1,75) = 26.95, *p* < .001, *partial-eta-squared* = 0.26. I compared 700ms (*M* = 0.78, *SE* = 0.02) to 1000ms and 1300ms, and 700ms was marginally less significant than the longest exposure durations combined, *F*(1,75) = 4.71, *p* = .03, *partial-eta-squared* = 0.06. I compared the short exposure durations, and 400ms did not differ from 700ms (*p* = .07), *F*(1,75) = 3.32, *p* = .07, *partial-eta-squared* = 0.04.

For the interaction of Exposure Duration and TargetPresence, I broke the data on TargetPresence and tested the simple effect of Exposure Duration for each level of TargetPresence, and followed up significant simple effects with additional tests. When the target was Present, there was a significant simple effect of Exposure Duration, *F*(2,75) = 17.18, *p* < .001, *partial-eta-squared* = 0.31. I compared 400ms (*M* = 0.75, *SE* = 0.02) to 700ms (*M* = 0.76,

[31] The 14 original and 11 added participants did not differ on average recognition accuracy (*p* = .18), assessment accuracy (*p* = .39), number of training trials (*p* = .95) or training time (*p* = .59).

SE = 0.02), which did not differ (*p* = .69). Similarly, I compared 1000ms (*M* = 0.81, *SE* = 0.02) to 1300ms (*M* = 0.82, *SE* = 0.02), which also did not differ (*p* = .44). However, when I compared the short exposure durations to the long exposure durations, 400ms and 700ms were significantly less accurate than 1000ms and 1300ms, *F*(1,75) = 18.92, *p* < .001, *partial-eta-squared* = 0.20. For Absent, there was also a significant simple effect of Condition, *F*(2,75) = 6.02, *p* = .004, *partial-eta-squared* = 0.14. I compared 400ms (*M* = 0.77, *SE* = 0.02) to 1300ms (*M* = 0.80, *SE* = 0.02), which did not differ (*p* = .12). Then I compared 700ms (*M* = 0.81, *SE* = 0.02) to 1000ms (*M* = 0.82, *SE* = 0.02), which did not differ (*p* = .59). However, when I compared the shortest and longest exposure durations to the middle ones, 400ms and 1300ms was marginally less accurate than 700ms and 1000ms, *F*(1,75) = 5.32, *p* = .02, *partial-eta-squared* = 0.07. The interaction of Exposure Duration and Target Absence was driven by different simple effects of Exposure Duration for Absent than Present.

False Alarm Rate. Figure 49 showed effects of Condition and Exposure Duration on false alarm rate. It appeared that Familiar had the lowest false alarm rate across exposure durations, and that Shuffled had a lower false alarm rate than New Shapes. To test these apparent effects, I conducted an ANOVA of Condition by Exposure Duration on false alarm rate, which revealed a main effect of Condition, *F*(2,75) = 6.02, *p* = .004, *partial-eta-squared* = 0.14, and a marginal main effect of Exposure Duration, *F*(3,225) = 2.61, *p* = .05, *partial-eta-squared* = 0.03, and no interaction (*p* = .93).

Custom hypothesis tests in ANOVA of all pairwise comparisons following up on the main effect of Condition revealed that Familiar (*M* = 0.13, *SE* = 0.03) had a significantly lower false alarm rate than New Shapes (*M* = 0.27, *SE* = 0.03), *F*(1,75) = 12.00, *p* = .001, *partial-eta-squared* = 0.14. Shuffled (*M* = 0.21, *SE* = 0.03) had a marginally higher false alarm rate than

Familiar, *F*(1,75) = 3.78, *p* = .06, *partial-eta-squared* = 0.05. Shuffled did not differ from New Shapes (*p* = .13). Learning was only demonstrated for Familiar.

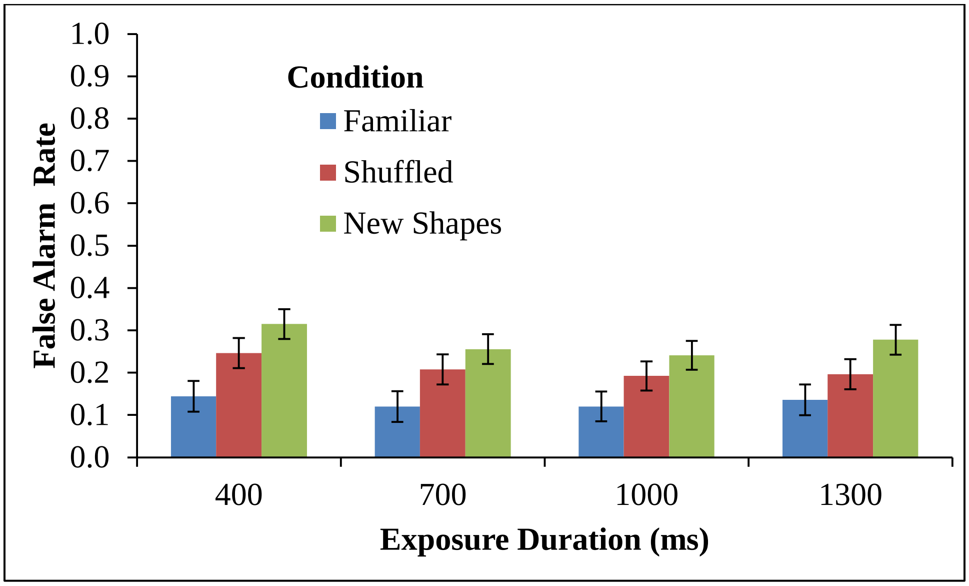

Figure 49. Effects of Condition and Exposure Duration on false alarm rate. Error bars indicate standard error of the mean.

For the marginal effect of Exposure Duration, custom hypothesis tests of pairwise comparisons revealed that 400ms (*M* = 0.24, *SE* = 0.02) had a marginally higher false alarm rate than 1000ms (*M* = 0.18, *SE* = 0.02), *F*(1,75) = 7.72, *p* = .007, *partial-eta-squared* = 0.09. No other comparisons were significant (all *p*'s > .03).

Sensitivity. Figure 50 showed effects of Condition and Exposure Duration on sensitivity. It appeared that Familiar was most sensitive across exposure durations, and Shuffled was more sensitive than New Shapes. To test these apparent effects, I conducted an ANOVA of Condition by Duration on sensitivity (d'). This revealed a main effect of Condition *F*(2,75) = 19.10, *p* <

.001, *partial-eta-squared* = 0.34, and a main effect of Exposure Duration *F*(3,225) = 7.76, *p* < .001, *partial-eta-squared* = 0.09, and no interaction (*p* = .26).

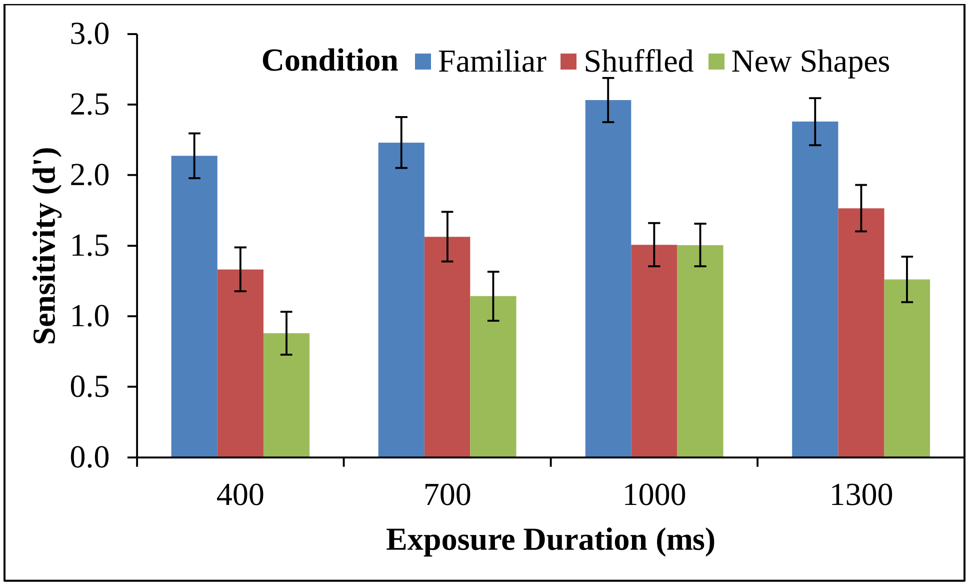

Figure 50. Effects of Condition and Exposure Duration on sensitivity. Error bars indicate standard error of the mean.

