



Familiarity acts as a reduction in objective complexity

Jing Zhang^{1,2} · Xiaonan L. Liu³ · Matt So² · Lynne M. Reder^{2,4}

Published online: 3 June 2020
© The Psychonomic Society, Inc. 2020

Abstract

High-complexity stimuli are thought to place extra demands on working memory when processing and manipulating such stimuli; however, operational definitions of complexity are not well established, nor are the measures that would demonstrate such effects. Here, we argue that complexity is a relative quantity that is affected by preexisting experience. Experiment 1 compared cued-recall performance for Chinese and English speakers when the stimuli involved Chinese features that varied in the number of strokes or involved Ethiopic features unfamiliar to both groups. Chinese pseudocharacters (two radicals) had half the strokes of Chinese pseudowords (two characters). The response terms were English words familiar to both groups. English speakers performed equivalently with the Ethiopic and pseudocharacters, but much worse on the pseudowords. In contrast, Chinese speakers performed equivalently with pseudowords or pseudocharacters, but worse with Ethiopic cues. Experiment 2 showed that the lack of a complexity effect for Chinese speakers was not due to greater ease of rehearsal of pseudowords compared with pseudocharacters. Experiment 3 ruled out that Chinese speakers are just better at learning paired associates involving Mandarin by demonstrating that while complexity did not affect them, other features of the stimuli did. Taken together, it appears that complexity is not an absolute property based on the number of visual elements, but rather a relative property affected by one's prior knowledge.

Keywords Complexity · Working memory · Familiarity · Prior experience

Complexity has been an object of study in many fields, including visual perception (Oliva, Mack, Shrestha, & Peeper, 2004; Palumbo, Ogden, Makin, & Bertamini, 2014; Yoon, Lim, & Ji, 2015), auditory perception (Eerola, Himberg, Toiviainen, & Louhivuori, 2006; Hannon, Soley, & Ullal, 2012; North & Hargreaves, 1995), and memory (Alvarez & Cavanagh, 2004; Chen, Li, & Liu, 2017; Halford, Wilson, & Phillips, 1998; Kemps, 1999). Researchers have long tried to quantify

complexity to either measure or control its effects (Attneave, 1957; Forsythe, Mulhern, & Sawey, 2008; French, 1954; Jakesch & Leder, 2015; Purchase, Freeman, & Hamer, 2012). However, since complexity appears in so many contexts, multiple definitions have been put forth and remain controversial (Edmonds, 1995; Forsythe, 2009; Nadal, Munar, Marty, & Cela-Conde, 2010; Rao & Lohse, 1993; Xing & Manning, 2005). For example, Snodgrass and Vanderwart (1980) defined complexity as the amount of detail or intricacy of lines in a picture, while Heaps and Handel (1999) defined complexity as “the degree of difficulty in giving a verbal description (on page 303).”

Regardless of these differences in how complexity is defined, most would agree that complexity affects behavior (e.g., McDougall, De Bruijn, & Curry, 2000; Rock, Halpern, & Clayton, 1972; Rosenholtz, Li, & Nakano, 2007; Sweller, 2010; Tuch, Bargas-Avila, Opwis, & Wilhelm, 2009). In particular, stimuli of higher complexity typically lead to longer processing times and worse performance on tasks, including diminished memory span (Alvarez & Cavanagh, 2004; Eng, Chen, & Jiang, 2005; Song & Jiang, 2006). Higher complexity stimuli are often considered to be an additional strain on working memory, consuming more working memory resources to process and manipulate the information (Liu, Chen, Liu, & Fu, 2012; Luria, Sessa, Gotler, Jolicoeur, & Dell'Acqua, 2010).

Jing Zhang and Xiaonan L. Liu contributed equally and should be considered co-first authors

✉ Jing Zhang
zhangjing1026@seu.edu.cn

✉ Lynne M. Reder
reder@cmu.edu

¹ School of Mechanical Engineering, Southeast University, Nanjing 211189, China

² Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, USA

³ Center for Neuroscience, University of California, Davis, Davis 95618, CA, USA

⁴ Center for the Neural Basis of Cognition, Pittsburgh, PA, USA

Although these results do not seem controversial, recent empirical studies (Reder, Liu, Keinath, & Popov, 2016; Shen, Popov, Delahay, & Reder, 2018) from our lab suggest that the effects of complexity on memory disappear with familiarization. For example, in one study (Reder et al., 2016), subjects who were previously unfamiliar with Chinese characters were trained to recognize these characters over a period of weeks using a visual search task with some characters presented 20 times more often than other characters. These characters were randomly assigned to frequency conditions, irrespective of their complexity. After the training sessions, subjects were asked to learn associations of two Chinese characters with one English word. The results showed that associations involving high-frequency characters were better remembered than those involving low-frequency characters. Moreover, subjects also showed better performance with high-frequency characters in a working memory task (*N*-back task) than with low-frequency characters.

Although those studies did not explicitly manipulate complexity, their results suggest that the effect of complexity might be modulated by whether the information has been both chunked and the chunk practiced enough that it becomes stronger (more familiar). When a stimulus is highly complex and has many parts, the process of chunking results in a reduction in the number of parts, leading to an effective reduction in “subjective” complexity. The notion that it is easier to hold information in working memory when it can be “chunked” into meaningful or organized units has been known for over half a century (e.g., Miller, 1956; Simon, 1974). More recently, we have argued that when these chunks become stronger, they consume less of limited working memory resources, making it easier to complete tasks with familiar chunks (Popov & Reder, 2020; Reder et al., 2016; Reder, Paynter, Diana, Ngiam, & Dickison, 2007; Shen et al., 2018).

While the notion that familiarity modulates the effect of complexity on memory performance seems intuitive, no prior studies have directly examined this, nor shown that relationship, to the best of our knowledge. In fact, as reviewed above, the literature tends to define complexity as an absolute quantity based on counting or classifying the presenting features and components (Chikhman, Bondarko, Danilova, Goluzina, & Shelepin, 2012; Donderi, 2006; Forsythe, Sheehy, & Sawey, 2003; García, Badre, & Stasko, 1994; Machado et al., 2015). Therefore, the experiments in this article are intended to test the hypothesis that complexity is, in fact, a relative quantity and that preexisting experience and knowledge affects memory performance on stimuli that differ in objective complexity (e.g., the primitive features or number of strokes, number of radicals).

Here, we compare the performance of two groups of subjects on stimuli that are novel to everyone, but for which the constituent parts of these stimuli vary in familiarity between the two groups. One group has had extensive experience with

Chinese characters, and the other group has not. The stimuli are novel pairings for which the constituents are highly familiar for only Chinese subjects. Chinese pseudowords consist of two real Chinese characters and Chinese pseudocharacters consist of two real Chinese radicals. The stimuli to be learned will be unfamiliar for both language groups, but for native English speakers, these novel combinations are unfamiliar both at the level of combination and at the level of the constituent elements. For Chinese speakers, the stimulus cues are also completely novel but the components or constituents of these stimuli are familiar to them, either as characters or radicals. These novel stimuli are studied with different English word response terms that must be recalled when the stimulus is later presented as part of a cued-recall task.

As a control for potential individual differences between the two groups on other dimensions, both groups also study pairs involving Ethiopic pseudowords as cues. Since these stimuli are equally unfamiliar for both groups, this provides a reference for ability to learn novel stimuli.

