Journal of Experimental Psychology: General

Perceptual Similarity Affects the Learning Curve (but Not Necessarily Learning)

Tim Wifall, Bob McMurray, and Eliot Hazeltine

Online First Publication, December 10, 2012. doi: 10.1037/a0030865

CITATION

Wifall, T., McMurray, B., & Hazeltine, E. (2012, December 10). Perceptual Similarity Affects the Learning Curve (but Not Necessarily Learning). *Journal of Experimental Psychology: General*. Advance online publication. doi: 10.1037/a0030865

Perceptual Similarity Affects the Learning Curve (but Not Necessarily Learning)

Tim Wifall, Bob McMurray, and Eliot Hazeltine
University of Iowa

What role does item similarity play in motor skill acquisition? To examine this question, we used a modified version of the chord learning task (Seibel, 1963) that entails producing simultaneous finger key presses, similar to playing a chord on a piano. In Experiment 1, difficulty, as indexed by response time (RT) to a particular chord on the first session, was held constant, and chords that were similar to other chords had longer RTs after practice than dissimilar chords. In Experiment 2, we used chords that produced different initial RTs to show that similarity affected asymptotic RT rather than the size of RT decrement achieved with practice. In Experiment 3, we eliminated differences in perceptual similarity by using Chinese characters for stimuli while retaining differences in motoric similarity, which resulted in nearly identical asymptotes for similar and dissimilar chords. Thus, the density effect observed in Experiments 1 and 2 appears to stem from competition triggered by similar stimuli. Because performance differences were immediately re-established when stimulus similarity was introduced in Experiment 3 during transfer sessions, competition appears to emerge among learned, central representations that can be coactivated by multiple stimuli.

Keywords: skill acquisition, learning, similarity, response selection, practice

Skill acquisition is central to human behavior; the simplest everyday tasks would become hopelessly onerous without the benefit of experience. Research on skill acquisition has uncovered some general principles describing how performance improves with practice and how different training protocols (e.g., availability of feedback, structure of training sessions) affect the rate of improvement (e.g., Anderson, 1982; Logan, 1992; Magill & Hall, 1990; Newell & Rosenbloom, 1981; Schmidt & Bjork, 1992; Shepard, Hovland, & Jenkins, 1961). However, there has been little consideration of how the properties of the items that are being learned affect the learning curve. Most studies either attempt to use items for which many properties are uniform or treat variation in the items as a nuisance variable (e.g., Anderson, 1982; Logan, 1988; Newell & Rosenbloom, 1981).

In the present study, we investigated how one property—similarity among to-be-learned items—affects the practice-related changes in performance. In particular, we investigated the mechanism by which similarity affects performance. To characterize the effects of similarity, we focused on three related questions. One key question is, Does similarity primarily relate to changes in the

rate at which items are acquired or to changes real-time task operations? We refer to this as the *locus question*. However, before the locus question can be addressed, one must establish whether similarity facilitates or impedes the learning of trained items (termed the *direction question*), and do perceptual similarity (i.e., similarity among the stimulus cues for the items) and response similarity (i.e., similarity among the appropriate motor outputs) play the same role (termed the *domain question*)?

Given the small amount of work that has been done on this topic to date, these three questions are meant to provide a starting point for addressing the role that similarity plays in skill learning. The answer to each of these broad questions likely depends in part on the particulars of the task. For example, the roles of stimulus and response similarity may be contingent on the relative amounts of variation among the stimuli and responses and the mappings between them. However, given the dearth of available data and the range of possible tasks, we chose a task in which stimulus and response similarity could be roughly equated, and the mappings were either orderly, with similar stimuli mapped to similar responses (Experiments 1 and 2), or completely arbitrary (Experiment 3). Moreover, as we note in the General Discussion, our findings have close parallels with other domains, most obviously in language learning and categorization, suggesting that they reflect fairly general principles.

Little research has been conducted on the role of similarity in skill learning, but there is reason to expect that its contribution may be complex. In some cases, similarity facilitates transfer (e.g., Gandolfo, Mussa-Ivaldi, & Bizzi, 1996), such that learning about one item can facilitate generalization to other, similar items. However, similarity might also produce contextual interference, which diminishes retention after a consolidation interval (e.g., Shea & Morgan, 1979), and similarity between items may enhance competition among items, making it more difficult to retrieve the

Tim Wifall, Bob McMurray, and Eliot Hazeltine, Department of Psychology and the Delta Center, University of Iowa.

We gratefully acknowledge Tana Truelove and Jodi Smith for their assistance in collecting the data. We also thank Richard Ivry and Andrew Heathcote for extremely helpful comments on an earlier version of the article. The research was supported by National Institute of Deafness and Other Communication Disorders Grant DC008089 to Bob McMurray.

Correspondence concerning this article should be addressed to Tim Wifall, E11 Seashore Hall, Department of Psychology, University of Iowa, Iowa City, IA 52241. E-mail: timothy-wifall@uiowa.edu

appropriate response (see [Palmeri, 1997](#)). In short, we note that most real-world tasks involve stimuli and responses that both vary in terms of their similarity to one another. There are multiple theoretical accounts to suggest interitem similarity plays a role in learning, but the domain over which it plays out and its locus are by no means clear. Even at the broadest level, it is not known whether similarity impedes or facilitates learning. Thus, our goal of this study was to investigate this basic issue and to probe the mechanisms that underlie these effects. First we examine how existing accounts of skill acquisition address the issue of similarity and then describe a model task that can be used to address our three questions.

Similarity and Theories of Skill Acquisition

A core behavioral phenomenon of skill acquisition can be described as the power law of learning. This suggests that decrements in response time (RT) with practice are large at first but diminish as training progresses. This finding is so ubiquitous that it has become a benchmark for theories of skill acquisition ([Anderson, 1982](#); [Logan, 1992](#)). Theories of skill acquisition propose different underlying mechanisms, some of which make distinct predictions with regard to similarity, although this is seldom directly tested.

For example, one prominent theory, chunking theory ([Newell & Rosenbloom, 1981](#)), focuses on tasks with complex encoding or output demands. According to chunking theory, the diminishing reductions in RT stem from the formation of hierarchically organized chunks that encompass increasingly large sets of task components. The lowest level chunks contain a single element of a stimulus (e.g., one movement in a complex movement), while a highest level chunk contains the whole movement. Once a chunk is encoded, its elements do not have to be retrieved individually, so performance is speeded. Because lower level chunks (i.e., chunks consisting of small number of items) occur more frequently than higher level chunks (i.e., chunks consisting of larger number of items), practice initially produces large gains, but the benefits become smaller as learning focuses on more specialized, larger chunks that occur in a smaller subset of items.

At its core, chunking theory holds that improvements in performance stem from the formation of chunks as items that contain them are practiced. Thus, the more often a chunk appears, the more quickly it should be encoded, and the sooner it will benefit performance. Moreover, the explanation assumes that practicing a chunk in one context facilitates performance of that chunk in another context. Otherwise, it would not be able to account for the decelerating reductions in RT. Therefore, a set of similar items that share chunks should be learned more quickly than dissimilar items that do not share chunks because in the former case, the chunks appear more often in more items, whereas in the latter case, the chunks appear less frequently and benefit the performance of fewer other items.

The prediction that similar chunks should facilitate performance has also been made in artificial grammar learning ([Servan-Schreiber & Anderson, 1990](#)). Here the competitive chunking mechanism is straightforward; once a chunk is learned for one string (of letters), that same chunk can aid in learning another string that uses that same chunk (as long as that the underlying grammar is the same for both sentences). For example, when a participant learns the chunk *VP* for the string *TTXVP*, learning this

chunk can facilitate learning the string *VXVPX*. In fact, [Servan-Schreiber and Anderson \(1990\)](#) proposed that familiarity is based on the number of shared chunks between a given item and those in memory and that more familiar strings are learned more quickly because shared chunks facilitate learning. It is not clear how similar the mechanisms underlying motor and grammar learning are, but this example illustrates the how chunking mechanisms enable similarity among items to facilitate learning.

Chunking theory offers clear answers for at least two of three questions we pose here. With respect to direction, there is no interference in chunking theory, so similarity benefits performance rather than producing costs. That is, chunking theory requires that chunks occurring in one item be potentially beneficial to the performance of another item. To the extent that two items share chunks, there should be transfer of learning. With respect to locus, chunking theory predicts that similarity affects the rate of encoding: more similar items will contain share more chunks, and these shared chunks will occur at a higher frequency and be learned more quickly. Finally, with respect to domain, chunking theory has been primarily applied to skilled performance and thus clearly holds that similarity among the responses can play a critical role. However, it is possible to apply chunking theory to perceptual learning tasks, leaving its predictions on this last issue open. Nonetheless, despite these clear predictions for motor tasks, there has been little empirical work to directly test them.

An alternative account of skill learning, the instance theory of automaticity ([Logan, 1988, 1992](#)), makes distinct claims about how interitem similarity might modulate the effect of practice. Instance theory proposes that practice-related decrements in RT stem from a race between a rule-based process and the retrieval of specific memory traces of previous responses to that stimulus. Each encounter with a stimulus initiates a race between an algorithmic process for determining the response and a memory trace of each prior instance. The first to finish produces the response. As the number of memory traces increases, so does the probability that at least one of the traces will be recalled before the algorithmic process is finished. Thus, RT decreases with practice, and these decrements diminish because the benefit of additional memory traces decreases as the number of existing memory traces increases.

Instance theory makes no explicit claims about the role of similarity. However, given the mechanism by which RTs are reduced according to instance theory, the effects of similarity must occur during acquisition or during retrieval of previously encoded items. For example, it might be assumed that more similar items are less memorable, and instance theory does allow for item memorability to affect the rate of learning. [Logan \(1988\)](#) suggested that less memorable (i.e., more similar, possibly) items are less likely to be encoded—thereby reducing the learning rate—or have prolonged retrieval times—slowing online processing. However, no empirical work has resolved which of the possible outcomes is obtained with less memorable items.

Modifications of instance theory applied to categorization tasks do make more explicit predictions regarding similarity. [Palmeri \(1997](#); see also, [Nosofsky & Palmeri, 1997](#)) asked participants to categorize dot patterns based on the number of dots they contained and manipulated both the similarity within and between categories. Similarity among items belonging to the same category facilitated

learning, whereas similarity between items belonging to distinct categories slowed learning.

Palmeri (1997) used an exemplar-based random walk model to simulate the retrieval process within instance theory. In the model, similarity affects the speed of retrieving exemplars associated with a particular category. A new item triggers a retrieval processes by which similar items are likely to be recalled more quickly than dissimilar items. In this way, similarity among items associated with the same category facilitates performance, while similarity among items associated with different categories impairs performance. As the retrieval effect is based on the items presently within memory, the effects of similarity gradually increase as more items are added to memory over the course of practice. Thus, the model provides clear predictions for our three questions: it holds that similarity can both facilitate or impair performance (depending on whether it is between- or within-category); that these effects are primarily related to similarity among the stimuli, not among the responses; and that they relate to changes in online processing at the time of performance (in this case, matching perceptual cues with an exemplar in memory), not to differences in how similar or dissimilar stimuli are learned.

While Palmeri (1997) offered explicit answers to our questions, it is not entirely clear how these ideas scale to more general learning situations for a number of reasons. First, the Palmeri (1997) model addresses a specific form of skill learning—learning to categorize items—and similarity among the responses plays no role in performance. Thus, this issue is not addressed by the model, nor is it addressed by the empirical work, as response similarity was not manipulated (but this is not to say that the model cannot take into account response similarity or multiple-response alternatives). Second, because the task requires the compression of complex, highly dimensional inputs to simple, unidimensional outputs, it is perhaps not surprising that similarity among items that belong to the same response category improves performance, whereas similarity among items that belong to different response categories impairs performance. These factors, which essentially describe the coherence of the categories, are known to affect the ease of categorization (cf. Goldstone, Lippa, & Shiffrin, 2001). However, it is not clear whether these principles hold in tasks that do not require such a reduction in dimensionality.

It is clear from this discussion that a crucial factor that has yet to be explored is the complexity or the multidimensionality of the motor output and that this factor may change the story entirely. In tasks involving more complex output requirements, such as playing a musical instrument, there are many competing inputs and outputs to sort out, and a different set of constraints may be important. In such cases, similarity among the inputs may or may not relate to similarity among the responses, and the number of categories may alter the relative importance of between- and within-category similarity.

In this light, while both chunking theory and instance theory aim to explain practice-related improvements in performance, they are typically applied to tasks on different ends of this perception–action chain. Categorization tasks, such as those used by Palmeri (1997), have been used to test the predictions of instance theory. These emphasize similarity only among the stimuli, not among the responses. Thus, the pattern of facilitation and interference observed by Palmeri (1997) may stem from the requirement to map complex sets of inputs to arbitrary and simple outputs with no

relation to each other. In contrast, empirical work on chunking theory typically uses tasks with complex response demands (e.g., Gobet et al., 2001). Thus, chunking tasks may emphasize factors associated with motor production, which in turn may mask the effects of competition during retrieval. However, to date, none of these tasks have been applied to the issues raised by interitem similarity, despite the clear predictions of chunking theory.