I followed up on significant effects via custom hypothesis tests in ANOVA. For the main effect of Condition, I conducted all pairwise comparisons. I found that Familiar (*M* = 2.32, *SE* = 0.13) had significantly higher sensitivity than Shuffled (*M* = 1.54, *SE* = 0.13), *F*(1,75) = 36.72, *p* < .001, *partial-eta-squared* = 0.19. Familiar also had significantly higher sensitivity than New Shapes (*M* = 1.20, *SE* = 0.13), *F*(1,75) = 36.72, *p* < .001, *partial-eta-squared* = 0.33. Shuffled showed marginally higher sensitivity than New Shapes, *F*(1,75) = 3.54, *p* = .06, *partial-eta-squared* = 0.05. Learning was strongest in the trained condition, but did transfer to Shuffled as well.

I conducted all pairwise comparisons using custom hypothesis tests in ANOVA to follow up on the main effect of Exposure Duration. 400ms (*M* = 1.45, *SE* = 0.09) was less sensitive than 1000ms (*M* =1.85, *SE* = 0.09), *F*(1,75) = 25.78, *p* < .001, *partial-eta-squared* = 0.26. 400ms was also less sensitive than 1300ms (*M* =1.80, *SE* = 0.10), *F*(1,75) = 18.18, *p* < .001, *partial-eta-squared* = 0.20. No other comparisons were significant (all *p*'s > .03).

Appendix C: Experiment 3 Results (Excluding Late Participants)

Results

Recognition

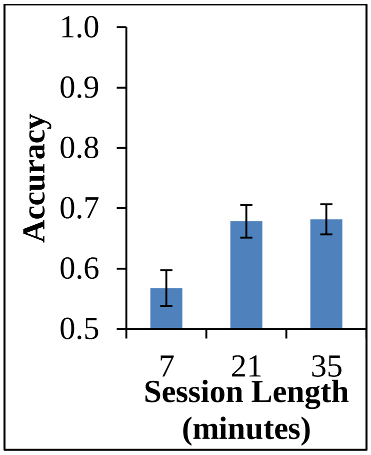

Figure 51. Experiment 3 Recognition accuracy by Session Length against chance. Error bars indicate standard error of the mean.

Figure 51 shows that participants demonstrated accuracy significantly higher than chance (*M* = 0.64, *SE* = 0.02) on recognition of pairs in familiarization in the recognition test, *t*(192) = 8.38, *p* < .001, Cohen's *d* = 2.76. As apparent in the distribution of accuracy for participants (in Figure 52), many participants did not pass the recognition test. Participants were divided into two groups by their accuracy on recognition: "Recognizers" (*n* = 116) scored above 50% and "Nonrecognizers" (*n* = 77) scored at or below 50%. An independent-samples *t*-test of recognition group on recognition accuracy revealed that Recognizers (*M* = 0.79, *SE* = 0.01) had significantly higher accuracy than Nonrecognizers (*M* = 0.40, *SE* = 0.02), *t*(191) = 20.40, *p* < .001, Cohen's *d* = 1.47.

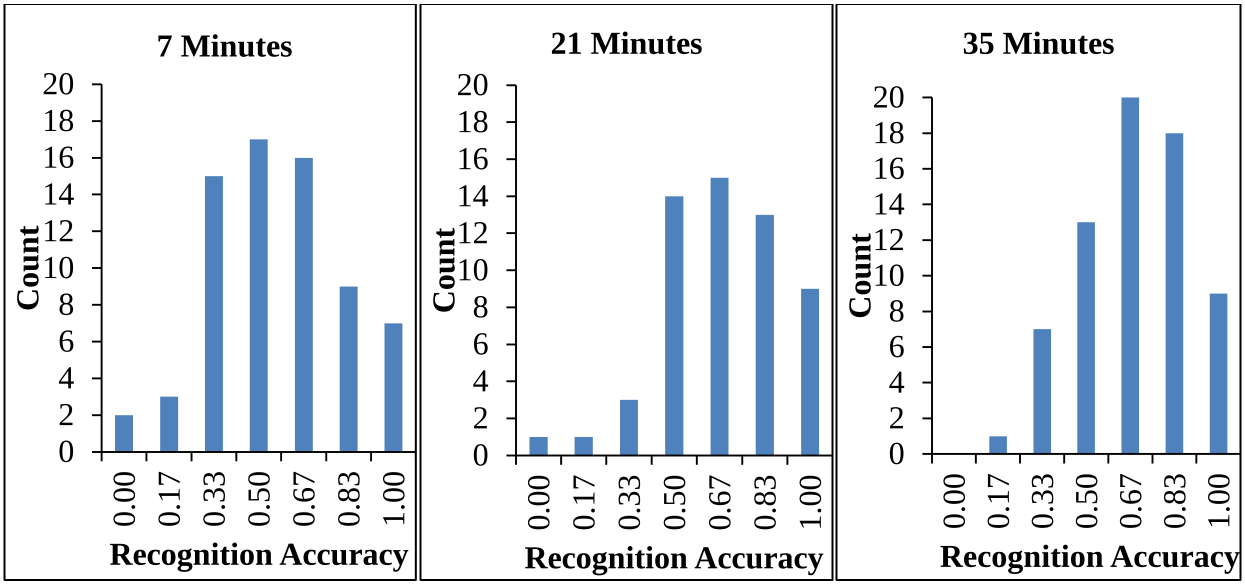

Figure 52. Experiment 3 frequency histograms of recognition accuracy by Session Length: 7 minutes (left panel), 21 minutes (center panel), and 35minutes (right panel). Error bars indicate standard error of the mean.

An ANOVA of Condition and Session Length on recognition accuracy showed no main effect of Condition (*p* = .85), no interaction of Condition and Session Length (*p* = .49), and a main effect of Session Length, *F*(2,184) = 4.91, *p* = .008, *partial-eta-squared* = 0.05. Custom hypothesis tests in ANOVA revealed that 21 minutes (*M* = 0.67, *SE* = 0.03) and 35 minutes (*M* = 0.68, *SE* = 0.03) did not differ (*p* = .71). However, participants were significantly less accurate on the recognition test with 7 minutes (*M* = 0.57, *SE* = 0.03) of familiarization than when they had a longer session length, *F*(1,184) = 9.39, *p* = .003, *partial-eta-squared* = 0.05. This is consistent with piloting (and why 21 minutes of familiarization was used for Experiment 1). However, participants with the 7-minute session length still showed higher recognition than chance, *t*(68) = 2.44, *p* = .02, Cohen's *d* = 0.29. Decreasing the session length decreased recognition.

Participants in all three versions of the psychophysical assessment pairs seen in familiarization equally well. This is unsurprising because all participants received the same recognition test (relative to their familiarization shape set), but it does indicate that there were not significant differences across participants in different assessment versions by chance. It was also important that Session Length did not interact with (assessment) Condition for recognition.

Table 2.

Session Length and Recognition Group on actual and expected counts.