Experiment 1

Subjects studied novel symbol combinations (never seen before as a pair) along with English words. There were three lists, each with different types of stimulus pairs that served as a unique cue to an arbitrarily assigned English word. After studying a list, subjects were given a cued-recall task that required them to try to recall the English word that had been studied with the current probe stimulus. After subjects attempted to recall each cue on the current list, another test list was presented until all lists had been tested.

Method

Subjects

Fifty-one U.S. college students from Carnegie Mellon University participated in Experiment 1 for partial course credit and an additional bonus up to \$5, depending on performance (accuracy) on the task. Thirty-four subjects were native English speakers with no prior experience with the Chinese language, except through casual exposure. There were 17 native Chinese speakers, fluent in Mandarin, raised in China, and educated there at least through high school. These subjects were also sufficiently fluent in English to matriculate at a good American university. One Chinese subject was dropped from the experiment because his performance was two standard deviations below the mean of the remaining subjects. This left 16 Chinese subjects and 34 English subjects.

Design and materials

This experiment used a 3 (stimulus types) \times 2 (native languages) mixed design such that both groups of subjects were exposed to all three stimulus types: Ethiopic pseudowords, Chinese pseudocharacters, and Chinese pseudowords. An Ethiopic pseudoword consisted of two Ethiopic characters, while a Chinese pseudocharacter consisted of two Chinese radicals randomly combined to form a novel character never seen by either subject group. Chinese pseudowords consisted of two randomly combined Chinese characters with the constraint that they did not inadvertently form a real word in Chinese (see Fig. 1). Note that each Chinese pseudoword contained two characters, and each Chinese character contained at least two radicals such that Chinese pseudowords contained at least twice the number of strokes as a pseudocharacter. Therefore, if complexity of Chinese pseudowords and Chinese pseudocharacters is defined by the number of strokes contained in the cue, Chinese pseudowords are at least twice as complex as Chinese pseudocharacters.

The design involved a total of 16 Ethiopic characters, 16 Chinese characters and 16 Chinese radicals. As noted above, each list was comprised of only one of the three types of stimuli. For lists involving Chinese pseudocharacters, each radical was used twice in two different symbol pairings, and each pair combination was studied with a different English word. Likewise, each character in the Chinese pseudoword condition and each Ethiopic character in the Ethiopic pseudoword condition were also in two different combinations within a list, with each pairing associated with a different English word. By using each symbol in two different pairings within a list, subjects were forced to memorize both elements of the pair (not just the left or right symbol of a pair) that was associated with a unique English word. Given that each character/radical was repeated exactly twice in each list and bound with a different character/radical for the other association, each character or radical of a pair should contribute equally to the learning of the association. Each list contained exactly 16 cues associated with a different English word.

All Chinese characters and radicals were chosen from Levels 1 and 2 (i.e., medium-frequency and high-frequency characters) of the *Standard List of Common Characters in Modern Chinese* (State Language Work Committee, 1988), with between six and 12 strokes for characters (mean = 8.9) and between two and five strokes for radicals (mean = 3.8) and displayed in Song typeface. All the Ethiopic characters were chosen from online Unicode Entity Codes for Ethiopic Language (<http://www.personal.psu.edu/ejp10/symbolcodes/bylanguage/ethiopicchart.html>). Forty-eight English words were randomly assigned to those symbol pairs for each subject, forming 48 pair combinations. These English words were chosen from the MRC Psycholinguistic Database, had familiarity ratings of 600 or higher, and had a word length between three and six letters (http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm). The size of each character/radical on the screen was 130 \times 130 pixels, while the screen resolution was 1,280 \times 800. The viewing distance was approximately 50 cm.

Procedure

Subjects repeatedly studied the same 48 pair combinations across six rounds to test learning. The order of the three lists was the same for each round, but the order of the 16 pair combinations in each study list and test list was randomized for each subject. Figure 2 illustrates an example round, consisting of study and test phase for each of the three stimulus types. During study trials, each symbol pair was shown along with its English word for 3 seconds before the next pair combination automatically appeared. After studying all 16 pair combinations in a list, test trials immediately began for the items on that list. The subject was cued with one of the studied symbol pairs and prompted to enter its associated English word. The 16 possible English words for that list were displayed on the left and right sides of the screen in a standardized layout, in random order across rounds. There was no time limit for responses. After entering the answer, subjects received visual feedback that indicated whether their response was correct or not. This continued until all 16 of the studied

Ethiopic Character		Chinese Character		Chinese Radical	
ጁ	ገ	校	就	彳	冂
Ethiopic Pseudo-Word		Chinese Pseudo-Word		Chinese Pseudo-Character	
ጁ ገ		校 就		彳 冂	

Fig. 1 Examples of Ethiopic pseudoword, Chinese pseudocharacter, and Chinese pseudoword symbol pairs

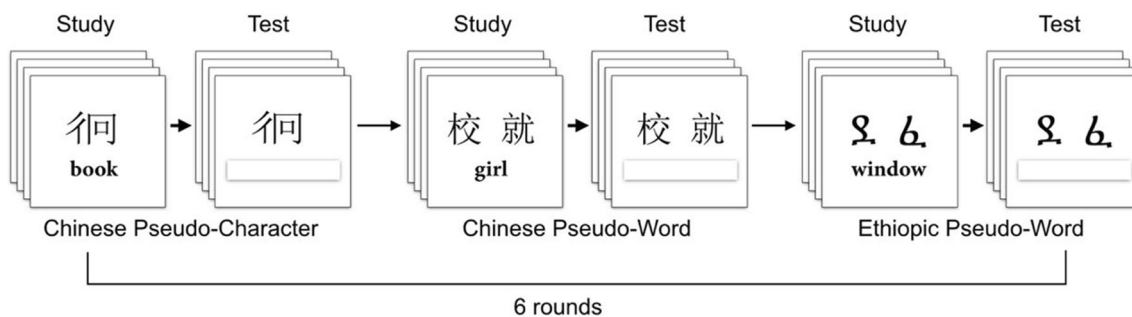


Fig. 2 Example round consisting of a study and test phase for each of the three stimulus types. There were six rounds total of the three types of stimulus lists

pair combinations for that list were tested. There were six rounds total for each of the three types of stimulus lists, and there was a 1-minute break between each round. The entire experiment lasted approximately 1 hour.

Results and discussion

To ensure that the sample size achieved adequate power, a power analysis was conducted using the “pwr package” in R (Champely, 2016; Cohen, 1988). Because each subject had six rounds of study and test trials, we assumed a large effect size (i.e., $\eta_p^2 = .14$; Cohen, 1992). The achieved power was .898, with $\alpha = 0.05$.

We analyzed the accuracy data via logistic mixed-effects regressions (Baayen, Davidson, & Bates, 2008; Jaeger, 2008). Figure 3 shows the proportion correct on the cued-recall tests as a function of language group and stimulus type of cue for each round. The main effect of round was significant, $\Delta AIC = -2,008$; $LLR \chi^2(1) = 2,018.6$; $p < .001$, such that accuracy increased from Round 1 to Round 6. Because we are not interested in the role of practice of the pairs (round), further analyses focus on the interaction between language group and stimulus type. There was a significant interaction between

stimulus type and language group, $\Delta AIC = -166$, $LLR \chi^2(1) = 169.412$, $p < .001$, such that Chinese speakers performed basically the same in the Chinese pseudocharacter condition, the Chinese pseudoword condition, and the Ethiopic condition, $\Delta AIC = 2.7$, $LLR \chi^2(1) = 1.297$, $p = .523$. In contrast, performance for English speakers differed significantly depending on stimulus condition, $\Delta AIC = -493$, $LLR \chi^2(1) = 496.65$, $p < .001$. Native English speakers performed best in the Ethiopic condition, slightly worse in the Chinese pseudocharacter condition, and worst in the Chinese pseudoword condition. In summary, this pattern suggests that, although a Chinese pseudoword contains two characters or twice the number of radicals as a pseudocharacter (i.e., it is “objectively” more complex), the effect of complexity is moderated by familiarity for Chinese speakers.