Manipulating response similarity may be crucial to understanding skill learning more broadly. Many real-world tasks involve responses with multiple overlapping features. For example, a spoken word is composed of multiple articulator movements; musical performance on an instrument like the piano requires selecting from chords and sequences with highly similar response demands; and pitching a baseball requires selecting from a range of similar and dissimilar grips and motor patterns to achieve a spin and velocity with the appropriate properties. In some of these cases, motor similarity is also highly related to perceptual similarity. For example, the motor pattern of a piano chord is analogous to the visual pattern of the notes on the page that serve as stimulus as well as the sound itself, the perceptual outcome of this action. Thus, two chords that are similar visually will also require similar motor outputs. Similarly, words that sound similar are typically produced with similar articulatory configurations.

However, the relationships are not always clear. For example, the production of guitar chords is less related to the notes on the page because the fingering is more arbitrary, and the desired motor features of a baseball swing relate to the perceptual properties of the pitch in a complex manner. In these cases, the possible responses may share some features but differ in terms of other features, and the relationship among the features of the response may or may not parallel relationships among the stimuli. It is possible, then, that similarity may work differently when the responses are complex and vary in terms of similarity like the stimuli.

Across these various cases, it is likely that learning processes make use of both stimulus and response similarity during skill acquisition, and in the present experiment, we aimed to assess how similarity among stimuli and among responses affect performance when each stimulus is associated with a unique response. To start, we examined the most straightforward arrangement in which similarity among the stimuli mirrors similarity among the responses.

Chord Learning

To address our three questions, we needed a task with specific properties. First, it should capture a high-dimensional to high-dimensional mapping, as this may closely resemble a variety of learning situations such as language, music, and complex motor skill learning. Second, it should place demands on both perceptual and motor systems by involving complex inputs and outputs. Third, it should be possible to systematically control similarity of both stimuli and responses, and fourth, to do so independently.

Seibel's (1963) classic chord learning task fits this bill. In the task, participants see a visual stimulus indicating which spatially corresponding response keys should be pressed simultaneously (see Figure 1), in a fashion similar to how a pianist would play a chord (see Drake & Palmer, 2000). In this way, the task can capture similarity among input stimuli (chords can have similar

Figure 1. Typical stimulus display and response board (with Xs and dashes) used by each participant.

visual cues) and motor output (chords can share fingers).¹ Previous work has demonstrated that learning in this task (as measured by RT) reveal diminishing returns, and such findings have been interpreted as support for chunking theory (Newell & Rosenbloom, 1981).

Subsequent research has elaborated on these findings, indicating that learning relates to the encoding of specific chords rather than improved processing of the individual elements within a chord (Hazeltnine, Aparicio, Weinstein, & Ivry, 2007). These findings are consistent with results from other skill acquisition studies examining mirror reading (Masson, 1986) and typing (Crump & Logan, 2010). Learning in these cases is based on features specific to practiced items; there is no evidence of transfer to items that do not share features with the items in the practiced set. Note that item-specific transfer is consistent with chunking theory and versions of instance theory (see Crump & Logan, 2010). More important, the fact that participants are learning item-based representations means that this task can be quite sensitive to similarity effects due to perceptual encoding, motor output, or learned central representations (the locus and domain questions).

While these studies have looked at how similarity between practiced and transfer items affects performance on the transfer items, they have not examined how similarity among the practiced items affects learning. The discrete set of elements (fingers and perceptual positions) used in the chord task allows an easy manipulation of similarity as researchers can control the number of elements shared by each chord. Thus, they can manipulate interitem similarity in a complex task that involves highly dimensional input and outputs. In this way, researchers can directly test how similarity affects the shape of the learning curve and rate of learning and, by manipulating the perceptual input, can disentangle effects of perceptual and motor similarity.

Using this model task in Experiment 1, we addressed our first question (direction) by examining the effect of similarity on learning between a similar and a dissimilar set of chords. In Experiment 2, we then confirmed this result in a between-subjects design and further contrasted the effects of similarity between chords by manipulating the difficulty of producing them individually. Fi-

nally, in Experiment 3, we addressed the domain and locus of the effects. To examine the domain question, we held perceptual similarity constant among the chords and manipulated the response similarity. By doing so, we could determine whether the differences observed in Experiments 1 and 2 derive from similarity relationships among the stimuli or among the responses. To address the locus question, we tested performance in transfer sessions in which unpracticed perceptual stimuli were used to cue practiced chords. This manipulation allowed us to determine whether the effects of similarity relate to differences in the real-time processing of the perceptual stimuli or differences in the rate at which representations of these chords are built (learning). Thus, Experiments 1 and 2 characterized the effect of similarity, and Experiment 3 explored where and how this effect changes task operations.

Characterizing the Learning Function

Across these experiments, our primary measure of learning is change in RT over training. As a result, mathematically describing the learning curve is crucial for characterizing the effects of similarity. An impressive body of work has attempted to provide a formal mathematical description of the decrements in RT typically observed across practice (e.g., Heathcote, Brown, & Mewhort, 2000; Newell & Rosenbloom, 1981; Palmeri, 1997).

The most widely cited description of the changes in RT is given by the power function (Newell & Rosenbloom, 1981):

$$RT = a + bN^{-c}. \quad (1)$$

Here, a is asymptotic performance, b is the difference between initial and asymptotic performance, N is the amount of practice, and c is the learning rate parameter.² The power law has been used to characterize learning in a myriad of domains, including choice-RT tasks (Morin & Forrin, 1962), generation of geometry proofs (Neves & Anderson, 1981), and cigar rolling (Crossman, 1959).

However, there is an active debate as to whether the reductions in RT with practice are best described by the power law or an exponential function (see Heathcote et al., 2000), particularly when individual items are analyzed separately (Myung, Kim, & Pitt, 2000). The exponential function is given by

$$RT = a + be^{-cN}. \quad (2)$$

As in the power law, a again represents asymptotic performance, b the amount that learning can reduce RT, and c describes the rate at which asymptotic level performance is approached as a proportion of the distance to the asymptote. Both formalisms describe functions with diminishing returns over learning. However, the exponential function assumes a constant relative learning rate—that is, a constant proportion of the remaining possible speedup is

¹ Intriguingly, this task potentially allows one to manipulate a third factor: the similarity among the perceptual consequences of the motor output (e.g., if the piano chord resulted in the generation of a sound). Although this was not investigated here, it may offer an intriguing analogue to situations like language, reading and music.

² An additional parameter can be added to the to the power function, E , representing the participant's prior learning before performing in the experiment. However, most researchers have been successful fitting the data with the three parameter version of the power function with E fixed at zero (e.g., Logan, 1988). Also, b assumes that $N = 1$ at the start of practice.

achieved on each trial—whereas the power function assumes a decreasing rate of learning across practice.

Our goal here was not to enter this debate, but rather to use these parameters to examine how interitem similarity affects learning. For example, in the exponential formulation, similarity could potentially change all three parameters. That is, similarity might affect the asymptote, a , such that similar items have a higher or lower ultimate level of performance than dissimilar ones (implying a locus in online processing, rather than learning). Or similarity might affect b , such that more or less overall improvement (i.e., difference between initial and final RTs) is possible for similar (or dissimilar) items. Finally, similarity might affect c , which determines the shape of the learning function, with similar items being acquired faster (or slower) than dissimilar ones (implying a locus of the effect in learning). Which component of the function is affected by similarity may help reveal something of the mechanism by which similarity exerts its effects or its locus.

While both the power law and the exponential function capture all three of these fundamental properties of the learning curve, its proper characterization is essential to ensuring that these parameter estimates are meaningful. Thus, we fit both functions to our data and found a better fit for the exponential function, as is typically found computing learning curves for individual items (Heathcote et al., 2000). We were then able to describe how similarity changed performance with respect to the three parameters.

Experiment 1

Experiment 1 tested the effect of interitem similarity (chord density) on learning in a version of the chord task in which stimulus and response similarity were functionally equivalent. Our goal was to determine whether similarity hindered or facilitated learning. We created two sets of chords that differed in terms of similarity and compared performance of these sets across practice. Each chord used three fingers across both hands. Similarity was measured as the number of pairs of key presses shared by two chords. The dense set consisted of chords that shared a pair with at least four other practiced chords, whereas the sparse set consisted of chords that shared no pairs with any other practiced chord (see Table 1). These chords were practiced over the course of eight hour-long sessions.

Table 1
Grammars Used in Experiments 1, 2, and 3

Experiments 1 and 3		Experiment 2			
Dense	Sparse	Dense		Sparse	
BEF (4)	ACI (0)	AEG (3)	CEF (3)	ADH (1)	BCI (1)
BEG (4)	ADF (0)	AEH (2)	BDI (2)	ADF (1)	GHJ (1)
BEH (6)	AGJ (0)	AGJ (3)	AHJ (3)	ABG (1)	FGJ (1)
BFH (4)	CDJ (0)	BGJ (2)	BHJ (3)	ACG (1)	BFH (1)
BGH (4)	CFG (0)	CDF (3)	CDG (2)	AEI (1)	FHI (1)
EFH (4)	DGI (0)	EFI (2)	DFI (3)	EGI (1)	CDJ (1)
EGH (4)	FIJ (0)	CEG (3)	BHI (2)	BCE (1)	DEJ (1)

Note. Letters in the grammar (ABC) refer to abstract elements of a chord that were mapped to specific key presses differently for different participants (e.g., “A” could refer to the left index finger for Participant 1 and the right ring finger for Participant 2). The number in parentheses shows the number of other chords that share a pair with that chord.

Grammars specified the relationships among elements in the chords without specifying the actual key presses. For example, ABC and ABD are two chords that share a pair of elements (AB); however, A, B, C, and D could potentially refer to any four fingers. This level of abstraction enabled us to better control production difficulty, as we describe in the following.

We created a set of grammars consisting of seven chords that shared pairs with each other and seven chords that shared no pairs with any other chord. We termed the chords that shared pairs with other chords *dense chords* (as they derived from a dense neighborhood of chords). Chords that shared no pairs with other chords were termed *sparse chords* (see Table 1 for list of the grammars). Note that the dense chords also consisted of elements that were used more frequently than the elements used in the sparse chords.

Independent of similarity, individual chords vary in terms of their production difficulty as indexed by RT. For example, a chord including the index finger and thumb may be less difficult than one consisting of the middle and ring finger. We wanted to equate these sources of difficulty between the sparse and dense chord sets, so we started by restricting the set of possible chords to only those involving three button presses. However, even within these chords, there was likely to be substantial differences in production difficulty.

Thus, we conducted a single-session pilot experiment in which 16 participants produced each of the 120 possible three-finger chords. Each chord was tested eight times for a total of 960 trials (divided into 32 blocks of 30 trials), producing a library of 120 chords (all possible three-finger combinations) and the corresponding mean RT (averaged across all 16 subjects). Within this simple task, there was substantial variability between chords, with mean RTs for individual chords ranging from approximately 900 ms to 2,100 ms, confirming that there would be a substantial effect of production difficulty on these chords (this will be explored more explicitly in Experiment 2).

To map the grammars to actual key presses, we exhaustively searched the 3,628,800 possible mappings between the 10 letters in the grammars and 10 actual fingerings for assignments that would be closely matched in terms of the mean RT according to the 120-chord library. The three mappings that produced the most similar mean RTs for the sparse and dense sets of chords, according to the pilot data (Mapping 1: $M_{\text{sparse}} = 1,651$ ms, $SD = 98$ ms; $M_{\text{dense}} = 1,640$ ms, $SD = 97$ ms; Mapping 2: $M_{\text{sparse}} = 1,601$ ms, $SD = 97$ ms; $M_{\text{dense}} = 1,591$ ms, $SD = 89$ ms; Mapping 3: $M_{\text{sparse}} = 1,601$ ms, $SD = 97$ ms; $M_{\text{dense}} = 1,606$ ms, $SD = 83$ ms), were counterbalanced across the six participants (see Appendix). In this way, we were able to test how item similarity affected learning while holding each item’s production difficulty fairly on the first session constant.

While this procedure controls for production difficulty, we could not control the frequency of individual elements in our within-subject design (but see Experiment 2). As noted previously, one consequence of the way density was instantiated in the grammars was that dense items were composed of elements that appeared more frequently than the elements in the sparse items. In other words, chords that share pairs with other chords also share more individual elements with other chords. This was unavoidable because pairs were repeated in the dense items but not in the sparse set. As a consequence, we were not able to distinguish between the roles of similarity at the element level and similarity at the structural (pair) level in Experiment 1. However, more frequently used elements should provide an extra benefit to the dense chords,

because these elements have been practiced more frequently, so if this is an issue, we would expect learning to be enhanced for the dense chords. Moreover, the design used in Experiment 2 does distinguish between these two forms of similarity.