	Nonrecognizers		Recognizers		
Session Length	Actual	Expected	Actual	Expected	Total
7 Minutes	37	(27.5)	32	(41.5)	69
21 Minutes	19	(22.5)	37	(33.7)	56
35 Minutes	21	(27.1)	47	(40.9)	68
Total	77		116		193

I investigated the relationship of Session Length and Recognition Group using a chi-squared test. There was a significant association between Session Length and Recognition Group, $\chi^2(2) = 8.59$, *p* = .01. Looking at Table 2, it appears that the association of Session Length and Recognition Group was due to a higher percentage of participants in the 7 minutes group failing the recognition test than in the other session lengths.

Psychophysical Assessment: Main Analyses

Accuracy. Figure 53 showed effects of Condition, Session Length, and Exposure Duration on accuracy. It appeared that for 7 minutes there was only learning at the longest exposure duration, and that for 21 minutes Familiar had the highest accuracy across exposure durations but this “flipped” to Shuffled having the highest accuracy across exposure durations for 35 minutes. To test the apparent effects, I conducted a four-way mixed ANOVA of Condition by Session Length by Exposure Duration by Target Presence on accuracy. I found significant main effects of Condition, *F*(2,184) = 4.90, *p* = .008, *partial-eta-squared* = 0.05, and Exposure Duration, *F*(3,552) = 3.14, *p* = .03, *partial-eta-squared* = 0.02. There was also a significant main effect of Target Presence, such that participants were more accurate when the target was Absent (*M* = 0.78, *SE* = 0.01) than when the target was Present (*M* = 0.71, *SE* = 0.01), *F*(1,184) = 18.18, *p* < .001, *partial-eta-squared* = 0.09. I also found three significant interactions: Condition interacted marginally with Exposure Duration *F*(6,552) = 2.01, *p* = .06, *partial-eta-squared* = 0.02; the effect of Condition marginally depended upon the combined effects of Session Length and Exposure Duration, *F*(12,552) = 1.60, *p* = .09, *partial-eta-squared* = 0.03; and Exposure Duration interacted with Target Presence, *F*(3,552) = 5.72, *p* = .001, *partial-eta-squared* = 0.03. All other effects were non-significant (all *p*’s > .12).

Custom hypothesis tests in ANOVA were used to investigate the main effects. For the main effect of Condition, I compared Familiar (*M* = 0.76, *SE* = 0.01) and Shuffled (*M* = 0.77, *SE* = 0.01), which did not differ in accuracy (*p* = .50). I compared New Shapes (*M* = 0.71, *SE* = 0.02) to Familiar and Shuffled, and found that New Shapes showed significantly lower accuracy than Familiar and Shuffled combined, *F*(1,184) = 9.34, *p* = .003, *partial-eta-squared* = 0.05. The learning transferred from Familiar to Shuffled.

For the main effect of Exposure Duration, I compared the shorter exposure durations and found that 400ms (*M* = 0.73, *SE* = 0.01) and 700ms (*M* = 0.74, *SE* = 0.01) did not show different accuracy (*p* = .53). Similarly, I compared the longer exposure durations, and found that 1000ms (*M* = 0.76, *SE* = 0.01) and 1300ms (*M* = 0.76, *SE* = 0.01), did not differ in accuracy (*p* = .97). I compared 400ms and 700ms to 1000ms and 1300ms, and found that the shorter exposure durations combined showed marginally lower accuracy than the longer exposure durations combined, *F*(1,184) = 8.47, *p* = .002, *partial-eta-squared* = 0.05.

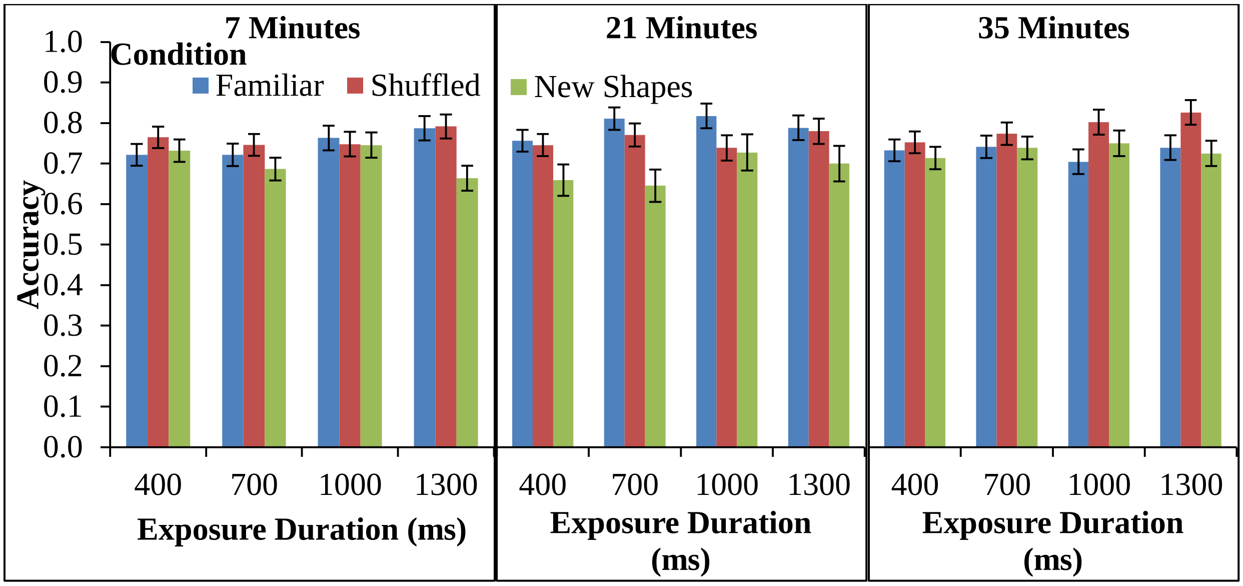

Figure 53. Condition by Session Length by Exposure Duration on accuracy (collapsed across Target Presence). Bar heights indicate adjusted marginal means and error bars indicate standard error of the mean.

Custom hypothesis tests were also used to examine the interaction of Condition and Exposure Duration, by testing the simple effect of Condition at each Exposure Duration, and following up on any significant simple effects with further custom hypothesis tests. At 400ms, there was a marginal simple effect of Condition, *F*(2,184) = 2.41, *p* = .09, *partial-eta-squared* = 0.03. I compared Familiar (*M* = 0.74, *SE* = 0.02) and Shuffled (*M* = 0.75, *SE* = 0.02), which did not differ in accuracy (*p* = .44). When I compared Familiar and Shuffled to New Shapes (*M* = 0.70, *SE* = 0.02), I found that together they were more accurate than New Shapes, *F*(1,184) = 4.22, *p* = .04, *partial-eta-squared* = 0.02. At 700ms, there was a significant simple effect of Condition, *F*(2,184) = 5.20 *p* = .006, *partial-eta-squared* = 0.05. Again, I compared Familiar (*M* = 0.76, *SE* = 0.02) and Shuffled (*M* = 0.76, *SE* = 0.02), which did not differ (*p* = .81). I compared Familiar and Shuffled to New Shapes (*M* = 0.69, *SE* = 0.02), Familiar and Shuffled showed higher accuracy than New Shapes, *F*(1,184) = 10.33, *p* = .002, *partial-eta-squared* = 0.05. There was also a simple effect of Condition at 1300ms, *F*(1,184) = 7.45, *p* = .001, *partial-eta-squared* = 0.08. I compared Familiar (*M* = 0.77, *SE* = 0.02) and Shuffled (*M* = 0.80, *SE* = 0.02), which did not differ in accuracy (*p* = .27). I compared Familiar and Shuffled to New Shapes (*M* = 0.70, *SE* = 0.02), and found that together they showed higher accuracy than New Shapes, *F*(1,184) = 13.66, *p* < .001, *partial-eta-squared* = 0.07. There was no simple effect of Condition at 1000ms (*p* = .68). The interaction of Condition and Exposure Duration was driven by the three significant simple effects of Condition (at 400ms, 700ms, and 1300ms).