We would like to conclude that complexity effects on learning are modulated by familiarity with the stimuli; however, there is an alternative explanation for these results. Specifically, it could be that Chinese pseudowords were not subjectively less complex for Chinese speakers than for English speakers, but rather that pseudowords were pronounceable, giving them a memorization advantage. That is, it would be easier to subvocally rehearse pseudowords than

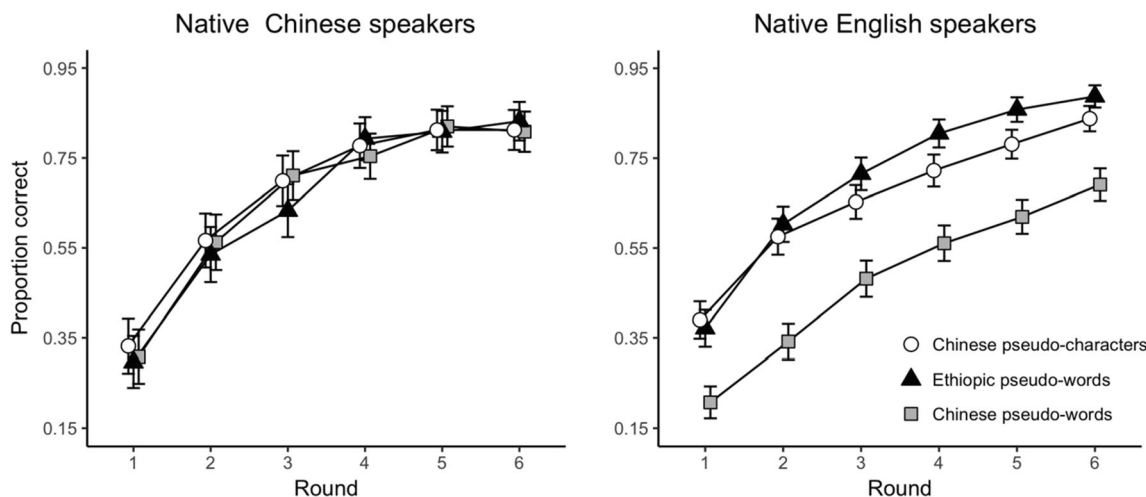


Fig. 3 Mean performance of the two language groups over six rounds of practice. The rounds of study–test learning are shown on the x-axis with separate plots for each of the three stimulus types. Error bars represent 95% confidence intervals

pseudocharacters. In general, memory researchers agree that rehearsal facilitates learning (e.g., Atkinson & Shiffrin, 1968; but see Lewandowsky & Oberauer, 2015, for an alternative perspective). Since real Chinese characters, unlike pseudocharacters, are pronounceable, native Chinese speakers could more easily pronounce the two chunks of the pseudowords compared with the two chunks of the pseudocharacters. This subvocal rehearsal advantage might compensate for any advantage of fewer strokes for the pseudocharacters. To assess the merit of this alternative explanation, Experiment 2 was designed to eliminate the rehearsal advantage for pseudowords over pseudocharacters.

Experiment 2

A plausible conclusion of Experiment 1 was that complexity in terms of number of strokes did not matter when the stimuli were highly familiar and unitized into chunks. However, that finding was subject to a potential artifact of a rehearsal advantage for Chinese pseudowords compared with pseudocharacters. To determine the plausibility of the alternative explanation, we used stimuli that effectively removed any possible rehearsal advantage. This method involved using only homophonous Chinese pseudowords. If all stimuli have the same pronunciation, subjects cannot use rehearsal to remember which Chinese pseudoword was associated with a given English word. We only used native Chinese speakers as subjects since the rehearsal advantage does not apply to non-Chinese speakers. Apart from the change in some of the stimuli and only using Chinese speakers, Experiment 2 was the same as Experiment 1.

Method

Subjects

A new group of 16 U.S. college students from Carnegie Mellon University and the University of Pittsburgh participated in this study. As defined in Experiment 1, all subjects were native Chinese speakers raised in China and educated in China at least through high school. They all also spoke English. In exchange for participation, subjects received a payment between \$10 and \$14, depending on performance (accuracy) on the task.

Design and materials

The Ethiopic pseudowords and Chinese pseudocharacters used in Experiment 2 were the same as those of Experiment 1. However, we selected 16 new homophonous Chinese characters; eight of them were all pronounced “zhi,” and the other eight were all pronounced “shu.” The 16 Chinese pseudowords in Experiment 1 were replaced with 16 new Chinese pseudowords that used homophones for the left and right

characters such that all Chinese pseudowords would be pronounced in the same way, “zhi shu” (see Table 1). Although the pronunciation was identical, the characters differed for each pair. Like in Experiment 1, each character was used in two different symbol pairings so that subjects were forced to memorize both elements of the pair, not just left or right character of the pseudoword pair. All the homophonous Chinese characters were chosen from Levels 1 and 2 (i.e., medium-frequency and high-frequency characters) of the *Standard List of Common Characters in Modern Chinese* (State Language Work Committee, 1988), with stroke counts between seven and 13 (mean = 9.8). They were displayed in Song typeface.

Procedure

The procedure for Experiment 2 was the same as Experiment 1.

Results and discussion

A power analysis, using the “pwr package” in R (Champely, 2016; Cohen, 1988), indicated that the achieved power was .942, with α set at .05 and a large effect size assumed (i.e., $\eta_p^2 = .14$; Cohen, 1992).

Figure 4 plots the accuracy by round in Experiment 2 for the three types of stimuli. As in Experiment 1, there is a main effect of round, $\Delta AIC = -1,041$; LLR $\chi^2(1) = 1,050.991$; $p < .001$, such that performance improved over the six rounds. Importantly, native Chinese speakers did not show significant differences in performance among stimulus types, $\Delta AIC = 2.7$, LLR $\chi^2(1) = 1.297$, $p = .522$.

The goal of this experiment was to rule out the alternative explanation for the finding that complexity did not affect performance for Chinese speakers. That is, did pseudoword cues produce equivalent performance because they could be more easily rehearsed than the less complex stimuli and thereby offset any disadvantage from their greater number of features? The results of Experiment 2 ruled that out that explanation because cued recall performance was still as good for pseudowords when rehearsal could not benefit Chinese speakers. When all the pseudowords are pronounced the same, rehearsal cannot help a subject remember the appropriate English response term to a given character pair.