Given the expense and added time associated with multiple session experiments, we decided to use a fairly small sample of six participants. Collecting a larger number of observations from a smaller number of participants is often done in practice studies that attempt to characterize changes over the course of learning (e.g., Hazeltine, Ruthruff, & Remington, 2006; Schumacher, Seymour, Glass, Kieras, & Meyer, 2001; Van Selst, Ruthruff, & Johnston, 1999). Because Cohen's d and partial eta-squared (η_p^2) can be distorted when the sample size is small, we also report unbiased d (d_{unb}) and omega-squared (ω^2) as additional measures of effect size (see Fritz, Morris, & Richler, 2012). As we report here, these measures consistently indicate that the effects of similarity are moderate to large, validating our use of a smaller sample. Most important, we replicate our principal findings across the three experiments presented here, to ensure that our conclusions are not based on Type I error.

Method

Participants. Six participants from the Iowa City area performed the eight-session experiment. Each participant was paid \$10 per session. Participants completed at least four sessions per week (sometimes five) but no more than one session per day.

Stimuli and procedure. PC computers running E-prime (Psychology Software Tools, Sharpsburg, PA) presented the stimuli and recorded the responses. Participants responded manually on two custom-built five-key response boards (20 cm $\times$ 30 cm) with spatially compatible keys (2 cm wide, 10 cm long for the four fingers and 13 cm long for the thumb). Each participant adjusted the position of the response boards until he or she could place all 10 fingers comfortably on each key.

Throughout the experiment, a fixation cross in the middle of the screen was flanked on either side by five horizontal lines (each line corresponded to one digit). Each line was 1.1 cm in length, and the distance between the lines was 2.2 cm. The distance between the horizontal lines on the left compared with the ones on the right was 4.9 cm.

Each trial began with three Xs appearing above three horizontal lines indicating the keys to press for the chord for that trial. Participants were instructed to press the corresponding keys as quickly and accurately as possible. The stimuli remained visible until participants made a response or 3,000 ms had elapsed. Visual feedback presented on the computer screen was given for each incorrect response indicating why the chord was incorrect (e.g., too many keys hit, incorrect keys pressed, and so on). Key presses had to occur simultaneously, with individual keys being pressed within 150 ms of each other. The intertrial interval was 1,500 ms.

Design. Participants were trained on 14 chords, seven dense and seven sparse. Each training block consisted of 28 trials, with each chord presented twice in a random order. A session consisted of 32 blocks (896 trials, each session lasting approximately 1 hr), and there were eight sessions, resulting in a total of 7,168 trials across training. Therefore, each chord was presented a total of 64 times during a session and 512 times during the entire experiment. Subjects were randomly assigned to one of three mappings between grammars and actual fingers.

Results and Discussion

Proportion correct was high throughout the entire experiment for both dense and sparse chords (95% correct for both chord types across eight sessions). We submitted the accuracy data to a 2 (density) $\times$ 8 (session) within-subject analysis of variance (ANOVA). There was a significant effect of session, $F(7, 35) = 3.41$, mean square error (MSE) = .002, $\eta_p^2 = .41$, $\omega^2 = .74$, $p < .007$, with accuracy increasing over the eight sessions. The effect of density, $F(1, 5) = 1.38$, $MSE = .001$, $\eta_p^2 = .28$, $\omega^2 = .06$, $p = .292$, was not significant and the Density $\times$ Session interaction, $F(7, 35) = 2.19$, $MSE = .000$, $\eta_p^2 = .30$, $\omega^2 = .58$, $p = .06$, was marginally significant.

Our primary analysis examined RTs for each chord; RTs were computed as the maximum latency across the three individual fingers. We excluded any RT from analysis if it resulted from incorrect response, nonsimultaneous key presses (i.e., the time between the first and last key press was greater than 150 ms, 1% of the data), or was less than 300 ms and greater than 2,000 ms (4% of the data). Although our primary analytic approach was based on power-law and exponential fits, we first used ANOVA to characterize the overall pattern of performance.

Figure 2 shows mean RTs by session for both dense and sparse groups. The RTs were submitted to a 2 (density) $\times$ 8 (session) within-subject ANOVA.³ The analyses were performed two ways: once with participants as a random factor (F_1 or t_1) to show that findings should generalize across participants and once with chord (based on the grammar, not the mapping) as a random factor (an item effect, F_2 or t_2), to show that the findings should generalize across chords (as in the "item analyses" typically reported in many psycholinguistic studies). It was not possible to use the actual fingering as a random factor because participants used different fingerings based on their grammar.

There was a significant main effect of session, $F_1(7, 35) = 135.16$, $MSE = 3,211$, $\eta_p^2 = .96$, $\omega^2 = .99$, $p < .001$, $F_2(7, 84) = 854.84$, $MSE = 621$, $\eta_p^2 = .97$, $\omega^2 = .999$, $p < .001$, as RTs decreased over the eight sessions. Density did not produce a significant effect, $F_1(1, 5) = 2.64$, $MSE = 3,402.29$, $\eta_p^2 = .346$, $\omega^2 = .21$, $p > .05$, $F_2 < 1$, but the Session $\times$ Density interaction was significant, $F_1(7, 35) = 3.00$, $MSE = 681.21$, $\eta_p^2 = .375$, $\omega^2 = .70$, $p < .05$, $F_2(7, 84) = 3.84$, $MSE = 621$, $\eta_p^2 = .24$, $\omega^2 = .003$, $p = .001$, indicating practice affected sparse and dense chords differently. On the first session, no difference was observed between the dense and sparse chords RTs ($M_{dense} = 1,352$ ms; $M_{sparse} = 1,350$ ms, $t_s < 1$), demonstrating that our procedure for selecting sets of chords with similar production difficulties was successful. In contrast, on Session 8, there was a 63-ms advantage for sparse chords compared with the dense chords ($M_{dense} = 821$ ms; $M_{sparse} = 758$) that was significant by participants, $t_1(5) = 2.54$, $d = 0.54$, $d_{unb} = .46$, $p = .05$, but not by items, $t_2(12) = 1.37$, $d = 0.74$, $d_{unb} = 0.69$, $p = .19$.

³ Before running this ANOVA, we first asked if there were any systematic differences between the three grammar $\rightarrow$ key presses mappings, by running the same ANOVA with mapping as an additional between-subjects factor. No differences were found between the mappings, $F(2, 3) = 1.94$, $MSE = 188,490$, $\eta_p^2 = .56$, $\omega^2 = .0003$, $p = .29$, and mapping did not interact with session, $F < 1$, or density, $F(2, 3) = 3.17$, $MSE = 1,821$, $\eta_p^2 = .68$, $\omega^2 = .0007$, $p = .18$. Thus, we collapsed across the mappings.

Figure 2. Mean reaction times across the eight sessions for dense and sparse chords along with the fits to the exponential function for Experiment 1. The lines indicate the best fit with an exponential function.

To evaluate specifically how density affects participants' performance over learning, we fitted an exponential function to the RTs from each chord for each participant (see Equation 2; Heathcote et al., 2000; Logan, 1992; Palmeri, 1997) to the RTs from each chord for each participant using a constrained gradient descent method that minimized the least-squares error. We then subjected the values of the parameters to t tests on the effect of density. The a parameter was allowed to vary between 0 and the participant's shortest RT for that chord across the sessions. The b parameter was allowed to vary between 0 and infinity. The c parameter was allowed to vary between 0 and 100. Of the 84 fits performed (six participants $\times$ 14 chords), a value associated with an upper or lower bound was obtained only twice for a and never for c and b . We also fit each participant's data to the power function, but the exponential function provided a better overall fit.⁴

The sparse and dense chords did not differ in terms of c , the rate of learning, $t_1(5) < 1$; $t_2(5) = 1.29$, $d = 0.70$, $d_{\text{unb}} = 0.65$, $p = .23$. However, the amount of learning, b , did differ between sparse and dense chords, $t_1(5) = 2.78$, $d = 0.80$, $d_{\text{unb}} = 0.68$, $p < .05$; $t_2(12) = 3.12$, $d = 1.34$, $d_{\text{unb}} = 1.25$, $p < .01$, with more total learning estimated for the sparse chords. The asymptotic RT parameter, a , was also significantly different between sparse and dense chords, $t_1(5) = 2.09$, $d = 0.96$, $d_{\text{unb}} = .81$, $p = .09$; $t_2(12) = 4.08$, $d = 1.53$, $d_{\text{unb}} = 1.43$, $p < .001$, with sparse chords being projected to achieve RTs that were ultimately 143 ms faster than dense chords (see Table 2).

The inverse relationship between a and b is expected, given that we equated initial RT; when one set of chords has a lower asymptote, it must also show greater improvement. However, the direction of the effect is somewhat surprising. The dense items were composed of individual elements that appeared more frequently than the individual elements in the sparse items. According to chunking theory (e.g., Newell & Rosenbloom, 1981; Servan-Schreiber & Anderson, 1990), this frequency of appearance should have offered an advantage for the dense chords. However, rather than demonstrating an advantage, these chords were performed more slowly, and the difference increased over practice, leading to smaller estimates of the amount that RT could be ultimately reduced (b) and larger estimates of asymptotic RTs (a). This observation suggests that the frequency of the individual elements

plays a comparatively small role in performance (see Fiser & Aslin, 2001, 2002; MacKay, 1969).

Thus, we find evidence that item similarity can interfere with skill acquisition. There was no effect on c , suggesting that both chords were learned at the same rate, but items that were similar to other items showed larger estimates of asymptotic performance and lower estimates of the amount of learning than items that were not similar to others. This raises the possibility that similarity primarily affects real-time performance, not learning. However, these differences cannot be interpreted unambiguously. The initial RTs were the same for dense and sparse chords, so larger estimates of the amount of change in RT (b) lead to smaller estimates of the asymptotic RT (a) and vice versa. Therefore, to disentangle whether similarity primarily affects the amount of learning or asymptotic performance, Experiment 2 tested sparse and dense chords that differed in terms of their initial RTs. While this is important for understanding the observed pattern of data, it is also crucial for addressing our questions. If similarity primarily affects asymptotic performance, the locus of similarity effects would be attributed to real time processes like perception or action planning—similarity sets the upper bounds on what people can do. In contrast, if similarity affects the amount of learning, this implies a locus in learning and storage.

Experiment 2

The primary goal of Experiment 2 was to examine whether similarity primarily affects the magnitude of the RT decrement or the asymptotic RT. To do this, we manipulated initial difficulty so that there was an easy set and a hard set of chords. If similarity primarily determines the asymptote, then differences in initial performance should not affect a (asymptotic performance) but will affect b (change in RT) as there will be more room for improvement for difficult chords. Alternatively, if similarity primarily determines the amount that RTs can be reduced (i.e., sparse chords can undergo more learning), then b should be driven by similarity and not by initial difficulty, whereas initial difficulty may interact with similarity to determine ultimate performance (a).

A secondary goal was to test whether sharing individual elements or sharing pairs of elements is sufficient to drive the similarity effect. In Experiment 1, chords that shared pairs of elements with other chords were also composed of more frequent elements. The between-subjects design allowed us to create a grammar in which the element probabilities for the chord types were matched as closely as possible. Thus, differences in performance between the sparse and dense grammars could not be attributed to element frequencies but rather must be attributed to associations among the elements. If no density effect is observed in Experiment 2, it is

⁴ Each of the participants' data were also fitted to the power function using the same procedure. The exponential function provided a better fit than the power function as the RMS error was significantly lower for the exponential function than the power function (27.82 vs. 32.63), $t(83) = 3.44$, $d = 0.33$, $d_{\text{unb}} = .33$, $p < .001$. We can also evaluate the functions on the basis of the actual fitted values, for example, if the a parameter (asymptote) is estimated to be lower than is reasonable. In this regard, the exponential function estimated the a value to be less than 200 ms for less than 1 chord per participant, compared to 8 chords per participant for the power function. Thus, the exponential fits also achieved more reasonable estimates.

Table 2
Individual Values for Parameters c , b , and a and R^2 for Experiment 1

Participant	Rate of learning (c)		Amount of learning (b)		Asymptote (a)		Goodness of fit (R^2)	
	Dense	Sparse	Dense	Sparse	Dense	Sparse	Dense	Sparse
1	0.34	0.34	985	1161	901	742	0.99	0.99
2	0.41	0.28	702	890	857	606	0.98	0.97
3	0.52	0.29	1079	1322	817	424	0.98	0.99
4	0.78	0.42	791	757	940	826	0.93	0.97
5	0.55	0.54	1050	1059	663	643	1.00	1.00
6	0.35	0.61	870	1145	557	637	0.99	0.99
Mean	0.49	0.42	913	1056	789	646	0.98	0.99
SEM	0.07	0.06	67	90	67	61	0.01	0.004

Note. SEM = standard error of the mean.

possible that simply sharing individual items with other chords counts for similarity and drives the density effect observed in Experiment 1. However, if a similarity effect is observed in Experiment 2, then it would indicate that similarity in terms of the number of shared pairs is sufficient to drive the effect.