In looking at Figure 52, it appeared that the three-way interaction of Condition, Session Length, and Exposure Duration was due to Familiar showing the highest sensitivity for 21 minutes and Shuffled showing the highest sensitivity for 35 minutes, and to learning for the 7-minute exposure duration being apparent only for the longest exposure duration. To test the interaction, I broke the data on Exposure Duration and tested simple interactions for Condition and Session Length at each exposure duration. There was no simple interaction of Condition and Session Length at 400ms (*p* = .52), 700ms (*p* = .12), or 1300ms (*p* = .34). However, there was a marginal interaction of Condition and Session Length at 1000ms, *F*(4,184) = 2.27, *p* = .08, *partial-eta-squared* = 0.05. I used custom hypothesis tests to investigate the marginal interaction at 1000ms by evaluating simple simple effects of Condition for each session length at 1000ms via custom hypothesis tests in ANOVA and followed up on significant simple simple effects with additional tests. There was no simple simple effect of Condition for 7 minutes (*p* = .91) or 21 minutes (*p* = .15). However, there was a marginal simple simple effect of Condition for 35 minutes of familiarization, *F*(2,184) = 2.55, *p* = .08, *partial-eta-squared* = 0.03. I followed this up with all pairwise comparisons. Familiar (*M* = 0.70, *SE* = 0.03) was less accurate than Shuffled (*M* = 0.80, *SE* = 0.03), *F*(1,184) = 5.08, *p* = .03, *partial-eta-squared* = 0.03. New Shapes (*M* = 0.75, *SE* = 0.03) did not differ in accuracy from Familiar (*p* = .30) or Shuffled (*p* = .24).

I also directly tested the apparent “flip” from Familiar having the numerically highest accuracy across exposure durations for 21 minutes to Shuffled having the numerically highest accuracy across exposure durations for 35 minutes by examining the interaction of Condition and Session Length in an ANOVA of Condition by Session Length (21, 35) by Exposure Duration by Target Presence on accuracy. There was a marginal interaction of Condition and Session Length across exposure durations and levels of Target Presence, *F*(2,118) = 2.75, *p* = .07, *partial-eta-squared* = 0.05. There was a simple effect of Condition at 21 minutes, *F*(2,118) = 3.34, *p* = .04, *partial-eta-squared* = 0.05, and I followed up with all pairwise comparisons. Familiar (*M* = 0.79, *SE* = 0.02) was more accurate than New Shapes (*M* = 0.68, *SE* = 0.03), *F*(1,118) = 6.68, *p* = .01, *partial-eta-squared* = 0.05. Shuffled (*M* = 0.76, *SE* = 0.02) was marginally more accurate than New Shapes, *F*(1,118) = 3.09, *p* = .08, *partial-eta-squared* = 0.03. Familiar and Shuffled did not differ (*p* = .32). There was no reliable simple effect of Condition at 35 minutes (*p* = .16).

I also directly tested the apparent effect of the learning for the 7-minute session length being restricted to the longest exposure duration. I tested the simple interaction of Condition and Exposure Duration for 7 minutes of familiarization[32]. For 7 minutes of Familiarization, there was a simple interaction of Condition and Exposure Duration on accuracy, *F*(6,198) = 2.94, *p* = .009, *partial-eta-squared* = 0.08. I followed up with custom hypothesis tests in ANOVA of simple simple effects of Condition at each exposure duration, and followed up significant simple simple effects with further custom hypothesis tests. There was a significant simple simple effect of Condition at 1300ms, *F*(2,66) = 5.59, *p* = .006, *partial-eta-squared* = 0.14. I compared Familiar (*M* = 0.79, *SE* = 0.03) and Shuffled (*M* = 0.79, *SE* = 0.03), and found that they did not differ in accuracy (*p* = .91). Then I compared Familiar and Shuffled together to New Shapes (*M* = 0.66, *SE* = 0.03), and found that the combination showed higher accuracy than New Shapes, *F*(1,66) = 11.15, *p* = .001, *partial-eta-squared* = 0.15. There was not simple simple effect for 400ms, 700ms, or 1000ms (all *p*'s > .23). The three-way interaction was driven by the highest performance for Shuffled at 35 minutes and 1000ms, by the transfer from Familiar to Shuffled at 21 minutes, and by the learning only at 1300ms for 7 minutes.

Custom hypothesis tests were also used to investigate the interaction of Exposure Duration and Target Presence, by testing the simple effect of Target Presence at each exposure duration. There was a significant simple effect of Target Presence at 400ms, such that participants were more accurate on trials when the target was Absent (*M* = 0.78, *SE* = 0.01) than when the target was Present (*M* = 0.68, *SE* = 0.01), *F*(1,184) = 24.04, *p* < .001, *partial-eta-*

[32] There was no simple interaction of Condition and Exposure Duration at 21minutes (*p* = .22) or 35 minutes (*p* = .20).

squared = 0.12. At 700ms, participants were also more accurate for Absent (*M* = 0.77, *SE* = 0.01) than Present (*M* = 0.70, *SE* = 0.02), *F*(1,184) = 13.65, *p* < .001, *partial-eta-squared* = 0.07. 1000ms showed the same pattern of higher accuracy for Absent (*M* = 0.79, *SE* = 0.01) than Present (*M* = 0.72, *SE* = 0.02), *F*(1,184) = 13.89, *p* < .001, *partial-eta-squared* = 0.07. In contrast, there was no simple effect of Target Presence for 1300ms (*p* = .33). The interaction of Exposure Duration and Target Presence was driven by the three significant simple effects.

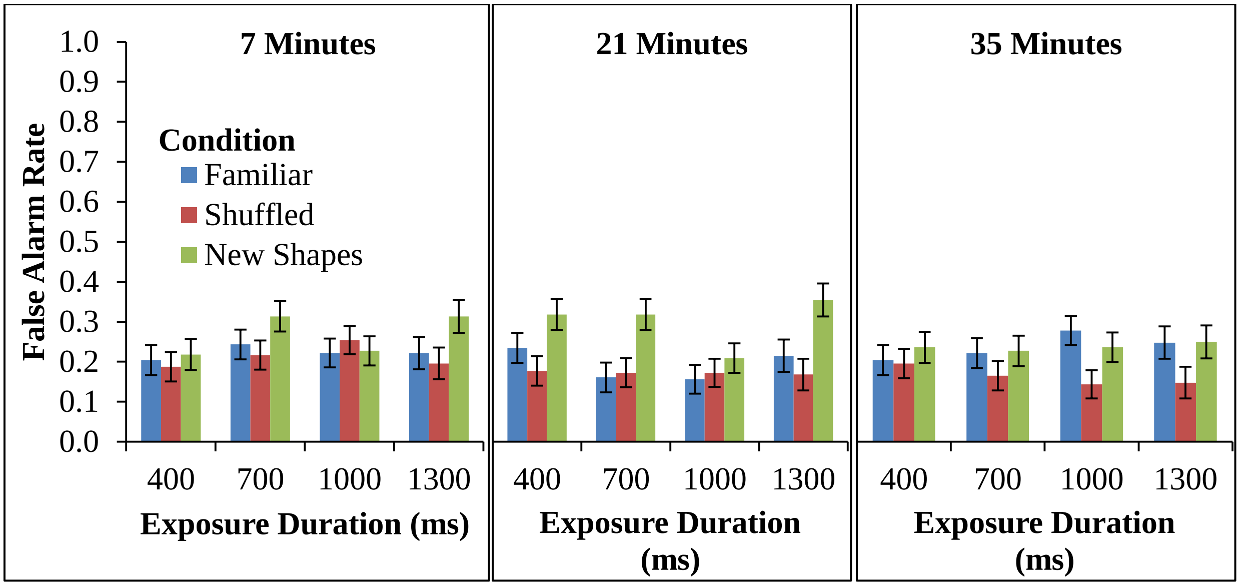

Figure 54. Effects of Condition, Session Length, and Exposure Duration on false alarm rate. Bar heights indicate adjusted marginal means and error bars indicate standard error of the mean.