Table 1 The list of 16 homophonous Chinese pseudowords (all pronounced “zhi shu”)

16 homophonous Chinese pseudo-words			
致树	质竖	志述	制恕
帜树	智竖	治述	置恕
智束	治漱	置数	致墅
帜束	质漱	志数	制墅

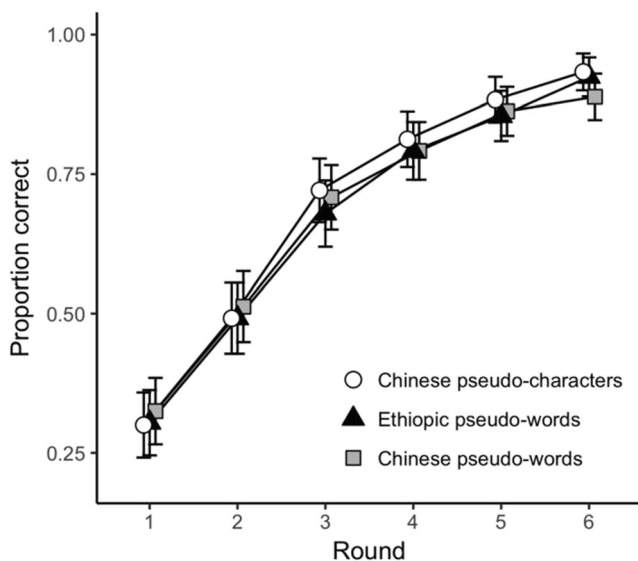


Fig. 4 Mean performance of native Chinese speakers in Experiment 2 over six rounds. Error bars represent 95% confidence intervals

While Experiment 2 provided additional support for the conclusion that complexity should not be defined by the number of visual elements in a stimulus, and instead should take into account the user's familiarity with the stimuli to be processed, the results do raise other questions. Zhang and Simon (1985) found that homophones performed equivalently to radicals that also could not be pronounced, and we still found the expected equivalence. On the other hand, their study showed a memory advantage for items where rehearsal could help performance (i.e., nonhomophones). Comparing across studies, it is not clear whether there would be a residual advantage for items that could be pronounced and rehearsed for Chinese speakers. Our Experiment 2 did not compare within subject nor within the same experiment a contrast between pseudowords that had distinct pronunciations versus pseudowords that were homophones. Experiment 3 was designed to compare these conditions in a within-subjects design to determine whether (a) we can demonstrate an effect of rehearsal advantage at the same time that we show the absence of a complexity disadvantage for Chinese subjects, and (b) the same set of Chinese stimulus cues generate opposite patterns for native English speakers compared with native Chinese speakers based on prior experience. In other words, can we demonstrate that Chinese subjects still are unaffected by "objective complexity" when rehearsal advantage is removed, while non-Chinese subjects are strongly affected by this objective complexity factor?

Experiment 3

In Experiment 3, we crossed the two within-subjects factors (homophonous vs. nonhomophonous characters and high vs.

low complexity, as defined by number of strokes) with language group, creating a mixed design. We expected the first factor to affect performance for native Chinese speakers, but not the complexity factor (number of strokes) and the opposite pattern for native English speakers, unfamiliar with Chinese. For native English speakers, homophones should not matter, since they do not know the characters' pronunciation, but the complexity factor should again strongly affect performance.

Method

Subjects

Subjects were 41 college students recruited from Carnegie Mellon University and the University of Pittsburgh. As defined in Experiment 1, 21 subjects were native English speakers, with no training in Chinese, and 20 were native Chinese speakers who were all raised in China, educated in China at least through high school, and also spoke English. Two subjects were dropped from the experiment because one English subject did not complete the study and another English subject's performance was at chance. This left 20 Chinese subjects and 19 English subjects. In exchange for participation, subjects received a payment between \$8 and \$16, depending on performance on the task.

Design and materials

We used a mixed design, with the between-subjects factor being native language: Chinese versus English. The two within-subjects factors were complexity (high vs. low) and whether the pronunciation of the pair of Chinese characters were identical for each pair in a list (homophonous) or unique. This design yielded four conditions for the pseudoword lists: HD for high-complexity with different pronunciation; HS for high-complexity with the same pronunciation; LD for low-complexity with different pronunciation, and LS for low-complexity with same pronunciation. In addition to these four stimulus conditions, as in Experiment 1, we included a fifth stimulus type, Ethiopic pseudowords (referred to as PE).

We selected 144 Chinese characters that were not used in Experiments 1 and 2, with 36 characters for each condition. Complexity was defined by the number of strokes for a given character, determined using criteria specified in the online Xinhua Dictionary (<https://zd.diyifanwen.com/zidian/bh/>). Low-complexity characters were defined as having no more than six strokes (mean = 4.72), and high-complexity characters were selected to have at least 10 strokes (mean = 12.08). Within the 36 characters for each condition, there were four subgroups of nine characters each. For the homophonous conditions (HS and LS), all nine characters in a subgroup had the same pronunciation (e.g., "shi," "zhi," "jian"). For each condition, instead of the same 16 pseudowords (characters) for all

subjects in Experiment 1, 12 characters were randomly selected to generate 12 unique pseudowords for each subject, six of which came from one subgroup and the other six from a separate subgroup. The randomized assignment of characters to each condition for each subject enabled us to further control for the potential confound that would be caused by the nature of characters, such as frequency and meaning. As in Experiments 1 and 2, each character was used twice in two different symbol pairings so that subjects were forced to memorize both elements of the pair, not just the left or right character. All 144 Chinese characters were chosen from Levels 1 and 2 (i.e., medium-frequency and high-frequency characters) of the *Standard List of Common Characters in Modern Chinese* (State Language Work Committee, 1988). One hundred and twenty English words were randomly assigned to those symbol pairs for each subject, forming 120 pair combinations. All of the English words were chosen from the MRC Psycholinguistic Database, had familiarity ratings of 600 or higher, and had a word length between three and six letters (http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm). All the Chinese characters were set in Kaishu typeface. The size of each character on the screen was 130×130 pixels, and the screen resolution was $1,280 \times 800$. The viewing distance was approximately 50 cm.

Procedure

Experiment 3 used the same procedure described in Experiment 1, with a few modifications, as noted here. Subjects studied five different lists, corresponding to the five different conditions denoted in Fig. 5. There were three rounds of study, followed by test for each of the five lists. The five lists were presented in a random order that changed for each subject on each round. After performing all study–test phases for the five conditions, the entire experiment was repeated with a totally different set of stimuli representing the same conditions. We refer to these as Block 1 and Block 2, with each block consisting of 15 study–test lists. There was a 1-

minute break between each round and a 5-minute break between the two blocks. The entire experiment lasted approximately 70 minutes.

Results and discussion

A power analysis using the “pwr package” in R (Champely, 2016; Cohen, 1988) indicated that the achieved power was .864, with α set at .05, and a large effect size assumed (i.e., $\eta_p^2 = .14$; Cohen, 1992).

Figure 6 plots the mean performance for the cued-recall tests as a function of language group and stimulus type for each round after collapsing over Block 1 and Block 2. Performance for each of the five stimulus types is plotted in the left panel for native Chinese speakers and in the right panel for native English speakers. The results showed a significant main effect of round, $\Delta AIC = -738$, LLR $\chi^2(1) = 740.16$, $p < .001$, such that both groups performed more accurately in later rounds, regardless of stimulus type, as the combinations became more familiar. Because we are primarily interested in the interaction effects between pronunciation and complexity, further analyses did not include the round factor. Consistent with Experiment 1, Chinese speakers performed better in the Chinese conditions than English speakers, $\Delta AIC = -4$, LLR $\chi^2(1) = 6.627$, $p = .01$. On the contrary, English speakers showed better performance in the Ethiopic control condition than Chinese speakers at a trend level, $\Delta AIC = -1$, LLR $\chi^2(1) = 3.011$, $p = .083$.