Moreover, treating similarity as a between-subjects variable had two added benefits. First, it enabled us to keep the number of practiced chords the same as in Experiment 1 while adding another factor without having a very small number of chords per condition. Second, because all of the chords performed by a given participant were equivalently similar, we were able to test whether the density effect was driven by relative novelty. Logan (1988) demonstrated that more memorable items are acquired at a faster rate than less memorable items. Across trials, the sparse chords are more likely to be more memorable as they are more different from the chords appearing on prior trials than are the similar chords. As a result, participants may have directed more attention to them. Thus, we manipulated similarity between participants. If each chord practiced by a given participant had the same similarity, then differences in the amount of learning could not be attributed to more attention being directed to particular chords. Initial difficulty was manipulated as a within-subject variable.

To construct the chord sets, we first generated two sets of grammars, one dense and one sparse, to be used by different groups of participants (see Table 1 for the grammar; see Appendix for the mappings). For both grammars, each element appeared in four of the 14 chords, except for two elements that appeared in five chords. Next, we selected mappings for the grammars using an exhaustive search of the chord library so that the fastest seven chords in a set would be expected to produce RTs 300 ms shorter than the remaining seven slowest chords for both the sparse and dense sets.

Method

Participants. Sixteen participants (eight in each group) from the Iowa City area completed eight sessions of the experiment. Each participant was paid \$10 per session. Participants performed at least four sessions per week and not more than one session per day.

Stimuli and procedure. The stimuli and procedure were the same as in Experiment 1 except that different chords sets were used. The participants were divided into two groups, one receiving dense sets of chords and the other receiving sparse sets of chords. Each set was used by two participants.

Results and Discussion

The proportion of correct trials was high throughout the experiment for all conditions (above 90% for all four conditions). An ANOVA on accuracy showed no difference between the sparse and dense chords, $F < 1$, but there was a difference between easy and difficult chords, $F(1, 14) = 28.92$, $MSE = .002$, $\eta_p^2 = .67$, $\omega^2 = .15$, $p < .0001$, as participants were more accurate in producing easy chords than hard chords. There was also an effect of session indicating improved accuracy over practice, $F(7, 98) = 13.93$, $MSE = .002$, $\eta_p^2 = .50$, $\omega^2 = .48$, $p < .001$. The Session $\times$ Density interaction was not significant, $F < 1$, but the Difficulty $\times$ Session interaction was significant, $F(7, 98) = 8.12$, $MSE < .001$, $\eta_p^2 = .37$, $\omega^2 = .27$, $p < .001$, and the three-way interaction was also significant, $F(7, 98) = 2.24$, $MSE < .001$, $\eta_p^2 = .14$, $\omega^2 = .05$, $p < .05$. Early in training, the hard chords (both sparse and dense) show the greatest improvement and the easy dense chords show the least improvement as performance was already high for this set of chords. Neither of these interactions are significant, $F < 1$, when only the final five sessions are considered, indicating the interactions were driven by early differences between easy and difficult chords.

RTs were computed and trimmed as in Experiment 1. Errors due to nonsimultaneous key presses accounted for 1% of the errors, and RTs of less than 300 ms or greater than 2,000 ms accounted for 2% of the errors. Figure 3 shows RTs across the eight sessions for both the dense and sparse groups broken down by initial difficulty. RTs were submitted to a 2 (density) $\times$ 2 (difficulty) $\times$ 8 (session) mixed ANOVA with density as a between-subjects factor and difficulty and session as within-subject factors.⁵ As expected, the main effect of difficulty was significant, $F_1(1, 14) = 139.25$, $MSE = 4,249$, $\eta_p^2 = .91$, $\omega^2 = .07$, $p < .001$, $F_2(1, 24) = 19.96$, $MSE = 37,270$, $\eta_p^2 = .45$, $\omega^2 = .002$, $p < .001$. Difficult chords ($M = 1,268$) were produced more slowly than easy chords ($M = 1,168$), validating our manipulation of difficulty. The effect of session, $F_1(7, 98) = 248.37$, $MSE = 4,449$, $\eta_p^2 = .95$, $\omega^2 = .86$, $p < .001$; $F_2(7, 168) = 1,404.97$, $MSE = 700$, $\eta_p^2 = .98$, $\omega^2 = .98$,

⁵ Before running this ANOVA, as in Experiment 1 we determined if there were any systematic differences between the three grammar $\rightarrow$ key presses mappings by running the same ANOVA with mapping as an additional between-subjects factor. No differences were found across the different mappings for either the sparse or dense set of chords, $F_s < 1$, so we averaged across this factor.

$p < .001$, was also significant, demonstrating learning. The Difficulty $\times$ Session interaction, $F_1(7, 98) = 9.43$, $MSE = 375$, $\eta_p^2 = .40$, $\omega^2 = .03$, $p < .001$; $F_2(7, 168) = 5.13$, $MSE = 700$, $\eta_p^2 = .18$, $\omega^2 = .003$, $p < .001$, and the three-way interaction, $F_1(7, 98) = 8.00$, $MSE = 375$, $\eta_p^2 = .36$, $\omega^2 = .02$, $p < .001$; $F_2(7, 168) = 5.36$, $MSE = 700$, $\eta_p^2 = .18$, $\omega^2 = .003$, $p < .001$, were significant, indicating differences in the response to practice across the four conditions.

To determine the precise nature of the effects, we again fitted an exponential function to the data from each chord from each participant. In the 112 fits performed (eight participants $\times$ 14 chords), a value associated with an upper or lower bound was obtained eight times for the a parameter and never for the c and b parameters. We also fit the data to the power function, and the exponential function again provided better fits overall.⁶ As in Experiment 1, the parameters were examined in ANOVAs, with similarity as a between-subjects factor and difficulty as a within-subject factor (using either participant or item as a random effect).

We first examined asymptotic performance (a). Sparse chords had a significantly lower asymptotic performance (a) than dense chords ($M_{\text{sparse}} = 585$ ms; $M_{\text{dense}} = 691$ ms); $F_1(1, 14) = 5.43$, $MSE = 16,470$, $\eta_p^2 = .28$, $\omega^2 = .23$, $p < .05$; $F_2(1, 28) = 25.43$, $MSE = 3,080$, $\eta_p^2 = .51$, $\omega^2 = .48$, $p < .001$ (Figure 4; see Table 3 for individual asymptotes). There was no effect of difficulty ($F_s < 1$) as easy and hard chords had similar a parameters ($M_{\text{easy}} = 630$ ms; $M_{\text{hard}} = 646$ ms), and the interaction was not significant, $F_s < 1$.

We next analyzed the amount of learning, b . Contrary to what was found in Experiment 1, the amount of learning was not different for sparse and dense chords ($M_{\text{sparse}} = 995$ ms; $M_{\text{dense}} = 962$ ms; $F_s < 1$; see Table 3 for individual values). There was a main effect of difficulty as hard chords (1,021 ms) had larger b values than easy chords (936 ms), $F_1(1, 14) = 8.85$, $MSE = 7,141$, $\eta_p^2 = .38$, $\omega^2 = .30$, $p = .01$; $F_2(1, 28) = 7.31$, $MSE = 7,312$, $\eta_p^2 = .23$, $\omega^2 = .17$, $p < .05$. The hard chords started with much higher RTs (across both dense and sparse) and ultimately reached similar asymptotes. Finally, the interaction between production difficulty and similarity was marginally significant, $F_1(1, 14) = 3.56$, $MSE = 7,141$, $\eta_p^2 = .19$, $\omega^2 = .09$, $p = .08$; $F_2(1, 28) = 2.89$, $MSE = 7,312$, $\eta_p^2 = .11$, $\omega^2 = .05$, $p = .1$, with a larger difference

Figure 3. Mean reaction times across the eight sessions for dense-easy/hard and sparse-easy/hard chords along with the fits to the exponential function for Experiment 2. The lines indicate the best fit with an exponential function.

Figure 4. The a , b , and c parameters for Experiments 1–3. The error bars represent standard error of the mean. Exp = Experiment; E = easy; H = hard.

between easy–sparse and easy–dense chords than difficult–easy and difficult–hard chords.

Finally, we examined the learning rate, c . The difference between sparse and dense chords ($M_{\text{dense}} = 0.44$; $M_{\text{sparse}} = 0.57$; see Table 3 for individual values) was not significant by participants, $F_1(1, 14) = 2.87$, $MSE = .043$, $\eta_p^2 = .17$, $\omega^2 = .06$, $p = .11$, although it was by items, $F_2(1, 28) = 23.94$, $MSE = .005$, $\eta_p^2 = .50$, $\omega^2 = .39$, $p < .01$. There was also a significant difference between easy and hard chords, $F_1(1, 14) = 10.48$, $MSE = 0.004$, $\eta_p^2 = .43$, $\omega^2 = .32$, $p < .01$; $F_2(1, 28) = 8.04$, $MSE = .005$, $\eta_p^2 =$

⁶ Each of the participants' data were also fitted to the power function as in Experiment 1. The exponential function had a significantly lower RMS error than the power function (28.08 vs. 29.68), $t(223) = 2.47$, $d = 0.11$, $d_{\text{unb}} = .11$, $p = .01$. Again as with Experiment 1, the exponential function provided a better qualitative fit as well. Less than 1 chord per participant had a values less than 200 for the exponential function, whereas the power function had on average 7 chords per participant that had a values less than 200.

Table 3

Estimated Rate and Amount of Learning, Asymptote, and Goodness of Fit for Experiment 2

Participant	Rate of learning (c)		Amount of learning (b)		Asymptote (a)		Goodness of fit (R^2)	
	Easy	Hard	Easy	Hard	Easy	Hard	Easy	Hard
Dense								
1	0.49	0.35	1310	1222	774	782	0.99	0.99
2	0.53	0.36	1223	1333	741	585	0.98	0.98
3	0.58	0.49	833	868	717	705	0.99	0.98
4	0.43	0.33	1154	1353	711	603	0.99	0.99
5	0.46	0.39	751	790	550	596	0.99	0.99
6	0.55	0.66	770	807	736	883	0.96	0.95
7	0.25	0.39	921	817	539	756	0.95	0.92
8	0.41	0.45	870	902	651	720	0.98	0.98
Mean	0.46	0.43	979	1011	677	703	0.98	0.97
SEM	0.04	0.04	82	93	34	40	0.01	0.01
Sparse								
9	1.12	0.90	881	1208	479	538	0.98	0.99
10	0.56	0.42	744	894	584	643	0.99	0.99
11	0.52	0.47	764	814	458	544	0.98	0.98
12	0.69	0.55	1138	1070	698	706	0.98	0.99
13	0.47	0.38	1055	1132	670	547	0.99	0.99
14	0.56	0.43	957	1303	526	424	0.99	1.00
15	0.51	0.43	807	933	745	769	0.98	0.98
16	0.63	0.50	764	918	511	513	0.99	0.99
Mean	0.63	0.51	889	1034	584	585	0.99	0.99
SEM	0.08	0.06	55	64	40	43	0.002	0.002

Note. SEM = standard error of the mean.

.25, $\omega^2 = .12$, $p < .01$, as easy chords were acquired at a faster rate than hard chords. The interaction between difficulty and similarity was not significant, $F_1(1, 14) = 2.94$, $MSE = 0.004$, $\eta_p^2 = .17$, $\omega^2 = .07$, $p = .11$; $F_2(1, 28) = 2.25$, $MSE = .005$, $\eta_p^2 = .09$, $\omega^2 = .02$, $p > .1$, with the difficulty manipulation affecting sparse chords more than dense chords.

In sum, similarity affected asymptotic RT rather than the magnitude of the RT decrease. It is impressive that while the difficult chords initially were performed 113 ms more slowly than the easy chords, the estimates of asymptotic RT were highly similar for the two levels of difficulty (easy = 629 ms vs. hard = 646 ms).⁷ Thus, it appears that the factors determining difficulty early in practice may be distinct from those determining difficulty later in practice (see Schmidt & Bjork, 1992). That is, practice may effectively address factors that relate to the difficulty of an item when considered in isolation, such as the unusual pairings of particular effectors within a chord, but lead to other forms of similarity-based interference between learned items. We return to this issue in the General Discussion.

Note that the similarity effect was obtained even though we equated the frequency of the individual key presses across the sparse and dense chords, so the frequency of the elements does not appear to drive the effect; the similarity effect depends on shared combinations of elements, not individual elements.