False Alarm Rate. Figure 54 showed effects of Condition, Session Length, and Exposure Duration on false alarm rate. It appeared that Familiar and Shuffled had lower false alarm rates across session lengths and exposure durations, except that Familiar was about the same as New Shapes at 35 minutes. I tested the apparent effects in an ANOVA of Condition by Session Length by Exposure Duration on false alarm rate, which showed a main effect of

Condition, *F*(2,184) = 5.06, *p* = .007, *partial-eta-squared* = 0.05; an interaction of Session Length and Exposure Duration, *F*(6,552) = 2.18, *p* = .04, *partial-eta-squared* = 0.02; a marginal interaction of Condition and Exposure Duration, *F*(6,552) = 1.89, *p* = .08, *partial-eta-squared* = 0.02; and no other effects (all *p*'s > .32).

For the main effect of Condition, I followed up with custom hypothesis tests of all pairwise comparisons in ANOVA. Shuffled (*M* = 0.18, *SE* = 0.02) had fewer false alarms than New Shapes (*M* = 0.27, *SE* = 0.02), *F*(1,184) = 10.12, *p* = .002, *partial-eta-squared* = 0.05. Similarly, Familiar (*M* = 0.22, *SE* = 0.02) had marginally fewer false alarms than New Shapes, *F*(1,184) = 3.61, *p* = .06, *partial-eta-squared* = 0.02. Familiar did not differ from Shuffled (*p* = .16). Learning in terms of reduced false alarming transferred from Familiar to Shuffled.

For the interaction of Condition and Exposure Duration, I conducted custom hypothesis tests of simple effects of Condition for each exposure duration, and followed up with additional tests for significant simple effects. There was no simple effect of Condition for 400ms (*p* = .12) or 1000ms (*p* = .49). For 700ms, there was a simple effect of Condition, *F*(2,184) = 4.86, *p* = .009, *partial-eta-squared* = 0.05. I compared Familiar (*M* = 0.22, *SE* = .02) and Shuffled (*M* = 0.19, *SE* = 0.02), which did not differ in accuracy (*p* = .44). I compared Familiar and Shuffled to New Shapes (*M* = 0.29, *SE* = 0.03), and found that together they showed a lower false alarm rate than New Shapes , *F*(1,184) = 9.11, *p* = .003, *partial-eta-squared* = 0.05. Similarly, there was a significant simple effect of Condition for 1300ms, *F*(2,184) = 7.06, *p* = .001, *partial-eta-squared* = 0.07. I conducted all pairwise comparisons. Shuffled (*M* = 0.17, *SE* = 0.02) showed a lower false alarm rate than New Shapes (*M* = 0.31, *SE* = 0.03), *F*(1,184) = 14.13, *p* < .001, *partial-eta-squared* = 0.07. Familiar (*M* = 0.23, *SE* = 0.03) also showed a lower false alarm rate than New Shapes at 1300ms, *F*(1,184) = 4.67, *p* = .03, *partial-eta-squared* = 0.03. Familiar showed a marginally higher false alarm rate than Shuffled, *F*(1,184) = 3.05, *p* = .08, *partial-eta-squared* = 0.02. The interaction of Condition and Exposure Duration was driven by the different simple effects of Condition at 700ms and 1300ms.

For the interaction of Session Length and Exposure Duration, I tested for simple effects of Session Length at each exposure duration, but there were no reliable effects (all *p*'s > .20). Then I tested for simple effects of Exposure Duration at each session length by splitting the data on session length. At 21 minutes, there was a significant simple effect of Exposure Duration, *F*(3,159) = 2.84, *p* = .04, *partial-eta-squared* = 0.05. 400ms (*M* = 0.24, *SE* = 0.03) and 1300ms (*M* = 0.25, *SE* = 0.03) did not differ in false alarm rate (*p* = .93). Combined, the shortest and longest exposure durations and 700ms (*M* = 0.22, *SE* = 0.03) did not differ (*p* = .27). 1000ms (*M* = 0.18, *SE* = 0.02) showed a marginally lower false alarm rate than the other three exposure durations combined, *F*(1,53) = 9.14, *p* = .004, *partial-eta-squared* = 0.15. There was no simple effect of Exposure Duration for 7 minutes (*p* = .11) or 35 minutes (*p* = .27) of familiarization.

Sensitivity. Figure 55 showed effects of Condition, Session Length, and Exposure Duration on sensitivity. It appeared that Familiar showed the most sensitivity for 21 minutes across exposure durations, Shuffled showed the most for 35 minutes across exposure durations, and for 7 minutes, there was only learning at the longest exposure duration. To test these apparent effects, I conducted an ANOVA of Condition by Session Length by Exposure Duration on sensitivity, which demonstrated a significant main effect of Condition *F*(2,184) = 5.99, *p* = .003, *partial-eta-squared* = 0.06. It also showed a main effect of Exposure Duration *F*(3,552) = 3.53, *p* = .02, *partial-eta-squared* = 0.02, and a marginal interaction of Condition, Session Length, and Exposure Duration, *F*(12,552) = 1.62, *p* = .08, *partial-eta-squared* = 0.03. No other effects were found (all *p*'s > .10).

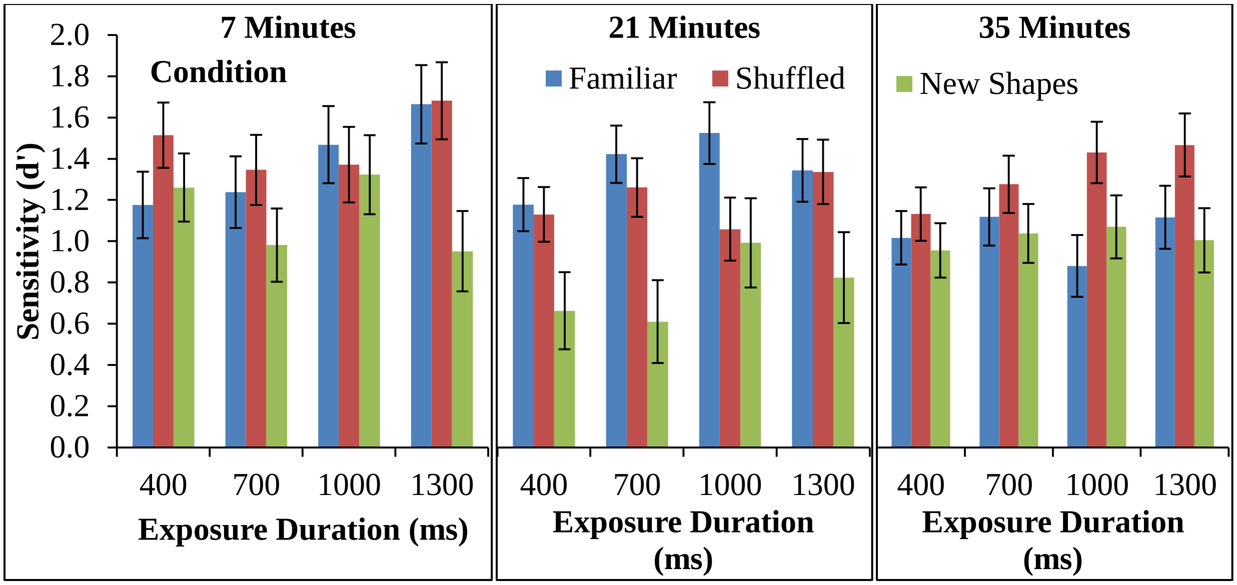

Figure 55. Effects of Condition, Session Length, and Exposure Duration on accuracy. Bar heights indicate adjusted marginal means and error bars indicate standard error of the mean.