Figure 7 shows the cued-recall performance of the two native language groups when the cues are Chinese characters, plotted as a function of pronunciation (x -axis) and complexity (separate lines for the two levels). Also plotted on the graph is performance for the Ethiopic pseudoword cues (triangles), which served as baseline stimuli for which neither group was familiar. First, we report the analyses that exclude the Ethiopic stimuli. As expected, there was a significant interaction between complexity and language group, $\Delta AIC = -32$, LLR $\chi^2(1) = 33.46$, $p < .001$, such that native English speakers were more accurate for less complex cues, $\Delta AIC = -42.9$,

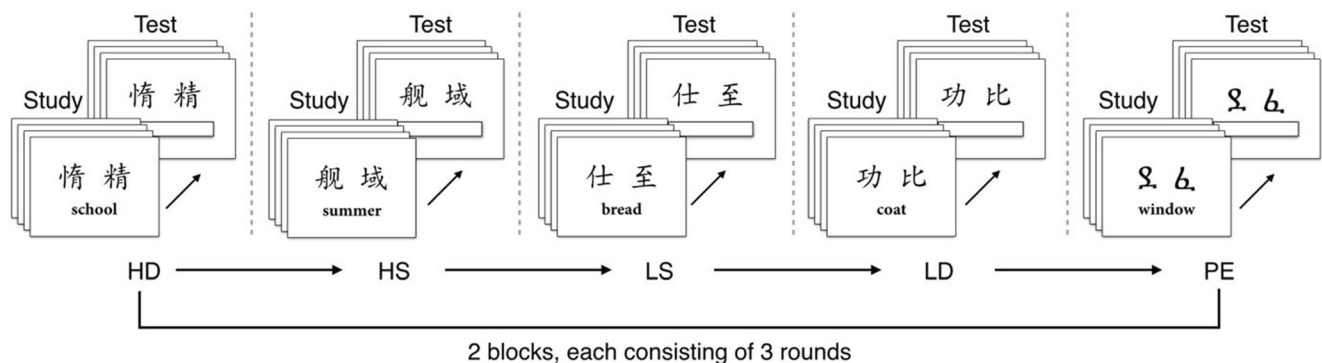


Fig. 5 Example round consisting of a study and test phase for each of the five stimulus types. Altogether, there were two blocks, each consisting of three rounds

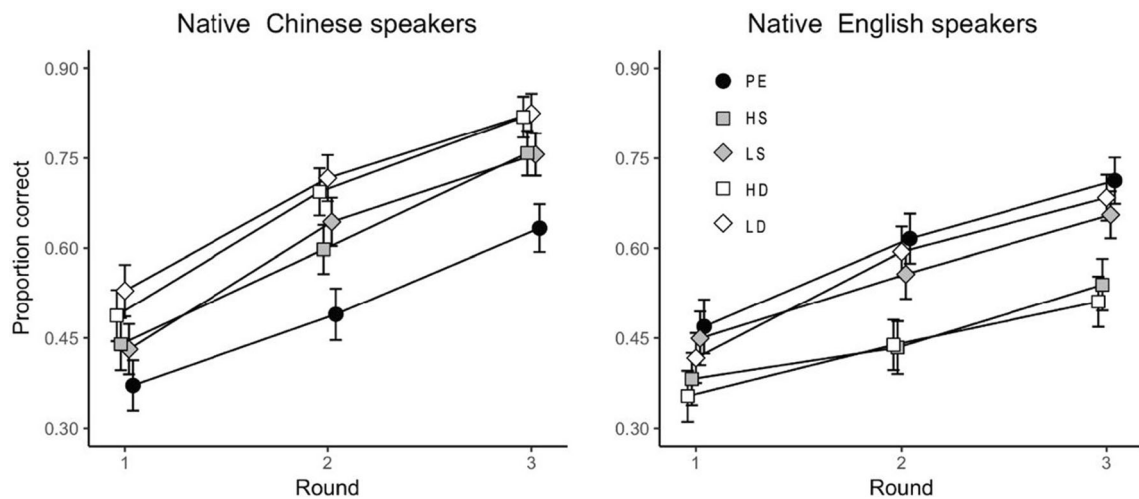


Fig. 6 Mean performance of two groups in Experiment 3 after collapsing over Block 1 and Block 2. Left panel plots accuracy for native Chinese speakers, and right panel plots accuracy for native English speakers. PE = Ethiopian pseudowords; HS = high-complexity, same pronunciation; LS =

low-complexity, same pronunciation; HD = high-complexity, different pronunciation; LD = low-complexity, different pronunciation. Error bars represent 95% confidence intervals

LLR $\chi^2(1) = 44.863, p < .001$, while Chinese speakers showed no effect of complexity, $\Delta AIC = 1.2$, LLR $\chi^2(1) = .851, p = .356$. There was also a significant interaction between pronunciation and language group, $\Delta AIC = -19$, LLR $\chi^2(1) = 20.842, p < .001$, such that Chinese speakers showed a difference in performance based on the pronunciation of the cues (performance was worse when the pronunciations were the same), $\Delta AIC = -14.5$, LLR $\chi^2(1) = 16.522, p < .001$; in contrast, English speakers showed no effect of whether Chinese characters were homophones, $\Delta AIC = 1.8$, LLR $\chi^2(1) = .1416, p = .707$. There was no significant two-way interaction between complexity and pronunciation on accuracy, $\Delta AIC = 1$, LLR $\chi^2(1) = 1.049, p = .306$, nor a three-way interaction between language group, complexity, and pronunciation, $\Delta AIC = 2$, LLR $\chi^2(1) = .303, p = .582$.

To summarize, native English speakers' performance was affected by the complexity of the cue stimuli, but not whether the pronunciation was the same. Conversely, performance of native Chinese speakers showed the opposite pattern: They were affected by whether the different cues shared the same pronunciation, but were unaffected by the complexity (number of strokes) of the stimuli. These results provide additional support for the results from Experiments 1: Complexity of Chinese characters only affects performance for those unfamiliar with Chinese characters.

The fact that native Chinese speakers were adversely affected when the stimuli were homophones while native English speakers were unaffected rules out the possibility that the weaker effect of complexity for native Chinese speakers was due to a ceiling effect caused by being better subjects. In

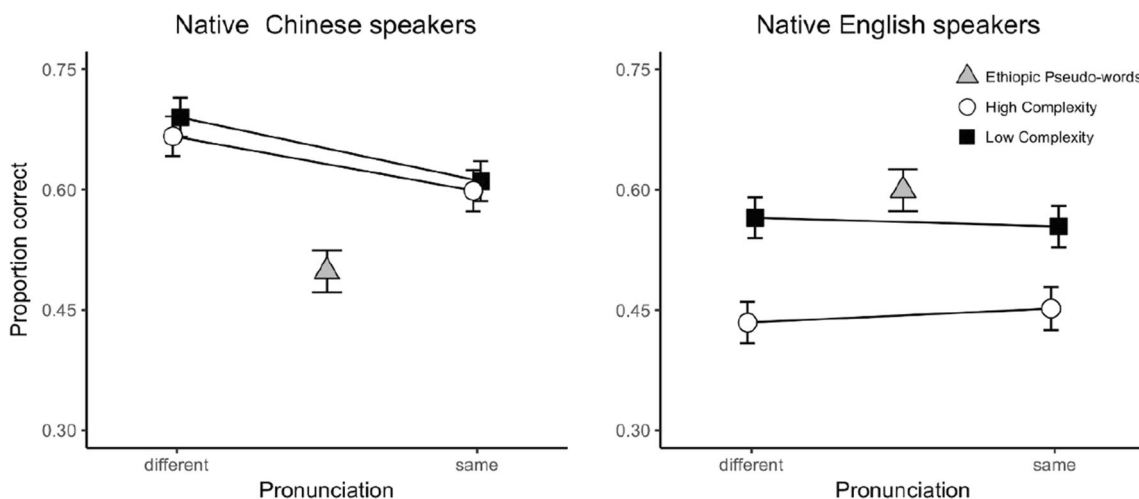


Fig. 7 Mean recall accuracy of English words as a function of the type of cues paired with words on a given list for the two native language groups. Error bars represent 95% confidence intervals

fact, Chinese speakers did no better, and arguably worse, when the cues were from a third language unknown to either group (Ethiopic stimuli). In the General Discussion, we provide an explanation for why native English speakers did slightly better when the cues were from a language unfamiliar to both groups.