We should also note that the similarity effect was obtained even though density was manipulated between participants, so it does not stem from relative differences between items within the learning set. While these findings indicate that the differences in learning do not stem from preferential attention to one set of stimuli

relative to another, it is possible that more attention is allocated to dissimilar chords regardless of the context. That is, participants in the dense group may have been attending to all of the stimuli to a greater degree than participants in the sparse group. If so, then similarity in an absolute sense modulates attention; when all items are dissimilar, increases in attention persist throughout the experiment. We return to this issue in Experiment 3.

The pattern of results clearly supports the hypothesis that similarity affects asymptotic performance. This pattern is consistent with the proposal that similarity interferes with online processing, given that it acts to set a lower bound on RT. That is, according to this hypothesis, similarity affects the processes engaged during performance of the task (as opposed to the learning processes governing encoding), and these processes change as the chords are learned. This hypothesis is directly tested in Experiment 3.

In contrast to Experiment 1, the c parameter (learning rate) was larger for sparse chords than for dense chords, and this difference was most pronounced for the easy chords. It may be that when initial difficulty is low, individuals can reach the asymptotic levels of performance quickly, particularly when there is little competition among chords to overcome. When the chords are more difficult to produce and overlap with other chords, learning may be slower as these factors may require practice to overcome. However, it is unclear why no such effect was observed in Experiment 1.

Experiment 3

Experiments 1 and 2 indicate that similarity primarily affects asymptotic RT rather than the amount of learning. This implies an answer to our locus question: the effect may derive from differences in real-time processing rather than differences in the rate at which chords are learned. One of the goals of Experiment 3 is to test this more directly by manipulating how chords are cued.

The domain question also remains to be answered. The similarity effect may emerge from the perceptual processing of the visual stimulus or from the generation of motor commands. With respect to perception, similar chords will share stimulus elements as well as response elements. That is, they will have an X in a shared position of the display (e.g., if both chords use the third digit, an X in the third slot), and as a result, this common stimulus element will be associated with multiple internal chord representations. Thus, when a particular stimulus (pattern of X s and dashes) is observed, these competing representations may be partially activated, and the competition must be resolved before the correct motor commands can be generated. Similarly, once the perceptual system disambiguates the chord stimulus, associations between the motor commands associated with the individual fingers may be strong, given that the same fingers participate in each chord. Therefore, associations between frequently coactive motor commands may result in competition during production. In both cases, one would expect these competitive effects to be small or nonexistent early in learning when little has been encoded about either chord and grow as the associations between internal chord representations and either visual stimuli or key presses are formed. In

⁷ This pattern is not obvious in Figure 3 because it does not depict the asymptotic RTs and because it is difficult to visually average the hard chords and average the easy chords across the two levels of density.

sum, competition arising during encoding or production could give rise to changes in the effect of similarity.

It is also possible that there are competitive interactions among central operations. By *central operations*, we refer to operations that depend on both stimulus and response properties and thus do not fit entirely into perceptual or motor stages of processing. For example, the time required to make a particular response to a particular stimulus depends not only on the properties of the stimulus and the properties of the responses but also on the set of mappings in which the stimulus–response association is embedded (e.g., Fitts & Seeger, 1953; Luce & Pisoni, 1998, for an analogous argument in language processing). Such differences in the time to make the same response to the same stimulus in different contexts must be attributed to changes in the duration of central operations, which presumably include the retrieval and implementation of the stimulus–response association but do not include stimulus classification and response production.

The duration of response selection processes may be more affected by practice than stimulus or response processes (see Pashler & Baylis, 1991; Ruthruff, Johnston, & Van Selst, 2001), so central operations may play a critical role in the patterns of RTs observed in the present experiments. If differences in the asymptote stem from interactions between central operations, then performance of the sparse and dense chords may depend on the amount of practice, the nature of the learned representations, and the task demands of the current trial. That is, the performance differences do not stem from interference between stimuli or interference between responses, but, rather interference occurring during central operations that mediate the translation of the stimuli into the responses (e.g., central representations of specific chords). Therefore, learning may cause changes in processing that are latent with one set of stimuli or responses, but emerge when another set of stimuli or responses is used. Such a pattern of data is possible because the learning-related changes are not associated solely with the particular practiced stimuli or responses but instead reflect changes in central operations or representations whose effects are determined by the current stimuli and responses.

With the task used in Experiments 1 and 2, we could not disentangle these possibilities (interference arising from perception, production, or central operations) because similarity among chords was preserved in both the stimuli and the responses. Thus, the data were consistent with all three potential domains of the learning-related changes. To differentiate among these alternatives, we manipulated the overlap at the response end while holding overlap at the stimulus end constant. This was achieved by asking participants to map arbitrary Chinese characters (Figure 5) to the chord responses. Chinese characters were randomly assigned to each chord for each participant, making it unlikely that any spurious similarity among the visual stimuli would be related

systematically to similarity among the responses. If the dense chords showed a larger asymptote than the sparse chords, as in Experiments 1 and 2, then similarity among either the responses or the central representations (i.e., representations including both stimulus and response information) must be responsible for the changes in the learning curve. Alternatively, if the asymptotes were equivalent for the two tasks, perceptual similarity (competition during visual encoding) must be driving the differences in the asymptote.

In this experiment, we returned to the more powerful within-subject design of Experiment 1 (along with the same chords). After the learning phase, participants completed transfer sessions to test whether central operations play a role in the asymptote differences. In these sessions, the stimuli were the spatial displays used in Experiments 1 and 2 (with which participants had minimal experience), in which similar motoric responses were cued by similar stimuli. New chords were included to serve as a control to confirm that the transfer of learning from the Chinese characters to the spatial stimuli for the trained chords. Demonstrating the presence of chord-specific learning is necessary for interpreting the data from the practiced chords, especially if no differences are observed between practiced sparse and practiced dense chords in the transfer sessions.

The presence of asymptote differences with the Chinese characters would indicate that the similarity effect is *not* stimulus based. If the differences stem from interference among learned representations of the responses, then similar differences should be observed in the transfer sessions when the responses remain the same. However, if asymptote differences are observed with the Chinese characters but not in transfer sessions, then we would conclude that the similarity effect reflects learning at a central level, because the differences would depend partly on the stimuli used during training, even if stimulus similarity does not play a role.

Alternatively, the absence of differences in the asymptote with the Chinese characters would indicate that the similarity effect is *not* response based, given that the responses are the same with the Chinese characters and spatial stimuli. If the asymptote differences stem from interference from learned representations at a purely perceptual level, then no differences between dense and sparse chords should be observed during either practice or the transfer sessions. This is what we predicted because similarity among the Chinese characters is presumably orthogonal to similarity among the motor responses (given that the mappings were randomized across participants). Thus, interference among the chords should be independent of their motoric similarity, and no differences in asymptote should be observed.

Note that in Experiment 1, no differences were observed with the spatial stimuli in the first session, so differences observed in the transfer sessions (but not in the practice sessions) would indicate that learning reflects changes in representations that are independent of the spatial stimuli. Therefore, if the Chinese characters produce no differences in the asymptote but the spatial stimuli show performance differences between the sparse and dense chords, then we would conclude that learning alters the amount of competition between central representations and that this competition between central representations is triggered by overlapping perceptual inputs. That is, the learned representations are not based solely on stimulus properties, but they are activated

鯪 杼 欸 莽 敲 麼 螞
 馭 在 是 到 得 些 糊

Figure 5. Chinese characters used in Experiment 3.

by stimulus properties, and competing representations can be activated by similar stimuli.

Method

Participants. Six participants from the Iowa City area completed eight sessions of the experiment. Each participant was paid \$10 per session. Participants performed at least four sessions per week and not more than one session per day. None of the participants spoke Chinese or could read any of the characters used in the Experiment.

Stimuli and procedure. Fourteen Chinese characters were mapped to the same 14 chords used in Experiment 1 (Figure 5). The procedure was the same as in Experiment 1 except for the use of the Chinese character stimuli and the feedback on incorrect responses. For each incorrect response, feedback was given so that participant would know not only that they produced the incorrect chord but also what the correct chord response was. The feedback displays were the same spatial stimuli as in Experiments 1 and 2 to convey the appropriate chord response. To familiarize participants with the correct response to the ambiguous Chinese characters, in the first six blocks of Session 1 and first two blocks of Session 2, participants were shown the Chinese character along with the spatial display of the chord. On the remaining blocks, only the Chinese characters were presented. Only these blocks with the Chinese character presented alone were used in the analyses.

After the eight training sessions, two transfer sessions (Sessions 9 and 10) were conducted in which the chords were signaled only by spatial stimuli as in Experiments 1 and 2. A set of 28 equally probable chords was used. These included the 14 chords used during training and a set of 14 new sparse chords selected from the library of chords so that the difficulty (indexed by mean RT) was approximately equal for all the new chords. It was necessary to make all the new chords sparse so as to not add different amounts of overlap with the new chords to the old chords. The new sparse chords shared only one pair of key presses with the other new chords and three pairs with the old chords, except for one chord that shared five pairs and one chord that shared one pair (see Table 4 for the grammar used to generate these new chords; see Table 1 for the grammars for the old chords).

Results and Discussion

An ANOVA examining the effects of session and density on accuracy revealed an effect of session indicating that participants become more accurate over the eight sessions, $F(1, 35) = 4.35$, $MSE = .029$, $\eta_p^2 = .47$, $\omega^2 = .80$, $p < .01$, and an effect of density, $F(1, 5) = 9.05$, $MSE = .001$, $\eta_p^2 = .64$, $\omega^2 = .57$, $p < .05$, with sparse chords showing greater accuracy ($M = .92$) than dense chords ($M = .90$). The interaction between session and density was also significant, $F(7, 35) = 2.33$, $MSE = .001$, $\eta_p^2 = .32$, $\omega^2 = .61$, $p = .05$; however, both of the main effects of density and the interaction become nonsignificant if one is evaluating only the final six sessions, $F < 1$. The lower accuracy for the early portion of the experiment is primarily driven by one participant who had 27% correct across the first two sessions; however, across the remaining six sessions, the accuracy for that participant was 96%. Thus, for all participants, the mean percentage correct was 96% over the final six sessions for dense chords and 97% for sparse chords, consistent with the previous two experiments.

Table 4

Grammars Used in Sessions 9 and 10 for the New Sparse Chords in Experiment 3

Sparse grammar
ABE (1,3)
ACD (1,3)
ACJ (1,3)
AEG (1,3)
AFI (1,3)
BDJ (1,1)
BFG (1,5)
BGI (1,3)
CFH (1,3)
CGH (1,3)
DEH (1,3)
DIJ (1,3)
EFI (1,3)
EHJ (1,3)

Note. The first number in parentheses shows the number of other new chords that share a pair with that chord and the second number is the number of old chords that share a pair with that chord.

RTs were computed and trimmed as in Experiments 1 and 2. Errors resulting from RTs of less than 300 ms or greater than 2,000 ms accounted for 5% of the errors and nonsimultaneous key presses accounted for 1% of the errors. RTs were submitted to a 2 (density) $\times$ 8 (session) within-subject ANOVA.⁸ There was a main effect of session, $F_1(7, 35) = 64.47$, $MSE = 7,131$, $\eta_p^2 = .93$, $\omega^2 = .99$, $p < .001$; $F_2(7, 84) = 274.63$, $MSE = 1,942$, $\eta_p^2 = .96$, $\omega^2 = 1.00$, $p < .001$, but neither density nor the interaction between density and session was significant, $F_s < 1$. That is, there was no evidence that participants learned the sparse and dense chords differently. RTs for the dense and sparse chords were 1,360 ms and 1,335 ms, respectively, on the first session, $t_s < 1$, and 805 ms and 810 ms, respectively, on the eighth session, $t_s < 1$, suggesting that the asymptotes for both chord types is not statistically different (see Figure 6).

Again, we fitted the RT data from each chord for each participant to exponential and power functions, and the exponential function again provided a better fit.⁹ The sparse and dense chords had asymptotes, a , that did not differ, $t_s < 1$, and the trend was in the opposite direction as the significant effect in Experiments 1 and 2; the dense chords had smaller asymptotes ($M_{\text{dense}} = 614$ ms; see Table 5 for individual asymptote values) than the sparse chords ($M_{\text{sparse}} = 644$ ms). These asymptote values are comparable to those observed in the previous experiments, as are the overall RTs, indicating that performing the task with the Chinese characters did

⁸ As in Experiments 1 and 2 we determined if there were any systematic differences between the mappings, by running the same ANOVA with mapping as an additional factor. No differences were found, $F_s < 1$, so we averaged across this factor.

⁹ The RTs from each chord for the participant were also fitted to the power function as in Experiments 1 and 2. The exponential function provided a better fit than the power function. The exponential function had a significantly lower RMS error than the power function (52.54 vs. 59.59), $t(83) = 5.04$, $d = 0.18$, $d_{\text{unb}} = .18$, $p < .001$. Again the exponential fit provided a better qualitative fit as well. The exponential fit had on average only 2 chords per participant with a values less than 200, whereas, the power function had on average 7 chords per participant with a values less than 200.