I conducted custom hypothesis tests to investigate the main effects. For Condition, compared Familiar (*M* = 1.46 *SE* = 0.08) and Shuffled (*M* = 1.54, *SE* = 0.08), which did not differ on sensitivity (*p* = .48). I compared Familiar and Shuffled to New Shapes (*M* = 1.22, *SE* = 0.10), and found that Familiar and Shuffled showed higher sensitivity than New Shapes, *F*(1,184) = 11.49, *p* = .001, *partial-eta-squared* = 0.06. The learning transferred from Familiar to Shuffled.

For Exposure Duration, the shorter exposure durations, 400ms (*M* = 1.28, *SE* = 0.06) and 700ms (*M* = 1.33, *SE* = 0.06), did not differ in sensitivity (*p* = .45). Similarly, 1000ms (*M* = 1.43, *SE* = 0.07) and 1300ms (*M* = 1.46, *SE* = 0.07) did not differ in sensitivity (*p* = .61). However, the shorter exposure durations showed significantly lower sensitivity than the longer exposure durations (1000ms, 1300ms) combined, *F*(1,184) = 10.61, *p* = .001, *partial-eta-squared* = 0.06.

In looking at Figure 54, it appeared that the three-way interaction of Condition, Session Length, and Exposure Duration was due to Familiar showing the highest sensitivity for 21 minutes and Shuffled showing the highest sensitivity for 35 minutes, and to learning for the 7-minute exposure duration being apparent only for the longest exposure duration. To test the interaction, I broke the data on Exposure Duration and tested simple interactions for Condition and Session Length at each exposure duration. There was no simple interaction of Condition and Session Length at 400ms (*p* = .36), 700ms (*p* = .21), or 1300ms (*p* = .56). However, there was an interaction of Condition and Session Length at 1000ms, *F*(4,184) = 3.08, *p* = .01, *partial-eta-squared* = 0.06. I used custom hypothesis tests to investigate the marginal interaction at 1000ms by evaluating simple simple effects of Condition for each session length at 1000ms via custom hypothesis tests in ANOVA and followed up on significant simple simple effects with additional tests. There was no simple simple effect of Condition for 7 minutes (*p* = .86). There was a simple simple effect of Condition for 21 minutes, *F*(2,184) = 3.17, *p* = .04, *partial-eta-squared* = 0.03. I followed up with all pairwise comparisons. Familiar (*M* = 1.91, *SE* = 0.19) showed significantly higher sensitivity than Shuffled (*M* = 1.32, *SE* = 0.19), *F*(1,184) = 4.75, *p* = .03, *partial-eta-squared* = 0.03. Familiar was also more accurate sensitive than New Shapes (*M* = 1.24, *SE* = 0.27), *F*(1,184) = 4.10, *p* = .04, *partial-eta-squared* = 0.02. Shuffled and New Shapes did not differ (*p* = .80). There was a simple simple effect of Condition for 35 minutes of familiarization, *F*(2,184) = 3.49, *p* = .03, *partial-eta-squared* = 0.03. I followed this up with all pairwise comparisons. Familiar (*M* = 1.10, *SE* = 0.19) was less accurate than Shuffled (*M* = 1.79, *SE* = 0.19), *F*(1,184) = 6.77, *p* = .01, *partial-eta-squared* = 0.04. New Shapes (*M* = 1.34, *SE* = 0.19) was marginally less sensitive than Shuffled, *F*(1,184) = 2.84, *p* = .09, *partial-eta-squared* = 0.02. New Shapes did not differ in accuracy from Familiar (*p* = .38).

I also directly tested the apparent “flip” from Familiar having the numerically highest accuracy across exposure durations for 21 minutes to Shuffled having the numerically highest accuracy across exposure durations for 35 minutes by examining the interaction of Condition and Session Length in an ANOVA of Condition by Session Length (21, 35) by Exposure Duration by Target Presence on accuracy. There was a marginal interaction of Condition and Session Length across exposure durations and levels of Target Presence, *F*(2,118) = 2.86, *p* = .06, *partial-eta-squared* = 0.05. There was a simple effect of Condition at 21 minutes, *F*(2,118) = 3.97, *p* = .02, *partial-eta-squared* = 0.06, and I followed up with all pairwise comparisons. Familiar (*M* = 1.71, *SE* = 0.14) was more accurate than New Shapes (*M* = 0.97, *SE* = 0.20), *F*(1,118) = 7.92, *p* = .006, *partial-eta-squared* = 0.06. Shuffled (*M* = 1.50, *SE* = 0.14) was more accurate than New Shapes, *F*(1,118) = 3.97, *p* = .05, *partial-eta-squared* = 0.03. Familiar and Shuffled did not differ (*p* = .32). There was no reliable simple effect of Condition at 35 minutes (*p* = .13).

I also directly tested the apparent effect of the learning for the 7-minute session length being restricted to the longest exposure duration. I tested the simple interaction of Condition and Exposure Duration for 7 minutes of familiarization[33]. For 7 minutes of Familiarization, there was a simple interaction of Condition and Exposure Duration on sensitivity, *F*(6,198) = 2.44, *p* = .03, *partial-eta-squared* = 0.07. Participants’ sensitivity for 7 minutes of familiarization at 1300ms showed a simple simple effect of Condition, *F*(2,66) = 4.34, *p* = .02, *partial-eta-squared* = 0.12. I compared Familiar (*M* = 1.66, *SE* = 0.19) to Shuffled (*M* = 1.68, *SE* = 0.19), which did not differ (*p* = .95). Then I compared Familiar and Shuffled to New Shapes (*M* = 0.95, *SE* = 0.20), and found that Familiar and Shuffled had significantly higher sensitivity than New Shapes,

[33] There was no simple interaction of Condition and Exposure Duration at 21minutes (*p* = .15) or 35 minutes (*p* = .31).

F(1,66) = 8.67, *p* = .004, *partial-eta-squared* = 0.12. For 7 minutes of familiarization, there was no simple simple effect of Condition for 400ms (*p* = .20), 700ms (*p* = .21), or for 1000ms (*p* = .82). The three-way interaction was driven by the "flip" at 1000ms from Familiar having the highest performance for 21 minutes to Shuffled having the highest at 35 minutes (and no differences at 1300ms), by the transfer from Familiar to Shuffled at 21 minutes, and by the learning only at 1300ms for 7 minutes.

References

Ahissar, M., & Hochstein, S. (1993). Attentional control of early perceptual learning. *Proceedings of the National Academy of Sciences*, *90*(12), 5718–5722.

Baker, C. I., Olson, C. R., & Behrmann, M. (2004). Role of Attention and Perceptual Grouping in Visual Statistical Learning. *Psychological Science*, *15*(7), 460–466. https://doi.org/10.1111/j.0956-7976.2004.00702.x

Bao, S. (2015). Perceptual learning in the developing auditory cortex. *European Journal of Neuroscience*, *41*(5), 718–724. https://doi.org/10.1111/ejn.12826

Barakat, B. K., Seitz, A. R., & Shams, L. (2015). Visual rhythm perception improves through auditory but not visual training. *Current Biology*, *25*(2), 60-61.

Bays, B. C., Turk-Browne, N. B., & Seitz, A. R. (2015). Dissociable behavioural outcomes of visual statistical learning. *Visual Cognition*, *23*(9–10), 1072–1097. https://doi.org/10.1080/13506285.2016.1139647

Belshaw, J. P. (1951). Economic survey of Asia and the Far East, 1949. *The Australian Quarterly 23*(1) 122-124.