General discussion

It has long been known that high-complexity stimuli are harder to process, remember, and reproduce from memory than are low-complexity stimuli (Alvarez & Cavanagh, 2004; Attneave, 1957; Bradley, Hamby, Löw, & Lang, 2007; Eng et al., 2005; Song & Jiang, 2006). However, one limitation of the literature is that complexity is commonly regarded as an absolute quantity that is defined by the number of features and components, while subjective factors that modulate the effect of complexity, such as preexisting knowledge of the stimuli, have been neglected. Here, our results demonstrate that objective complexity does not tell the whole story. Our results provide cases in which objective complexity (here, defined as number of strokes) failed to make a difference. Moreover, whether or not one finds an effect of complexity on learning and memory is modulated by the familiarity of the stimuli. Specifically, the effects of complexity exist only when subjects are unfamiliar with the stimuli, and once the stimuli are highly familiar to subjects, the effects of complexity disappear completely. In Experiment 1, we found evidence that the effects of complexity on learning were eliminated when the stimuli were highly familiar (for native Chinese speakers), but not for those unfamiliar with the stimuli. Experiment 2 replicated the lack of a complexity effect for native Chinese speakers while ruling out the potential confound that more complex stimuli only appeared as easy as simpler stimuli because they could be vocalized to aid rehearsal. Experiment 3 provided a demonstration that the same set of Chinese stimulus cues could generate opposite patterns of difficulty for native English speakers compared with native Chinese speakers based on prior exposure to the stimuli. Specifically, Chinese speakers' performance suffered when the cues were different pseudowords that were all pronounced the same way; however, their performance was unaffected by the complexity of the stimuli (as defined by number of strokes). Conversely, native English speakers' performance suffered when the stimuli were more complex but were unaffected by the pronunciation of the stimuli.

It is no surprise that subjects who do not speak Chinese would not be affected by whether Chinese characters share the same pronunciation, since English speakers could not pronounce them in the first place. Nevertheless, it is useful to demonstrate that English speakers are less affected by one of the dimensions of the Chinese stimuli than are Chinese

speakers given that they are more affected by another dimension of the stimuli. Although it may seem obvious that English speakers would be unaffected by features of the stimuli that they cannot perceive (the sounds of the characters), it is not obvious that Chinese speakers should be less affected by complexity than English speakers. In the latter case, all subjects can perceive the visual differences among stimuli. In fact, Chinese speakers' performance was virtually unaffected when the stimuli involved greater complexity (more strokes) than simpler stimulus cues. English speakers were considerably worse when the stimuli were more complex.

One explanation for the difference in effects of complexity for the two groups is offered by the chunking theory of Miller (1956) and Simon (1974). Chunking theory posits that when information can be grouped or chunked into fewer units, the resulting stimuli are more easily processed. For example, C-A-T can be considered as three separate chunks consisting of three different letters (and each letter consisting of features such as lines and parts of circles). Alternatively, CAT can be considered as one chunk that has letters as constituent parts and refer to a feline household pet. It seems reasonable to assume that, with time, native Chinese speakers likewise group the features of a character into chunks.

While the work of Simon and Miller demonstrated that more stimuli can be recalled when the information can be chunked, it does not explain why English speakers performed better in the Ethiopic condition than did Chinese speakers,¹ given that Ethiopic pseudowords were equivalently novel to both groups. More recent work (e.g., Reder et al., 2016; Reder et al., 2007; Shen et al., 2018) has extended this theory and shown that more familiar chunks are easier to combine together (into, for example, paired associates) and also to associate these pairs of strong chunks with an additional arbitrary stimulus. According to this theory, the strength of the chunks reflects their familiarity. Furthermore, as chunks become stronger, they deplete fewer working memory resources, allowing more to be devoted to binding stimuli together in long-term memory.

According to this elaboration of the chunking theory (see also Popov & Reder, 2020), the native English speakers have stronger/more familiar chunks for the English response terms, given their greater experience with English words. That means that there would be slightly more working memory resources available for the binding process for native English speakers when the cues were equivalent in familiarity, as was the case when the cues were Ethiopic characters. That is, while both groups expend working memory resources to encode the

¹ Given that we found a trend level difference in the Ethiopic control condition between two groups in Experiment 3, when we collapsed over Experiment 1 and Experiment 3 to examine the performance of two language groups in the Ethiopic condition, we found that English speakers performed significantly better in the Ethiopic condition than did Chinese speakers, $\Delta AIC = -2.4$, $LLR \chi^2(1) = 4.383$, $p = .036$.

Ethiopic characters and then to associate them with response terms, the processing and binding of the English response terms would consume less of the remaining resources for native English speakers, giving them a slight advantage² in binding.

More generally, the finding that familiarity acts as a reduction in complexity can be extended and used as an explanation for several other results. For example, there is an effect of number of syllables on working memory for pseudowords and low-frequency words, but not for high-frequency words, even when controlling for similarity and letter count (Ferrand, 2000; Ferrand & New, 2003; New, Ferrand, Pallier, & Brysbaert, 2006). This result may not seem intuitive and the cause is still debated (Juphard, Carbonnel, & Valdois, 2004), but it is predicted by our elaboration of the chunk theory. In that case, syllable count is the measure for complexity. Those effects are present for stimuli that are either novel or unfamiliar because they place greater demands on the limited resources of working memory; however, the effects disappear when stimuli are familiar and working memory resources are more plentiful (see Popov & Reder, 2020, for formal specifications of the theory.)

Other researchers have also reported situations in which familiarity with the stimuli removes the effects of complexity (Bethell-Fox & Shepard, 1988; Qian, Reinking, & Yang, 1994; Su & Samuels, 2010; Sun, Zimmer, & Fu, 2011). In addition, many studies on norms for pictures and icons have found a negative correlation between ratings of familiarity and ratings of complexity (Bonin, Peereman, Malardier, Méot, & Chalard, 2003; Cychowicz, Friedman, Rothstein, & Snodgrass, 1997; McDougall, Curry, & De Bruijn, 1999; Rossion & Pourtois, 2004; Sirois, Kremin, & Cohen, 2006; Snodgrass & Vanderwart, 1980), suggesting that the reduction of complexity is found in both performance and perception.