Figure 6. Mean reaction times across the 10 sessions for dense and sparse chords along with the fits to the exponential function for Experiment 1. The lines indicate the best fit with an exponential function.

not change the estimates of practiced performance despite the much more arbitrary stimulus–response mapping. Furthermore, the amount of learning (b), $t_1(5) = 1.10$, $d = 0.49$, $d_{\text{unb}} = 0.41$, $p = .33$; $t_2(12) < 1$, and rate of learning (c), $t_1(5) = 1.2$, $d = 0.49$, $d_{\text{unb}} = 0.41$, $p = .27$; $t_2(12) = 1.20$, $d = 0.65$, $d_{\text{unb}} = 0.61$, $p = .26$, did not differ between sparse and dense chords (see Table 5 for individual values).

These results from the training sessions are consistent with the claim that the effect of similarity observed in Experiments 1 and 2 is due to competition during the perceptual processing of the stimulus and not to competition from similar responses. In the dense condition, particular stimulus features (e.g., an X in a particular location) were associated with several chords, so that the presence of each feature may have activated multiple chord representations, giving rise to competition. In contrast, in the sparse condition, these features were mapped to a smaller number of chords such that any stimulus only activated a single chord strongly, which diminished competition. Because competition is only evident when the overlapping stimulus features are shared by chords and thus activate multiple central representations, the Chinese characters, which are uniquely associated with responses, do not produce a similarity effect. This explains why no similarity effect is observed early in practice; without training, there are no central representations to be coactivated, even when the chords are signaled by overlapping sets of stimulus features.

In this light, the results from the transfer sessions conform that the competition stems from changes that are not restricted to perceptual-level processing. We first analyzed the RTs during transfer for the old chords that were practiced during Sessions 1–8, this time cued with spatial arrays, rather than Chinese characters. We performed a 2×2 ANOVA to examine RT as a function of similarity and session (nine vs. 10). There was an effect of session, $F_1(1, 5) = 91.95$, $MSE = 1,097$, $\eta_p^2 = .95$, $\omega^2 = .89$, $p < .001$; $F_2(1, 12) = 150.63$, $MSE = 797$, $\eta_p^2 = .93$, $\omega^2 = .92$, $p < .001$, showing that RTs diminished from the first transfer session to the second. Most important, we found a significant effect of similarity for the participant analysis, $F_1(1, 5) = 6.82$, $MSE = 2,807$, $\eta_p^2 = .58$, $\omega^2 = .06$, $p < .05$, indicating that sparse chords (1,051 ms) were performed more quickly than dense chords (1,104 ms) once perceptual competition was introduced to the task. The difference between dense and sparse grammars was not significant in the item

analysis, $F_2(1, 12) = 1.40$, $MSE = 16,595$, $\eta_p^2 = .10$, $\omega^2 = .003$, $p = .26$, likely due to reduced power, as similarity was now a between-item effect, and there were few repetitions/subjects for each item. The interaction between session and similarity was not significant, $F_s < 1$.

The difference between the *old* dense and the *old* sparse chords was 56 ms across the two transfer sessions. Because no difference was observed in the initial session of Experiment 1, the observed differences in the transfer sessions of Experiment 3 most likely stemmed from the practice in Sessions 1–8. Thus, even though no differences were observed throughout practice with the Chinese characters, when these same chords were cued with spatial stimuli, the effect of similarity was immediately evident.

Independent confirmation that training with Chinese chords transfers to spatial stimuli comes from the consideration of the 14 *new* chords included in Sessions 9 and 10. Given that we found differences between sparse and dense chords when they are signaled by the spatial stimuli, it was important to demonstrate that these differences were the result of practice and not a direct result of the spatial stimuli. Therefore, it was necessary to show chord-specific learning (a difference between the *old* and *new* chords). As expected, the mean RT for the new chords was 1,146 ms, which was significantly slower than the sparse-*old* chords, $t(5) = 5.44$, $d = 0.62$, $d_{\text{unb}} = 0.52$, $p < .01$, and marginally significant for dense-*old* chords, $t(5) = 2.17$, $d = 0.27$, $d_{\text{unb}} = 0.23$, $p = .08$.

This finding addresses the locus question by indicating that training builds central representations that can be activated by multiple stimuli. Spatial stimuli, because they depict components of multiple chords (when they signal a dense chord), can activate multiple chord representations. When multiple chord representations are activated by a stimulus, RT is prolonged. Thus, with spatial stimuli, performance is sensitive to both the chord set that was trained and the stimulus currently used to cue the appropriate response. That is, the locus appears to be in the real-time processes by which central representations are activated by the perceptual input.

In order to pin these effects on learning during the practice sessions, we confirmed that these RT differences did not arise from any inherent differences between the specific chords that were used in this experiment. That is, these differences should not be observed when participants performed the same task with the same chords but without any prior learning. Thus, we tested an additional 12 participants in a single session using the exact same procedure as in Sessions 9 and 10 (all 28 chords without prior training with Chinese stimuli).¹⁰ We kept the distinction of old and new chords to be consistent with the analyses of Experiment 3 and to confirm that the effects observed in transfer sessions of that experiment do not stem from the particular grammars we chose. However, to be clear, for this single-session experiment, no previous training had occurred for any of the chords.

There was no difference in performance between the sparse-*old* ($M_{\text{sparse}} = 1,352$ ms) and the dense-*old* ($M_{\text{dense}} = 1,349$) chords, $t < 1$, confirming that when participants had not allowed to

¹⁰ This sample size was larger than in the other experiments because we wanted a more sensitive test of whether any differences might exist between the sparse and dense chords early in practice.

Table 5
Individual Values for Parameters c , b , and a and R^2 for Experiment 3

Participant	Rate of learning (c)		Amount of learning (b)		Asymptote (a)		Goodness of fit (R^2)	
	Dense	Sparse	Dense	Sparse	Dense	Sparse	Dense	Sparse
1	0.41	0.56	1133	1514	662	534	0.99	0.98
2	0.93	0.83	1122	1080	722	723	0.99	0.98
3	0.30	0.57	1235	950	313	647	0.95	0.99
4	0.38	0.17	1034	1266	743	401	0.95	0.84
5	0.50	0.71	988	1031	695	822	0.90	0.94
6	0.34	0.64	1417	1788	548	739	0.96	0.98
Mean	0.48	0.58	1155	1270	614	644	0.96	0.95
SEM	0.09	0.10	79	143	128	69	0.03	0.03

Note. SEM = standard error of the mean.

practice, density had no effect on RT for these particular chords.¹¹ Furthermore, there was no difference between the sparse-new ($M = 1,354$ ms) and the dense-old chords or sparse-new and sparse-old chords, $t_s < 1$. This was expected because little had been learned about either set (old or new) of chords.

These findings and our interpretation are consistent with Masson (1986). Masson did not examine similarity but did show that practicing reading typographical mirror-reversed words did not transfer to words composed of novel letters and transferred less to novel words composed of practiced letters than to practiced words (see also Crump & Logan, 2010; Hazeltine et al., 2007). Masson concluded that the features of the current word may trigger an instance-based retrieval process that partially benefits novel words with overlapping features. This conceptualization fits with our interpretation of the similarity effect, especially if one assumes that in Masson's reading task, the central representations of the words are well established before participants begin the experiment. These findings were extended to a typing task by Crump and Logan (2010), indicating that the framework holds when the task involves greater motor demands.

The similarity effect observed in the transfer sessions (56 ms) is nearly identical to the similarity effect observed in Experiment 1 (63 ms), suggesting that whether the stimuli were learned as spatial stimuli or Chinese characters did not affect the online encoding processes that produce the differences in the performance of dense and sparse chords when the stimuli are spatial. This finding argues against an attentional account of the advantage for dissimilar chords. According to the attentional account, items in the sparse set were considered more novel and therefore would have received more attention resulting in faster learning rates. However, in Sessions 1–8, no differences existed in the perceptual cues for sparse and dense chords, providing no reason for them to have received different amounts of attention during training. Nonetheless, when the participants performed the chords with the spatial stimuli, the difference in performance for the sparse and dense chords was immediately apparent. These differences must have been the result of the practice with the Chinese stimuli, because no such differences were observed during Session 1 or during the control experiment. In sum, it appears that the learned representations can be accessed by fairly novel stimuli.

This finding strongly suggests an answer to the locus question. The differences in performance between similar and dissimilar practice chords appear to relate to changes in online process-

ing—in this case activation and retrieval of the appropriate chord representation—rather than impaired learning for the similar items. In this way, the experiment provides concrete evidence for a locus in processes like retrieval (as proposed by Palmeri, 1997). The magnitude of the similarity effect observed for the practiced chords signaled by the spatial stimuli in the transfer sessions in Experiment 3 is nearly identical to that observed in Experiment 1, suggesting that learning proceeded in the same way for the sessions with the spatial and Chinese character stimuli. Nonetheless, during these training sessions, performance was markedly different for the two stimulus sets, with the Chinese characters showing no evidence of a similarity effect. Thus, it appears that although learning was similar under the two conditions, signaling the chords with spatial stimuli led to online competition producing the differences in performance. When the cues were dissimilar, the same learned responses were retrieved without interference.

General Discussion

The present study examined how the relational properties of individual practice items affect skill acquisition. We focused on three questions. First, does similarity facilitate or hinder performance? Second, if similarity does affect performance, what is the domain of the effect: perceptual, central, or motor processes? Finally, is the locus of effect expressed in the rate that items are acquired or in real-time processing? These questions were addressed by manipulating stimulus and response similarity and observing the effects on the parameters of the exponential fits to the data. The findings are summarized in Table 6.

In Experiment 1, we manipulated similarity within participants and found that similar (dense) chords had a higher asymptotic RT

¹¹ It is possible that differences observed in Sessions 9 and 10 might result from the spatial displays used in some of the early blocks in Sessions 1 and 2. However, there was no difference in performance between sparse and dense chords ($M_{\text{sparse}} = 1,460$ vs. $M_{\text{dense}} = 1,455$ ms) during the practice blocks with the spatial stimuli, $t < 1$ (the first six blocks of Session 1 and first four blocks of Session 2). Moreover, we can compare the data from Session 2 of Experiment 1, in which participants had approximately the same amount of exposure to the sparse and dense chords cued by spatial stimuli as in Sessions 9 and 10 of Experiment 3. As expected no differences existed between sparse and dense chords ($M_{\text{sparse}} = 1,120$ vs. $M_{\text{dense}} = 1,131$), $t < 1$. Thus, the differences observed in Sessions 9 and 10 appear to arise from practice with the Chinese character stimuli. We thank an anonymous reviewer for pointing out this possibility.

Table 6
Summary of Experiments 1–3

Experiment	Manipulation	Result
1	Within participants: stimulus + response similarity	Dense items: $\uparrow a$, $\downarrow b$
2	Within participants: initial difficulty	Easy chords: $\downarrow b$, $\uparrow c$
	Between participants: stimulus + response similarity	Dense chords: $\uparrow a$
3	Within participants: response similarity only	Dense items = sparse items
	Transfer: stimulus + response similarity	Dense items: $\uparrow RT$

Note. Parameters a , b , and c are from the fits of exponential functions to the reaction time (RT) data. a = asymptotic RT; b = change in RT; c = learning rate.

than dissimilar (sparse) chords when initial RT was controlled. Thus, the difference in RT for sparse and dense chords increased as practice progressed. This finding stands in contrast to the effects of other manipulations, such as stimulus–response compatibility (e.g., Van Selst et al., 1999) and the number of stimulus–response alternatives (e.g., Seibel, 1963), which have diminishing effects with practice. As practice increases, performance becomes more sensitive to interitem similarity.

In Experiment 2, we teased apart whether the similarity effect observed in Experiment 1 was primarily due to a difference in the amount of learning or a difference in asymptotic performance by manipulating initial chord difficulty within participants and chord similarity between participants. The results replicated those from Experiment 1; similarity led to increasing decrements in performance with practice using different chords, demonstrating that this effect does not stem from either preferential attention to a subset of items in the training set or the frequencies of the individual elements (digits) composing the chords. Moreover, by manipulating initial difficulty, we demonstrated that similarity primarily affects the ultimate asymptote and not the magnitude of the RT decrement. Furthermore, chords that were easier to produce were learned at a slightly faster rate than chords that were harder to produce, particularly for the dense group. Thus, both competition among items and difficulty in producing individual items appear to shape the learning curve.

In Experiments 1 and 2, perceptual and motor similarity covaried as spatially compatible Xs signaled the appropriate key presses. Therefore, to answer the domain question, in Experiment 3, we eliminated perceptual similarity by using Chinese characters instead of Xs and dashes. No differences between the sparse and dense chords were observed in the various learning parameters (a , b , and c) or in overall RTs, suggesting that competition during perceptual processing of the stimulus was driving the effect of similarity. However, when participants were transferred to spatial stimuli, RT differences between sparse and dense chords immediately appeared. Because training with one set of stimuli produced similarity effects when participants performed with another set of stimuli, we conclude that the learning-related changes occurred in central operations.