Bertels, J., Franco, A., & Destrebecqz, A. (2012). How implicit is visual statistical learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1425–1431. https://doi.org/10.1037/a0027210

Biederman, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(4), 640–645. http://doi.org/10.1037/0278-7393.13.4.640

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*(4), 2299–2310. https://doi.org/10.1121/1.418276

Brady, T. F., & Oliva, A. (2008). Statistical Learning Using Real-World Scenes Extracting Categorical Regularities Without Conscious Intent. *Psychological Science*, *19*(7), 678–685. https://doi.org/10.1111/j.1467-9280.2008.02142.x

Bratzke, D., Seifried, T., & Ulrich, R. (2012). Perceptual learning in temporal discrimination: Asymmetric cross-modal transfer from audition to vision. *Experimental Brain Research*, *221*(2), 205–210. https://doi.org/10.1007/s00221-012-3162-0

Bufford, C. A., Mettler, E., Geller, E. H., & Kellman, P. J. (2014). The psychophysics of algebra expertise: Mathematics perceptual learning interventions produce durable encoding changes. In P. Bellow, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of* *the 36th Annual Conference of the Cognitive Science Society* (pp. 272-277). Austin, TX. Cognitive Science Society.

Bufford, C. A., Thai, K. P., Ho, J., Xiong, C., Hines, C. A., & Kellman, P. J. (2016). Perceptual learning of abstract musical patterns: Recognizing composer style. *Proceedings of the 14th Biannual International Music Perception and Cognition Conference.*

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81. doi:10.1016/0010-0285(73)90004-2

Cheng, P. (2014). Copying equations to assess mathematical competence: An evaluation of pause measures using graphical protocol analysis. In P. Bellow, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 319-324). Austin, TX. Cognitive Science Society.

Conway, C. M., & Christiansen, M. H. (2006). Statistical Learning Within and Between Modalities: Pitting Abstract Against Stimulus-Specific Representations. *Psychological Science*, *17*(10), 905–912. https://doi.org/10.1111/j.1467-9280.2006.01801.x

Conway, C. M., & Christiansen, M. H. (2009). Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive Psychology*, *21*(4), 561–580. https://doi.org/10.1080/09541440802097951

Conway, C. M., Goldstone, R. L., & Christiansen, M. H. (2007). Spatial constraints on visual statistical learning of multi-element scenes. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 29, No. 29).

Cosman, J. D., & Vecera, S. P. (2014). Establishment of an attentional set via statistical learning. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 1–6. https://doi.org/10.1037/a0034489

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant Melodies: Statistical Learning of Nonadjacent Dependencies in Tone Sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(5), 1119–1130. https://doi.org/10.1037/0278-7393.30.5.1119

Dosher, B. A., & Lu, Z.-L. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proceedings of the National Academy of Sciences*, *95*(23), 13988–13993. https://doi.org/10.1073/pnas.95.23.13988

Emberson, L. L., & Rubinstein, D. Y. (2016). Statistical learning is constrained to less abstract patterns in complex sensory input (but not the least). *Cognition*, *153*, 63–78. https://doi.org/10.1016/j.cognition.2016.04.010

Endress, A. D., Nespor, M., & Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends in Cognitive Sciences*, *13*(8), 348–353. https://doi.org/10.1016/j.tics.2009.05.005

Epstein, W. (1967). *Varieties of perceptual learning*. McGraw-Hill Inc.

Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review*, *57*(2), 94–107. https://doi.org/10.1037/h0058559

Estes, W K. (1962). Learning Theory. *Annual Review of Psychology*, *13*(1), 107–144. https://doi.org/10.1146/annurev.ps.13.020162.000543

Estes, W. K., & Straughan, J. H. (1954). Analysis of a verbal conditioning situation in terms of statistical learning theory. *Journal of Experimental Psychology*, *47*(4), 225–234. https://doi.org/10.1037/h0060989

Fahle, M., & Poggio, T. A. (2002). *Perceptual Learning*. MIT Press.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Fiser, J. (2009). Perceptual learning and representational learning in humans and animals. *Learning & Behavior*, *37*(2), 141–153. https://doi.org/10.3758/LB.37.2.141

Fiser, J., & Aslin, R. N. (2001). Unsupervised Statistical Learning of Higher-Order Spatial Structures from Visual Scenes. *Psychological Science*, *12*(6), 499–504. https://doi.org/10.1111/1467-9280.00392

Fiser, J., & Aslin, R. N. (2005). Encoding Multielement Scenes: Statistical Learning of Visual Feature Hierarchies. *Journal of Experimental Psychology: General*, *134*(4), 521–537. https://doi.org/10.1037/0096-3445.134.4.521

Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: From behavior to neural representations. *Trends in Cognitive Sciences*, *14*(3), 119–130. https://doi.org/10.1016/j.tics.2010.01.003

Fiser, J., Scholl, B. J., & Aslin, R. N. (2007). Perceived object trajectories during occlusion constrain visual statistical learning. *Psychonomic Bulletin & Review*, *14*(1), 173–178. https://doi.org/10.3758/BF03194046

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117–125. https://doi.org/10.1016/j.tics.2014.12.010

Garrigan, P., & Kellman, P. J. (2008). Perceptual learning depends on perceptual constancy. *Proceedings of the National Academy of Sciences*, 105(6), 2248-2253.

Gibson, E. J. (1953). Improvement in perceptual judgments as a function of controlled practice or training. *Psychological Bulletin*, *50*(6), 401–431. https://doi.org/10.1037/h0055517

Gibson, E. J. (1969). Principles of perceptual learning and development. NY: Appleton-Century-Crofts.

Gibson, J. J. (1947). *Motion picture testing and research.* Retrieved from https://apps.dtic.mil/docs/citations/AD0651783

Gibson, J. J., & Gibson, E. J. (1955). Perceptual Learning: Differentiation or Enrichment? *Psychological Review*, *62*(1), 32–41. https://doi.org/http://dx.doi.org/10.1037/h0048826

Glicksohn, A., & Cohen, A. (2011). The role of Gestalt grouping principles in visual statistical learning. *Attention, Perception, & Psychophysics*, *73*(3), 708–713. https://doi.org/10.3758/s13414-010-0084-4

Goldstone, R. L. (1998). Perceptual Learning. *Annual Review of Psychology*, *49*(1), 585–612. http://doi.org/10.1146/annurev.psych.49.1.585

Goldstone, R. L. (2000). Unitization during category learning. *JEP: Human Perception and Performance 26*(1), 86-112.

Hall, G. (1991). Perceptual and associative learning. *Oxford psychology series*, (18), 29-107.

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values ofd′. *Behavior Research Methods, Instruments, & Computers*, *27*(1), 46-51.

Hawkey, D. J., Amitay, S., & Moore, D. R. (2004). Early and rapid perceptual learning. *Nature neuroscience*, *7*(10), 1055.

Helmholtz, H. L. F. V. (1864). On the normal motions of the human eye in relation to binocular vision. *Proceedings of the Royal Society of London*, (13), 186–199.

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, *130*(4), 658–680. https://doi.org/10.1037/0096-3445.130.4.658

Jannarone, R. J., Yu, K. F., & Takefuji, Y. (1988). Conjunctoids: Statistical learning modules for binary events. *Neural Networks*, *1*(4), 325–337. https://doi.org/10.1016/0893-6080(88)90006-8

Jiang, Y., & Chun, M. M. (2001). Selective attention modulates implicit learning. *The Quarterly Journal of Experimental Psychology Section A*, *54*(4), 1105–1124. https://doi.org/10.1080/713756001

Jones, J. L., & Kaschak, M. P. (2012). Global statistical learning in a visual search task. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(1), 152–160. https://doi.org/10.1037/a0026233

Karmarkar, U. R., & Buonomano, D. V. (2003). Temporal Specificity of Perceptual Learning in an Auditory Discrimination Task. *Learning & Memory*, *10*(2), 141–147. https://doi.org/10.1101/lm.55503

Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature*, *365*(6443), 250.