One potential concern with the current study is that the Chinese characters that formed pseudowords have semantic meanings in addition to pronunciations. One might wonder whether the meanings of the characters facilitated the memory for pseudowords even though the effect of pronunciation was controlled. However, based on a classic study by Zhang and Simon (1985), the effect of different meanings of Chinese homophones was very small compared with the problem of shared pronunciation. In addition, we further controlled the potential confounding of meaning of Chinese pseudowords by randomly combining two Chinese characters with the constraint that they did not inadvertently form a real word in Chinese. Thus, it is hard for subjects to form semantic representations of two random combined characters during the short presentation time (i.e., 3 seconds) for each pair.

² This slight advantage was far outweighed by the much larger working memory resource advantage for Chinese speakers when processing the cue terms that were Chinese, since the English speakers were not nearly as familiar with those stimuli as the Chinese speakers were with English.

In summary, the current study provides evidence that the effect of complexity on learning paired associates is modulated by familiarity with the stimuli. Moreover, the results support the argument that complexity is not an absolute property based on the number of visual elements, but rather is a relative property affected by one's prior knowledge with the stimuli. Once the stimuli are highly familiar, the effects of complexity go away. In addition, by including the Ethiopic pseudowords, our findings also support the theory that ability to learn novel associations among stimuli is affected by the strength, as well as the number, of chunks involved in the association.

Acknowledgements We are thankful to Ven Popov for his advice concerning data analyses and Michael Griffin for his thoughtful comments on the manuscript.

Open practices statement The data for all experiments in this study are available in the OSF repository (<https://osf.io/3ps4h/>), and none of the experiments was preregistered.

References

- Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*(2), 106–111. <https://doi.org/10.1111/j.0963-7214.2004.01502006.x>
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation*, *2*, 89–195. [https://doi.org/10.1016/S0079-7421\(08\)60422-3](https://doi.org/10.1016/S0079-7421(08)60422-3)
- Atneave, F. (1957). Physical determinants of the judged complexity of shapes. *Experimental Psychology*, *53*(4), 221–227. <https://doi.org/10.1037/h0043921>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 12–23. <https://doi.org/10.1037/0096-1523.14.1.12>
- Bonin, P., Peereman, R., Malardier, N., Méot, A., & Chalard, M. (2003). A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. *Behavior Research Methods, Instruments, and Computers*, *35*(1), 158–167. <https://doi.org/10.3758/BF03195507>
- Bradley, M. M., Hamby, S., Löw, A., & Lang, P. J. (2007). Brain potentials in perception: picture complexity and emotional arousal. *Psychophysiology*, *44* (3), 364–373. <https://doi.org/10.1111/j.1469-8986.2007.00520.x>
- Champely, S. (2016). pwr: Basic functions for power analysis [Computer software]. Retrieved from <https://cran.r-project.org/web/packages/pwr/pwr.pdf>
- Chen, X., Li, B., & Liu, Y. (2017). The impact of object complexity on visual working memory capacity. *Psychology*, *8*(6), 929–937. <https://doi.org/10.4236/psych.2017.86060>
- Chikhman, V., Bondarko, V., Danilova, M., Goluzina, A., & Shelepin, Y. (2012). Complexity of images: Experimental and computational estimates compared. *Perception*, *41*(6), 631–647. <https://doi.org/10.1068/p6987>

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology*, 65(2), 171–237. <https://doi.org/10.1006/jecp.1996.2356>
- Donderi, D. C. (2006). An information theory analysis of visual complexity and dissimilarity. *Perception*, 35(6), 823–835. <https://doi.org/10.1068/p5249>
- Edmonds, B. (1995). What is complexity? The philosophy of complexity per se with application to some examples in evolution. In F. Heylighen & D. Aerts (Eds.), *The evolution of complexity* (pp. 1–13). Dordrecht, Netherlands: Kluwer.
- Eerola, T., Himberg, T., Toiviainen, P., & Louhivuori, J. (2006). Perceived complexity of Western and African folk melodies by Western and African listeners. *Psychology of Music*, 34(3), 337–371. <https://doi.org/10.1177/0305735606064842>
- Eng, H. Y., Chen, D., & Jiang, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin and Review*, 12(6), 1127–1133. <https://doi.org/10.3758/BF03206454>
- Ferrand, L. (2000). Reading aloud polysyllabic words and nonwords: The syllabic length effect reexamined. *Psychonomic Bulletin & Review*, 7(1), 142–148. <https://doi.org/10.3758/BF03210733>
- Ferrand, L., & New, B. (2003). Syllabic length effects in visual word recognition and naming. *Acta Psychologica*, 113(2), 167–183. [https://doi.org/10.1016/S0001-6918\(03\)00031-3](https://doi.org/10.1016/S0001-6918(03)00031-3)
- Forsythe, A. (2009). Visual complexity: Is that all there is? In D. Harris (Eds.), *Engineering psychology and cognitive ergonomics* (pp. 158–166). Berlin, Germany: Springer. https://doi.org/10.1007/978-3-642-02728-4_17
- Forsythe, A., Mulhern, G., & Sawey, M. (2008). Confounds in pictorial sets: The role of complexity and familiarity in basic-level picture processing. *Behavior Research Methods*, 40(1), 116–129. <https://doi.org/10.3758/BRM.40.1.116>
- Forsythe, A., Sheehy, N., & Sawey, M. (2003). Measuring icon complexity: An automated analysis. *Behavior Research Methods, Instruments, & Computers*, 35(2), 334–342. <https://doi.org/10.3758/BF03202562>
- French, R. S. (1954). Identification of dot patterns from memory as a function of complexity. *Journal of Experimental Psychology*, 47, 22–26. <https://doi.org/10.1037/h0061098>
- García, M., Badre, A. N., & Stasko, J. T. (1994). Development and validation of icons varying in their abstractness. *Interacting With Computers*, 6(2), 191–211. [https://doi.org/10.1016/0953-5438\(94\)90024-8](https://doi.org/10.1016/0953-5438(94)90024-8)
- Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences*, 21(6), 803–831. <https://doi.org/10.1017/S0140525X98001769>
- Hannon, E. E., Soley, G., & Ullal, S. (2012). Familiarity overrides complexity in rhythm perception: A cross-cultural comparison of American and Turkish listeners. *Journal of Experimental Psychology: Human Perception and Performance*, 38(3), 543–548. <https://doi.org/10.1037/a0027225>
- Heaps, C., & Handel, S. (1999). Similarity and features of natural textures. *Journal of Experimental Psychology: Human Perception and Performance*, 25(2), 299–320. <https://doi.org/10.1037/0096-1523.25.2.299>
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434–446. <https://doi.org/10.1016/j.jml.2007.11.007>
- Jakesch, M., & Leder, H. (2015). The qualitative side of complexity: Testing effects of ambiguity on complexity judgments. *Psychology of Aesthetics, Creativity, and the Arts*, 9(3), 200–205. <https://doi.org/10.1037/a0039350>
- Juphard, A., Carbonnel, S., & Valdois, S. (2004). Length effect in reading and lexical decision: Evidence from skilled readers and a developmental dyslexic participant. *Brain and Cognition*, 55(2), 332–340. <https://doi.org/10.1016/j.bandc.2004.02.035>
- Kemps, E. (1999). Effects of complexity on visuo-spatial working memory. *European Journal of Cognitive Psychology*, 11(3), 335–356. <https://doi.org/10.1080/713752320>
- Lewandowsky, S., & Oberauer, K. (2015). Rehearsal in serial recall: An unworkable solution to the nonexistent problem of decay. *Psychological Review*, 122(4), 674–699. <https://doi.org/10.1037/a0039684>
- Liu, T., Chen, W., Liu, C. H., & Fu, X. (2012). Benefits and costs of uniqueness in multiple object tracking: The role of object complexity. *Vision Research*, 66, 31–38. <https://doi.org/10.1016/j.visres.2012.06.009>
- Luria, R., Sessa, P., Gotler, A., Jolicoeur, P., & Dell'Acqua, R. (2010). Visual short-term memory capacity for simple and complex objects. *Journal of Cognitive Neuroscience*, 22(3), 496–512. <https://doi.org/10.1162/jocn.2009.21214>
- Machado, P., Romero, J., Nadal, M., Santos, A., Correia, J., & Carballal, A. (2015). Computerized measures of visual complexity. *Acta Psychologica*, 160, 43–57. <https://doi.org/10.1016/j.actpsy.2015.06.005>
- McDougall, S. J., Curry, M.B., & De Bruijn, O. (1999). Measuring symbol and icon characteristics: Norms for concreteness, complexity, meaningfulness, familiarity, and semantic distance for 239 symbols. *Behavior Research Methods, Instruments, & Computers*, 31(3), 487–519. <https://doi.org/10.3758/BF03200730>
- McDougall, S. J., De Bruijn, O., & Curry, M. B. (2000). Exploring the effects of icon characteristics on user performance: The role of icon concreteness, complexity, and distinctiveness. *Journal of Experimental Psychology: Applied*, 6(4), 291–306. <https://doi.org/10.1037/1076-898X.6.4.291>
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97. <https://doi.org/10.1037/h0043158>
- Nadal, M., Munar, E., Marty, G., & Cela-Conde, C. J. (2010). Visual complexity and beauty appreciation: Explaining the divergence of results. *Empirical Studies of the Arts*, 28(2), 173–191. <https://doi.org/10.2190/EM.28.2.d>
- New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review*, 13(1), 45–52. <https://doi.org/10.3758/BF03193811>
- North, A. C., & Hargreaves, D. J. (1995). Subjective complexity, familiarity, and liking for popular music. *Psychomusicology: A Journal of Research in Music Cognition*, 14(1/2), 77–93. <https://doi.org/10.1037/h0094090>
- Oliva, A., Mack, M. L., Shrestha, M., & Peeper, A. (2004). Identifying the perceptual dimensions of visual complexity of scenes. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 1041–1046). Austin, TX: Cognitive Science Society.
- Palumbo, L., Ogden, R., Makin, A. D. J., & Bertamini, M. (2014). Examining visual complexity and its influence on perceived duration. *Journal of Vision*, 14(3), 1–18. <https://doi.org/10.1167/14.14.3>
- Popov, V., & Reder, L. M. (2020). Frequency effects on memory: A resource-limited theory. *Psychological Review*, 127(1), 1–46. <https://doi.org/10.1037/rev0000161>
- Purchase, H. C., Freeman, E., & Hamer, J. (2012). An exploration of visual complexity. In P. Cox, B. Plimmer, & P. Rodgers (Eds.),