Experiment 3 also addresses the locus question. The similarity effect observed with the spatial stimuli in the transfer sessions

suggests that training with the Chinese characters produced the same learning-related changes that lead to the similarity effect, even though performance with the characters did not produce a difference between sparse and dense chords. Thus, the similarity effect did not appear to derive from differences in the rate that items are encoded over practice (a locus in learning), but rather from competition among items cued by similar stimuli (a locus in performance). Dense chords were cued by stimuli associated with several chords, producing competition among multiple central representations. Sparse chords, however, were cued by stimuli associated with fewer chords, leading to less competition. That is, sparse and dense chords were likely learned at an equal rate, but the similarity effect that emerges from the expression of knowledge develops over the course of the experiment.

Accounting for Similarity-Based Effects

Usher and McClelland (2001) proposed an accumulator model of response selection in which RT depends on the amount of time required for sufficient evidence for an individual response to be accumulated. The amount of evidence for a particular response must exceed the amount of evidence for alternative responses by a particular threshold. This account can be applied to the current task: when a dense stimulus is presented, evidence is accumulated for multiple responses, driving competition and lengthening RT. Early in training, the stimuli may provide only weak evidence for the corresponding responses and provide even less evidence for responses signaled by similar stimuli. Thus, RTs are long, but there is little competition, so similarity has little effect. After practice, the stimuli may provide evidence more robustly for chord responses that share stimulus elements, including the appropriate response. This explains why the effect of similarity increases with practice; the strengthening of the response representations allows them to become activated by similar stimuli. When the spatial stimuli for dense chords were presented, they partially activated other similar chords, driving competition and increasing RTs for the dense chords. Thus, strengthening the representations speeds performance but also increases interference among chords cued by similar stimuli.

This account can explain a puzzling aspect of the present data: the similarity effect depends on both training and similarity among the stimuli, but it does not depend on whether training involved similar stimuli. That is, training with dissimilar stimuli increases the stimulus similarity effect. Such a pattern would be expected if the similarity effect is based on online competition between central representations. The source of the competition does not relate to perceptual encoding; that is, it does not occur at a perceptual level where the individual elements are detected. Rather, it derives from the mapping of perceptual features to central representations, as in an accumulator model.

Consider two similar chords: Chord A for which the digits 1–2–3 (the left pinky, ring, and middle fingers, respectively) are used, and Chord B for which the digits 1–2–5 (the left pinky, ring, and thumb, respectively) are used. After training, at the level of perceptual encoding, when participants see spatial stimuli corresponding to 1–2–3, there is no reason to encode a 5 because there is no ambiguity in the input. However, the 1 and 2 in the input partially activate both central representations for A and B. This competition must be resolved, leading to slower RTs. In sum, the

competition is occurring between representations at a central level but is driven by perceptual similarity.

Such an account is consistent with the absence of interference when Chinese characters served as input. These stimuli do not overlap in any systematic way. Therefore, Character 1 does not partially activate both A and B; it only activates A as it has no shared features with the character corresponding to B. Thus, competition emerges not because the learned representations are similar, but because they are *accessed* by similar inputs during online perception. The conclusion that perceptual cues can activate multiple representations is consistent with work in reading typography and reading as well (Crump & Logan, 2010; Masson, 1986). In these studies, transfer was based on shared perceptual features between practiced and novel items. As we will discuss later, this mechanism is analogous to how similarity and competition are hypothesized to operate in language, and as a result, this framing may be important for clinical problems such as specific language impairment or dyslexia in which unresolved competition has been hypothesized as a locus of their deficits (Harm & Seidenberg, 1999; Mainela-Arnold, Evans, & Coady, 2008; McMurray, Samuelson, Lee, & Tomblin, 2010).

In sum, to account for similarity effects, we propose that learning in this task involves the formation of central representations that specify response features and can be cued by multiple stimuli. This conclusion is consistent with findings in other motor learning tasks, particularly the serial reaction time (SRT) task. Several SRT studies have examined how learning a motor sequence with a set of stimuli benefits performance under conditions with either different stimuli or different responses (e.g., Grafton, Hazeltine, & Ivry, 1998; Hazeltine, 2002; Willingham, 1999; Ziessler & Nattkemper, 2001; for a review, see Abrahamse, Jiménez, Verway, & Clegg, 2010). Across these studies, the broad consensus is that sequence learning is not restricted to either associations among stimuli or associations among responses but rather includes abstract information that allows for the partial transfer of learning to a range of conditions that conceptually match the training regime.¹² In sum, in both domains, learning appears to include structural relations that can be accessed by novel stimuli (see Koch & Hoffmann, 2000). Future studies of chord learning should examine transfer to novel but related motor outputs.

Relation to Instance Theory

Interference among perceptually similar chords is consistent with some versions of instance theory, particularly with the exemplar-based random walk model of Palmeri (1997; see also Nosofsky & Palmeri, 1997). This model accounts for similarity effects by assuming that an item's similarity to an exemplar determines the rate at which the item contributes evidence toward a particular category. In the present experiments, each stimulus required a unique response, so there were no categories in the classical sense. Thus, it was unclear whether similarity would benefit or hinder performance. The present findings indicate that perceptual similarity may create competition among the responses by making the retrieval of particular chords more difficult.

In this way, the difference in the asymptote for sparse and dense chords extends the Palmeri (1997) results by showing that even when the task does not demand a reduction in the dimensionality (e.g., categorization), strong competitive interference can still be

seen, and this interference arises from perceptual similarity. In other words, the similarity effect does not appear to stem strictly from the coherence of the categories but can be observed even when similar (but different) stimuli cue similar (but different) responses. Moreover, it provides a concrete test of Palmeri's assumption that the locus of similarity effects lie in real-time processes like retrieval.

However, we see no reason that an exemplar account is the only way to model our findings—the competitive effect can also be observed in a host of other architectures: prototype models, connectionist models, dynamic systems, accumulators, and so forth. Indeed, exemplar models are rarely applied to complex skills, and it is not clear that a memory-based account is appropriate for such clearly procedural knowledge. Nonetheless, Palmeri's (1997) model is quite consistent with our claim that processing factors such as encoding and production must be accounted for as well as decision and learning processes.

Relation to Chunking Theory

The finding that interitem similarity slows performance more as learning progresses is contrary to a straightforward implementation of chunking theory, although additional assumptions may address the discrepancies. According to chunking theory, the diminishing benefits of practice relate to the decreasing frequency of more specific chunks, so that RT decrements depend on the frequency of the specific chunks. In the present data, participants' performance of chords that had more frequent chunks benefitted less from practice, challenging this account. Although the effect was eliminated when stimulus similarity was held constant, there was little evidence in this case for a learning advantage for chords sharing pairs of elements; in none of the experiments did chords that shared chunks with other chords show faster learning. In fact, in Experiments 1 and 3, the individual finger presses in the dense chords were more frequent than the finger presses in the sparse chords, yet no advantage was observed for these chords, even when perceptual similarity was held constant. This suggests that chunking theory may not apply to tasks like this one, in which the online processing demands created by competition among items cued by similar stimuli may simply outweigh any specific role for chunking in learning. If processes sensitive to this form of competition are added to chunking theory, so that multiple higher order chunks compete for shared lower order components, then the present data may be accommodated by such an account.

Implications for the Power Law of Learning

While not our primary focus, these data add to the body of work suggesting that exponential (rather than power law) functions may best describe changes in RT over learning (e.g., Heathcote et al., 2000; Myung et al., 2000). Perhaps more important, the present findings show how one can disentangle the underlying sources of changes in the parameters that describe this function. Asymptotic performance, *a*, appears to be driven by differences in real-time processing, driven by the extent to which possible stimuli activate multiple learned representations. In contrast, the amount of learn-

¹² We are grateful to an anonymous reviewer for pointing out this connection.

ing, *b*, is affected by initial task difficulty. More broadly, this is perhaps a cautionary message that changes in the functions that describe performance over an interval of training may not always derive from differences in the rate of encoding (see Schmidt & Bjork, 1992).

Relation to Word Learning

Our choice of tasks and items was inspired in part by the problem of learning words. While we do not make strong claims about word learning here, there are strong parallels between chord learning and word learning. In both domains, stimuli and responses are highly complex (as well as the stimulus–response mappings). Phonological similarity among words is relevant to both perceptual processing (e.g., word recognition) and production (e.g., speech production), although it is mediated by a set of arbitrary categories. Parallel activation and real-time competition are inherent to models of both these processes (Dell, 1986; Luce & Pisoni, 1998; McClelland & Elman, 1986; Rapp & Goldrick, 2000). To the extent that language is acquired by domain general principles (Elman, 1990; McMurray, 2007; Perruchet & Pacton, 2006; Saffran, Pollak, Seibel, & Shkolnik, 2007), it is useful to consider the parallels between our results and word learning.

In the present study, we quantified similarity in terms of shared elements (fingers) much as work in word learning has quantified similarity in shared phonemes. Indeed, our use of the term *density* was inspired by the construct of lexical neighborhood density (Luce & Pisoni, 1998), where words in high-density neighborhoods have many similar sounding neighbors and words in low-density neighborhoods have few. In both language and chord learning, the effect of similarity on learning could derive from perceptual processing, learning, or production, and it is important to disentangle these. By manipulating how the representations are accessed, we show clear evidence that this effect is due to the mapping between perceptual encoding and central representations.

If we consider this as a description of posttraining performance, such findings are analogous to predictions about adult spoken word recognition from the neighborhood activation and other models of spoken word recognition (Luce & Pisoni, 1998; McClelland & Elman, 1986). In these models, similar-sounding words are partially activated by shared elements in the input, leading to competition between lexical items (not between perceptual representations) and slowing word recognition (Apfelbaum, Blumstein, & McMurray, 2011; Luce & Pisoni, 1998; Magnuson, Dixon, Tanenhaus, & Aslin, 2007; Vitevitch, 2002). This is quite similar to our account of the present findings. Intriguingly, while similarity hinders perceptual encoding in word recognition, it facilitates speech production (Vitevitch, 2002). We should have been able to observe such a benefit in Experiment 3 because similarity in perceptual encoding was equivalent across chords but similarity in production was manipulated. However, we did not find any benefit, although there was a nonsignificant trend in this direction. It is possible that this benefit could be observed with a more sensitive design.

To extend this notion to learning, we found that the effect of similarity grows over training, as a result of strengthening links between overlapping perceptual input and central representations. Only a handful of studies have examined neighborhood density in word learning. Storkel, Armbrüster, and Hogan (2006) tested the

effect of density on adults' ability to acquire novel words in a task in which the words must be produced as well as comprehended (thus, similarity may play a role in both encoding and production). They found an advantage for high-density words, but their analyses focused on the quality of the learned representations, not changes over learning, and they did not conduct a detailed analysis of the learning curve. In contrast, a number of studies on younger infants have suggested that infants have a hard time learning words that are similar to each other or are similar to a known word in a comprehension paradigm (Stager & Werker, 1997; Swingley & Aslin, 2007).

However, the results from Experiment 3 raise the possibility that such findings may derive in part from differences in perceptual encoding and, thus, may only suggest that parallel activation/competition is operating at this age (see Swingley & Aslin, 2002). While it is not clear if our specific finding will extend to language learning, it is clear that to study language (or motor) *learning*, we must disentangle it from language *use* (cf. McMurray, Horst, Toscano, & Samuelson, 2007). As a result, differences in how we measure word learning may help account for the fact that some studies appear to show slower or more difficult learning for highly similar words, and others show facilitation. More important, chord learning may offer an intriguing model system for isolating the role of input and output similarity. This distillation is extremely difficult to achieve in language, where the similarity relations between how words sound (perception) are the result of how words are produced. That is, the input and output similarity are nonarbitrary and cannot be easily disentangled.

In principle, it should be possible to harness the similarity effect to improve the expression of developing skills, including language skills. We recently examined this in the context of reading. Apfelbaum, Hazeltine, and McMurray (2012) taught first-grade children grapheme–phoneme correspondences with training sets of words that entailed either similar or dissimilar consonant frames. Children who were taught with dissimilar consonant frames (analogous to our sparse chords) showed greater learning on both practiced and novel words and on practiced and novel tasks, analogous to the advantages observed for dissimilar groups in Experiments 1 and 2. The authors concluded that variable consonant frames facilitated statistical learning by increasing variability of irrelevant statistics.