Keller, F. S. (1943). Studies in International Morse Code. I. A new method of teaching code reception. *Journal of Applied Psychology*, *27*(5), 407–415. https://doi.org/10.1037/h0055363

Kellman, P. J. (2002). Perceptual learning. In R. Gallistel (Ed.), *Stevens' handbook of experimental psychology, volume 3: Learning, motivation, and emotion*. (3rd ed., pp. 259-299). NY: John Wiley & Sons.

Kellman, P. J. (2013). Adaptive and Perceptual Learning Technologies in Medical Education and Training. *Military Medicine*, *178*(10S), 98–106. http://doi.org/10.7205/MILMED-D-13-00218

Kellman, P.J. & Garrigan, P.B. (2009). Perceptual learning and human expertise. *Physics of Life Review, 6(*2), 53-84.

Kellman, P.J., Massey, C.M., & Son, J. (2010). Perceptual learning modules in mathematics: Enhancing students' pattern recognition, structure extraction, and fluency. *Topics in Cognitive Science*, 2(2), 285-305.

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters*, *461*(2), 145–149. https://doi.org/10.1016/j.neulet.2009.06.030

Knoke, D., & Burke, P. J. (1980). *Log-linear models* (Vol. 20). Sage.

Kok, E. M., de Bruin, A. B. H., Robben, S. G. F., & van Merriënboer, J. J. G. (2013). Learning radiological appearances of diseases: Does comparison help? *Learning and Instruction*, *23*, 90–97. doi:10.1016/j.learninstruc.2012.07.004

Lajoie, S. P. (1997). The use of technology for modeling performance standards in statistics. In J. Garfield & G. Burrill (Eds.), Research on the Role of Technology in Teaching and Learning Statistics (pp. 57-70). Voorburg, The Netherlands: International Statistical Institute.

Law, C. T., & Gold, J. I. (2009). Reinforcement learning can account for associative and perceptual learning on a visual-decision task. *Nature neuroscience*, *12*(5), 655.

Leek, M. R., & Watson, C. S. (1988). Auditory perceptual learning of tonal patterns. *Perception & Psychophysics*, *43*(4), 389–394. https://doi.org/10.3758/BF03208810

Lu, H., & Lee, A. (2013). Inferring hidden parts by learning hierarchical representations of objects. *Journal of Vision*, *13*(9), 782–782. https://doi.org/10.1167/13.9.782

McLaren, I. P. L., & Mackintosh, N. J. (2000). An elemental model of associative learning: I. Latent inhibition and perceptual learning. *Animal Learning & Behavior*, *28*(3), 211-246.

Mettler, E., & Kellman, P. J. (2014). Adaptive response-time-based category sequencing in perceptual learning. *Vision Research*, *99,* 111–123.

Nagarajan, S. S., Blake, D. T., Wright, B. A., Byl, N., & Merzenich, M. M. (1998). Practice-Related Improvements in Somatosensory Interval Discrimination Are Temporally Specific But Generalize across Skin Location, Hemisphere, and Modality. *Journal of Neuroscience*, *18*(4), 1559–1570. https://doi.org/10.1523/JNEUROSCI.18-04-01559.1998

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238. https://doi.org/10.1016/S0010-0285(03)00006-9

Notman, L. A., Sowden, P. T., & Özgen, E. (2005). The nature of learned categorical perception effects: a psychophysical approach. *Cognition*, *95*(2), B1–B14. doi:10.1016/j.cognition.2004.07.002

Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences*, *105*(7), 2745–2750. https://doi.org/10.1073/pnas.0708424105

Otsuka, S., Nishiyama, M., Nakahara, F., & Kawaguchi, J. (2013). Visual statistical learning based on the perceptual and semantic information of objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(1), 196–207. https://doi.org/10.1037/a0028645

Ottmar, E., Landy, D., Weitnauer, E., & Goldstone, R. (2015). Graspable Mathematics: Using Perceptual Learning Technology to Discover Algebraic Notation. *Integrating Touch-Enabled and Mobile Devices into Contemporary Mathematics Education*, 24–48. https://doi.org/10.4018/978-1-4666-8714-1.ch002

Pednault, E. P. D. (1998). Statistical Learning Theory. *MIT Encyclopedia of the Cognitive Sciences*, 798–800. MIT Press.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical Learning in a Natural Language by 8-Month-Old Infants. *Child Development*, *80*(3), 674–685. https://doi.org/10.1111/j.1467-8624.2009.01290.x

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in cognitive sciences*, *10*(5), 233-238.

Petrov, A. A., Dosher, B. A., & Lu, Z.-L. (2005). The dynamics of perceptual learning: An incremental reweighting model. *Psychological Review*, *112*(4), 715–743. https://doi.org/10.1037/0033-295X.112.4.715

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*(6), 855–863. https://doi.org/10.1016/S0022-5371(67)80149-X

Reber, A. S., & Millward, R. B. (1971). Event tracking in probability learning. *The American Journal of Psychology, 84*(1), 85–99. https://doi.org/10.2307/1421227

Robinson, C. W., & Sloutsky, V. M. (2007). Visual processing speed: Effects of auditory input on visual processing. *Developmental Science*, *10*(6), 734–740. https://doi.org/10.1111/j.1467-7687.2007.00627.x

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological review*, *84*(1), 1.

Seriès, P., & Seitz, A. (2013). Learning what to expect (in visual perception). *Frontiers in human neuroscience*, *7*, 668.

Slone, L. K., & Johnson, S. P. (2018). When learning goes beyond statistics: infants represent visual sequences in terms of chunks. *Cognition*, *178*, 92-102.

Song, Y., Peng, D., Lu, C., Liu, C., Li, X., Liu, P., et al. (2007). An event-related potential study on perceptual learning in grating orientation discrimination. *Neuroreport*, *18*(9), 945–948.

Seitz, A. R., Kim, R., van Wassenhove, V., & Shams, L. (2007). Simultaneous and Independent Acquisition of Multisensory and Unisensory Associations. *Perception*, *36*(10), 1445–1453. https://doi.org/10.1068/p5843

Suppes, P., & Atkinson, R. C. (1960). *Markov learning models for multiperson interactions*. Palo Alto, CA, US: Stanford Univer. Press.

Thai, K. P., Mettler, E., & Kellman, P. J. (2011) Basic information processing effects from perceptual learning in complex, real-world domains. In L. Carlson, C. Holscher, & T. Shipley (Eds.), Proceedings of the 33rd Annual Conference of the Cognitive Science Society (pp. 555-560). Boston, MA: Cognitive Science Society.

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*(2), B25–B34. https://doi.org/10.1016/j.cognition.2005.01.006

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The Automaticity of Visual Statistical Learning. *Journal of Experimental Psychology: General*, *134*(4), 552–564. https://doi.org/10.1037/0096-3445.134.4.552

Turk-Browne, N. B., & Scholl, B. J. (2009). Flexible visual statistical learning: Transfer across space and time. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(1), 195–202. https://doi.org/10.1037/0096-1523.35.1.195

Wagman, J. B., Carello, C., Schmidt, R. C., & Turvey, M. T. (2009). Is Perceptual Learning Unimodal? *Ecological Psychology*, *21*(1), 37–67. https://doi.org/10.1080/10407410802626027

Walker, G. (1879). The census of 1880. *The North American Review*, *128*(269), 393–404.

Watanabe, T., Náñez, J. E., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature*, *413*(6858), 844.

Watson, J. M. (1997). Assessing statistical thinking using the media. In *The Assessment Challenge in Statistics Education* (pp. 107–121). University of Tasmania: IOS Press.

Westheimer, G., & McKee, S. P. (1978). Stereoscopic acuity for moving retinal images. *Journal of the Optical Society of America*, *68*(4), 450-455.

Wickens, T. D. (2002). *Elementary signal detection theory*. Oxford University Press, USA.