- Diagrammatic representation and inference* (pp. 200–213). Berlin, Germany: Springer. https://doi.org/10.1007/978-3-642-31223-6_22
- Qian, G., Reinking, D., & Yang, R. (1994). The effects of character complexity on recognizing Chinese characters. *Contemporary Educational Psychology, 19*(2), 155–166. <https://doi.org/10.1006/ceps.1994.1014>
- Rao, A. R., & Lohse, G. L. (1993). Identifying high level features of texture perception. *CVGIP: Graphical Models and Image Processing, 55*(3), 218–233. <https://doi.org/10.1006/cgip.1993.1016>
- Reder, L. M., Liu, X. L., Keinath, A., & Popov, V. (2016). Building knowledge requires bricks, not sand: The critical role of familiar constituents in learning. *Psychonomic Bulletin & Review, 23*(1), 271–277. <https://doi.org/10.3758/s13423-015-0889-1>
- Reder, L. M., Paynter, C., Diana, R.A., Ngiam, J., & Dickison, D. (2007). Experience is a double-edged sword: A computational model of the encoding/retrieval trade-off with familiarity. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 48, pp. 271–312). New York, NY: Elsevier. [https://doi.org/10.1016/S0079-7421\(07\)48007-0](https://doi.org/10.1016/S0079-7421(07)48007-0)
- Rock, I., Halper, F., & Clayton, T. (1972). The perception and recognition of complex figures. *Cognitive Psychology, 3*(4), 655–673. [https://doi.org/10.1016/0010-0285\(72\)90025-4](https://doi.org/10.1016/0010-0285(72)90025-4)
- Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision, 7*(2), 1–22. <https://doi.org/10.1167/7.2.17>
- Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception, 33*(2), 217–236. <https://doi.org/10.1068/p5117>
- Shen, Z., Popov, V., Delahay, A. B., & Reder, L. M. (2018). Item strength affects working memory capacity. *Memory & Cognition, 46*(2), 204–215. <https://doi.org/10.3758/s13421-017-0758-4>
- Simon, H. A. (1974). How big is a chunk? *Science, 183*, 482–488. <https://doi.org/10.1126/science.183.4124.482>
- Sirois, M., Kremin, H., & Cohen, H. (2006). Picture-naming norms for Canadian French: Name agreement, familiarity, visual complexity, and age of acquisition. *Behavior Research Methods, 38*, 300–306. <https://doi.org/10.3758/BF03192781>
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Normal for name agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 174–215.
- Song, J. H., & Jiang, Y. (2006). Visual working memory for simple and complex features: An fMRI study. *NeuroImage, 30*(3), 963–972. <https://doi.org/10.1016/j.neuroimage.2005.10.006>
- State Language Work Committee. (1988). *Standard list of common characters in modern Chinese*. Beijing, China: Language Institute Publisher (in Chinese).
- Su, Y. F., & Samuels, S. J. (2010). Developmental changes in character-complexity and word-length effects when reading Chinese script. *Reading and Writing, 23*(9), 1085–1108. <https://doi.org/10.1007/s11145-009-9197-3>
- Sun, H., Zimmer, H. D., & Fu, X. (2011). The influence of expertise and of physical complexity on visual short-term memory consolidation. *Quarterly Journal of Experimental Psychology, 64*(4), 707–729. <https://doi.org/10.1080/17470218.2010.511238>
- Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*(2), 123–138. <https://doi.org/10.1007/s10648-010-9128-5>
- Tuch, A. N., Bargas-Avila, J. A., Opwis, K., & Wilhelm, F. H. (2009). Visual complexity of websites: Effects on users' experience, physiology, performance, and memory. *International Journal of Human Computer Studies, 67*(9), 703–715. <https://doi.org/10.1016/j.ijhcs.2009.04.002>
- Xing, J., & Manning, C. A. (2005). *Complexity and automation displays of air traffic control: Literature review and analysis*. Washington, DC: FAA Office of Aerospace Medicine.
- Yoon, S. H., Lim, J., & Ji, Y. G. (2015). Assessment model for perceived visual complexity of automotive instrument cluster. *Applied Ergonomics, 46*, 76–83. <https://doi.org/10.1016/j.apergo.2014.07.005>
- Zhang, G., & Simon, H. A. (1985). STM capacity for Chinese words and idioms: Chunking and acoustical loop hypotheses. *Memory & Cognition, 13*(3), 193–201. <https://doi.org/10.3758/BF03197681>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.