Conclusions

Understanding the factors that shape the learning curve has wide-reaching benefits, from informing the basic science of how practice changes perceptual–motor processes to guiding educational strategies and training programs. Here we examined a factor that has been given little consideration, interitem similarity. The results indicate that similarity plays a critical role in performance that increases as individuals become more practiced. Thus, the factors that determine difficulty early in training seem largely separate from those that dominate performance later in training (see Schmidt & Bjork, 1992). As an item is practiced, it can be performed more quickly, but it may also interfere with other items more readily because it is more easily activated. That is, competition appears to emerge among central representations that are coactivated by stimuli that are partially consistent with multiple practiced items. This account can be applied to a wide range of

domains, including reading and human-machine interfaces. For example, a straightforward recommendation is that devices should use signals that are as distinct as possible, even to cue events that require actions that are similar or that will become highly practiced. In such cases, it may be possible to design the tasks and change the practice regimes to yield steeper learning curves and better expert performance. For example, children learn phonics rules better if there is a greater variability among the rules compared with if the rules are very similar (Apfelbaum et al., 2012). By distilling these tasks into a domain that allows us to capture the relevant dimensionality and by manipulating the similarity inherent in both stimuli and responses, we can isolate the basic mechanisms of learning to achieve the kind of understanding necessary for such applications.

References

- Abrahamse, E. L., Jiménez, L., Verwey, W. B., & Clegg, B. A. (2010). Representing serial action and perception. *Psychonomic Bulletin & Review*, 17, 603–623. doi:10.3758/PBR.17.5.603
- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89, 396–406. doi:10.1037/0033-295X.89.4.369
- Apfelbaum, K., Blumstein, S. E., & McMurray, B. (2011). Semantic priming is affected by real-time phonological competition: Evidence for continuous cascading systems. *Psychonomic Bulletin & Review*, 18, 141–149. doi:10.3758/s13423-010-0039-8
- Apfelbaum, K., Hazeltine, E., & McMurray, B. (2012). Statistical learning in reading: Variability in irrelevant letters helps children learn phonics rules. *Developmental Psychology*. Advance online publication. doi:10.1037/a0029839
- Crossman, E. R. F. W. (1959). A theory of the acquisition of speed-skill. *Ergonomics*, 2, 153–166. doi:10.1080/00140135908930419
- Crump, M. J. C., & Logan, G. D. (2010). Episodic contribution to sequential control: Learning from a typist's touch. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 662–672. doi:10.1037/a0018390
- Dell, G. S. (1986). A spreading activation theory of retrieval in language production. *Psychological Review*, 93, 283–321. doi:10.1037/0033-295X.93.3.283
- Drake, C., & Palmer, C. (2000). Skill acquisition in music performance: Relations between planning and temporal control. *Cognition*, 74, 1–32. doi:10.1016/S0010-0277(99)00061-X
- Elman, J. (1990). Finding structure in time. *Cognitive Science*, 14, 179–211. doi:10.1207/s15516709cog1402_1
- Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, 12, 499–504. doi:10.1111/1467-9280.00392
- Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *PNAS: Proceedings of the National Academy of Sciences of the United States of America*, 99, 15822–15826. doi:10.1073/pnas.232472899
- Fitts, P. M., & Seeger, C. M. (1953). S–R compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology*, 46, 199–210. doi:10.1037/h0062827
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141, 2–18. doi:10.1037/a0024338
- Gandolfo, F., Mussa-Ivaldi, F. A., & Bizzi, E. (1996). Motor learning by field approximation. *PNAS: Proceedings of the National Academy of Sciences of the United States of America*, 93, 3843–3846. doi:10.1073/pnas.93.9.3843
- Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C. H., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, 5, 236–243. doi:10.1016/S1364-6613(00)01662-4
- Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition*, 78, 27–43. doi:10.1016/S0010-0277(00)00099-8
- Grafton, S. T., Hazeltine, E., & Ivry, R. B. (1998). Abstract and effector-specific representations of motor sequences identified with PET. *Journal of Neuroscience*, 18, 9420–9428.
- Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, 106, 491–528. doi:10.1037/0033-295X.106.3.491
- Hazeltine, E. (2002). The representational nature of sequence learning: Evidence for goal-based codes. In W. Prinz & B. Hommel (Eds.), *Attention and performance* (Vol. XIX, pp. 673–689). Oxford, English: Oxford University Press.
- Hazeltine, E., Aparicio, P., Weinstein, A., & Ivry, R. B. (2007). Configural response learning: The acquisition of a nonpredictive motor skill. *Journal of Experimental Psychology: Human Perception, and Performance*, 33, 1451–1467. doi:10.1037/0096-1523.33.6.1451
- Hazeltine, E., Ruthruff, E., & Remington, R. W. (2006). The role of input and output modality pairings in dual-task performance: Evidence for content-dependent central interference. *Cognitive Psychology*, 52, 291–345. doi:10.1016/j.cogpsych.2005.11.001
- Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review*, 7, 185–207. doi:10.3758/BF03212979
- Koch, I., & Hoffmann, J. (2000). Patterns, chunks, and hierarchies in serial reaction-time tasks. *Psychological Research*, 63, 22–35. doi:10.1007/PL00008165
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95, 492–527. doi:10.1037/0033-295X.95.4.492
- Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 883–914. doi:10.1037/0278-7393.18.5.883
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear & Hearing*, 19, 1–36. doi:10.1097/00003446-199802000-00001
- MacKay, D. G. (1969). The repeated letter effect in the misspellings of dysgraphics and normals. *Attention, Perception & Psychophysics*, 5, 102–106. doi:10.3758/BF03210531
- Magill, R. A., & Hall, K. G. (1990). A review of the contextual interference effect in motor skill acquisition. *Human Movement Science*, 9, 241–289. doi:10.1016/0167-9457(90)90005-X
- Magnuson, J. S., Dixon, J., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, 31, 133–156. doi:10.1080/03640210709336987
- Mainela-Arnold, E., Evans, J. L., & Coady, J. (2008). Lexical representations in children with SLI: Evidence from a frequency manipulated gating task. *Journal of Speech, Language, and Hearing Research*, 51, 381–393. doi:10.1044/1092-4388(2008/028)
- Masson, M. E. J. (1986). Identification of typographically transformed words: Instance-based skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 479–488. doi:10.1037/0278-7393.12.4.479
- McClelland, J., & Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86. doi:10.1016/0010-0285(86)90015-0
- McMurray, B. (2007, August 3). Defusing the childhood vocabulary explosion. *Science*, 317, 631. doi:10.1126/science.1144073
- McMurray, B., Horst, J. S., Toscano, J. C., & Samuelson, L. (2009). Towards an integration of connectionist learning and dynamical systems

- processing: Case studies in speech and lexical development. In J. P. Spencer, M. S. Thomas, & J. L. McClelland (Eds.), *Toward a unified theory of development: Connectionism and dynamic systems theory reconsidered*. (pp. 218–249). London, England: Oxford University Press.
- McMurray, B., Samelson, V. S., Lee, S. H., & Tomblin, J. B. (2010). Eye-movements reveal the time-course of online spoken word recognition language impaired and normal adolescents. *Cognitive Psychology*, 60, 1–39. doi:10.1016/j.cogpsych.2009.06.003
- Morin, R. E., & Forrin, B. (1962). Mixing two types of S–R associations in a choice reaction time task. *Journal of Experimental Psychology*, 64, 137–141. doi:10.1037/h0048688
- Myung, I. J., Kim, C., & Pitt, M. A. (2000). Toward an explanation of the power law artifact: Insights from response surface analysis. *Memory & Cognition*, 28, 832–840. doi:10.3758/BF03198418
- Neves, D. M., & Anderson, J. R. (1981). Knowledge compilation: Mechanisms for the automatization of cognitive skills. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 57–84). Hillsdale, NJ: Erlbaum.
- Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Erlbaum.
- Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, 104, 266–300. doi:10.1037/0033-295X.104.2.266
- Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 324–354. doi:10.1037/0278-7393.23.2.324
- Pashler, H., & Baylis, G. (1991). Procedural learning: 1. Locus of practice effects in speeded choice tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 20–32. doi:10.1037/0278-7393.17.1.20
- Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, 10, 233–238. doi:10.1016/j.tics.2006.03.006
- Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, 107, 460–499. doi:10.1037/0033-295X.107.3.460
- Ruthruff, E., Johnston, J. C., & Van Selst, M. (2001). Why practice reduces dual-task interference. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 3–21. doi:10.1037/0096-1523.27.1.3
- Saffran, J. R., Pollak, S. D., Seibel, R. L., & Shkolnik, A. (2007). Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition*, 105, 669–680. doi:10.1016/j.cognition.2006.11.004
- Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, 3, 207–217. doi:10.1111/j.1467-9280.1992.tb00029.x
- Schumacher, E. H., Seymour, T. L., Glass, J. M., Kieras, D. E., & Meyer, D. E. (2001). Virtually perfect time sharing in dual-task performance: Uncorking the central attentional bottleneck. *Psychological Science*, 12, 101–108. doi:10.1111/1467-9280.00318
- Seibel, R. (1963). Discrimination reaction time for a 1,023-alternative task. *Journal of Experimental Psychology*, 66, 215–226. doi:10.1037/h0048914
- Servan-Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 592–608. doi:10.1037/0278-7393.16.4.592
- Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 179–187.
- Shepard, R. N., Hovland, C. I., & Jenkins, H. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied*, 75, 1–42. doi:10.1037/h0093825
- Stager, C. L., & Werker, J. F. (1997, July 24). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, 388, 381–382. doi:10.1038/41102
- Storkel, H., Armbrüster, J., & Hogan, T. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research*, 49, 1175–1192. doi:10.1044/1092-4388(2006/085)
- Swingle, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, 13, 480–484.
- Swingle, D., & Aslin, R. N. (2007). Lexical competition in young children's word learning. *Cognitive Psychology*, 54, 99–132. doi:10.1016/j.cogpsych.2006.05.001
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, 108, 550–592. doi:10.1037/0033-295X.108.3.550
- Van Selst, M., Ruthruff, E., & Johnston, J. C. (1999). Can practice eliminate the psychological refractory period effect? *Journal of Experimental Psychology: Human Perception and Performance*, 25, 1268–1283. doi:10.1037/0096-1523.25.5.1268
- Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 735–747. doi:10.1037/0278-7393.28.4.735
- Willingham, D. B. (1999). Implicit motor sequence learning is not purely perceptual. *Memory & Cognition*, 27, 561–572. doi:10.3758/BF03211549
- Ziessler, M., & Nattkemper, D. (2001). Learning of event sequences is based on response-effect learning: Further evidence from a serial reaction task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 595–613. doi:10.1037/0278-7393.27.3.595

(Appendix follows)

Appendix

List of Chords

Mapping		Chord						
		Sparse				Dense		
Experiments 1 and 3								
Mapping 1	1 3 9	1 4 6			2 5 6		2 5 7	
	1 7 10	3 4 10			2 5 8		2 6 8	
	3 6 7	4 7 9			2 7 8		5 6 8	
	6 9 10				5 7 8			
Mapping 2	1 3 9	1 4 6			2 5 7		2 5 8	
	1 8 10	3 4 8			2 5 9		2 7 8	
	3 6 10	4 9 10			2 7 9		5 7 8	
	6 8 9				5 7 9			
Mapping 3	1 3 9	1 4 6			2 4 5		2 4 7	
	1 8 10	3 4 8			2 5 7		2 5 9	
	3 6 10	4 9 10			2 7 9		4 5 7	
	6 8 9				5 7 9			
Mapping		Sparse chord			Dense chord			
		Easy		Difficult	Easy		Difficult	
Experiment 2								
Mapping 1	6 8 9	2 3 9	3 4 0	1 3 7	1 5 7	1 5 8	2 4 6	2 5 7
	2 3 5	1 2 7	2 6 8	1 4 6	1 7 10	1 8 9	2 5 8	2 7 10
	1 5 9	6 7 10	5 7 9	7 8 10	2 4 0	3 9 10	3 4 10	3 4 6
	4 5 10		1 4 8		6 8 9		3 6 9	
Mapping 2	1 7 10	1 2 7	3 4 6	3 4 10	1 5 7	1 7 10	2 8 10	4 6 9
	6 8 9	5 8 9	2 6 10	3 7 8	1 8 10	2 8 9	3 4 6	3 5 6
	5 6 10	2 3 5	3 7 9	1 4 8	2 4 9	1 5 8	3 5 7	2 7 10
	2 4 5		1 4 9		5 6 9		3 4 7	
Mapping 3	1 5 7	1 5 9	3 4 10	3 4 7	2 4 10	5 8 9	2 8 10	3 4 6
	2 8 9	2 3 5	2 6 8	1 4 8	2 3 5	2 3 9	1 3 6	5 7 9
	2 4 5	1 6 10	1 3 8	7 9 10	5 6 10	1 2 7	4 7 8	1 4 8
	5 6 10		6 7 9		1 7 10		6 9 10	

Note. The mappings refer to the fingers used in the sparse and dense chords. For example, for the chord 1–3–9, the participant would use the left pinky, left middle, and right ring fingers. Each experiment had three different mappings.

Received June 16, 2012
Revision received October 14, 2012
Accepted October 15, 2012 ■