Journal of Memory and Language 66 (2012) 361–375

Contents lists available at SciVerse ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

A familiar finding: Pseudowords are more familiar but no less recollectable than words

Jason D. Ozubko [a,*], Andrew P. Yonelinas [b]

[a] Rotman Research Institute, Baycrest Centre, Canada
[b] University of California, Davis, United States

ARTICLE INFO

Article history:
Received 31 March 2011
revision received 31 October 2011
Available online 23 December 2011

Keywords:
Recognition memory
Pseudoword effect
Recollection and familiarity

ABSTRACT

The pseudoword effect is the finding that pseudowords (i.e., pronounceable nonwords) tend to give rise to more hits and false alarms than words. The familiarity-based account attributes this effect to the fact that pseudowords lack distinctive semantic meanings, which increases the inter-item similarity of pseudowords compared to words and thereby, increases the familiarity of pseudowords at test. However, studies examining recollection and familiarity of pseudowords have not consistently observed increased familiarity and several studies have reported effects on recollection rather than familiarity. One issue with these prior studies is that overall recognition accuracy is rarely equated between words and pseudowords. Examining the pseudoword effect in three experiments where we control for overall accuracy, we find greater familiarity for pseudowords and no recollective difference. In a fourth experiment, we find a recollection advantage for words when overall recognition accuracy is better for words than pseudowords. These results are consistent with the familiarity-based account of the pseudoword effect, as well as explaining why past studies have provided mixed support for this account.

Introduction

The pseudoword effect is the finding that, in the context of recognition memory experiments, pseudowords (i.e., pronounceable nonwords such as GRAWK or HENSION) tend to give rise to more hits and false alarms than words (Greene, 2004; Joordens, Ozubko, & Niewiadomski, 2008; Ozubko & Joordens, 2011; Wixted, 1992). One well supported type of explanation for this effect are so-called familiarity-based accounts (Ozubko & Joordens, 2011). Familiarity-based accounts of the pseudoword effect suggest, somewhat counter-intuitively, that the effect arises because pseudowords are more familiar than words, on average (Greene, 2004, 2007; Joordens et al., 2008; Ozubko & Joordens, 2008, 2011). Especially, these accounts are not suggesting that pseudowords are more pre-experimentally familiar than words, but instead that this familiarity advantage arises from study, as a result of the properties of pseudowords.

That is, familiarity-based accounts begin by noting that pseudowords differ from words in that they lack clear semantic meanings. As such, subjects cannot use semantic information to differentiate orthographically or phonemically similar pseudowords, and this ultimately leads to an inflated sense of familiarity. Consider the words HOUSE and HORSE. Despite the orthographic similarity, these words seem quite distinct due to their very different meanings. Now consider the pseudowords GRAWK and GLAWK. These pseudowords seem much more similar to one another, despite the fact that they differ to the same orthographic degree as HOUSE and HORSE. These cases exemplify the notion behind familiarity-based accounts of the pseudoword effect: Namely, by lacking clear semantic meanings, pseudowords are hypothesized to be more similar to one another than words. This increased inter-item similarity means that any given pseudoword test

* Corresponding author. Address: Rotman Research Institute, Baycrest Centre, 3560 Bathurst Street, Toronto, Ontario, Canada M6A 2E1.
E-mail address: jdozubko@rotman-baycrest.on.ca (J.D. Ozubko).

0749-596X/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2011.11.002

probe, old or new, will be more likely than a word test probe to match the study list as a whole. As a result, pseudowords feel more familiar to subjects than words, during a memory test (Ozubko & Joordens, 2008, 2011).

An alternative explanation for the pseudoword effect is the overcompensation account (e.g., Hockley & Niewiadomski, 2001; Stretch & Wixted, 1998). This account suggests that pseudowords appear more familiar than words in recognition data because subjects adopt a more lax response criterion for pseudowords. Specifically, the overcompensation account suggests that subjects believe pseudowords to be less memorable than words. Because of this fact, subjects adopt a lower response criterion for pseudowords and this in turn inflates both hits and false alarms of pseudowords. The overcompensation account actually represents one instantiation of a response-bias account of the pseudoword effect. That is, at a more general level the pseudoword effect could be accounted for by any account that suggested that subjects are using a more lax criterion for pseudowords than for words.

Both familiarity-based and response-bias accounts predict the basic pseudoword effect. However, a pseudoword effect has been observed in forced-choice recognition (Greene, 2004), a finding that is not predicted by simple criterion-shifting response-bias accounts but is compatible with familiarity-based accounts. Additionally, when pseudowords are made obviously more memorable at study, either through repetition or directed-forgetting of words, the pseudoword effect is not reduced or eliminated as the overcompensation account would predict (Greene, 2004). In another line of work, increasing the inter-item similarity (i.e., the mechanism proposed to underlie the pseudoword effect by the familiarity-based account) of cartoon characters in a recognition memory paradigm was found to increase both the hits and false alarms to those characters (Ozubko & Joordens, 2008). Finally, recent simulations comparing computational implementations of the overcompensation account and the familiarity-based account have found that familiarity-based accounts are compatible with many nuances regarding the pseudoword effect and recognition memory of nonwords, whereas overcompensation accounts miss such subtleties (Ozubko & Joordens, 2011).

Although it is very difficult to definitively rule out a response-bias account, it is our contention that the current evidence favors a familiarity-based account of the pseudoword effect. Given this fact, we will adopt this theoretical vantage point when discussing the pseudoword effect and experiments in this manuscript. The issue of the overcompensation account and response-bias accounts will be returned to in “General Discussion”. For now then, we will press forward and focus on the familiarity-based account of the pseudoword effect.

Generally speaking, familiarity-based accounts of the pseudoword effect suggest that the pseudoword effect results from a general inflation in familiarity at test. That is, during a recognition memory test, familiarity for old and new pseudowords should be greater than familiarity for old and new words respectively. Furthermore, the familiarity-based account makes no explicit predictions that the familiarity difference between words and pseudowords should be anything more than a global difference. Hence, the increase in familiarity should be a similar magnitude for old and new words and this difference in familiarity should ultimately act to increase both hits and false alarms of pseudowords above that of words, giving rise to the pseudoword word effect.

Looking at past studies that have directly investigated the subjective nature of recollection and familiarity for words and pseudowords, we find mixed evidence for the familiarity-based account. For example, one method that can be used to gauge the influences of recollection and familiarity is the remember/know paradigm (Tulving, 1985). The remember/know paradigm is a modification of normal yes/no recognition: subjects are asked to separate their “old” responses into “remember” and “know” responses. Subjects are to respond “remember” when episodic memory detail associated with study is present for a test probe and “know” when an item feels as though it was studied, but no episodic detail is present. Hence, the presence of episodic detail is used as a measure of recollection and is indexed by “remember” responses, whereas “know” responses index familiarity in the absence of recollection. Although “know” responses then, do not provide a direct measure of familiarity, using the independent remember/know (IRK) method (e.g., Jacoby, Yonelinas, & Jennings, 1997; Mangels, Picton, & Craik, 2001; Ochsner, 2000; Yonelinas & Jacoby, 1995), general familiarity can be estimated from “know” responses (see Experiment 1 for a more detailed discussion of this method).

Using the remember/know procedure Greene (2004), Rajaram, Hamilton, and Bolton (2002), and Whittlesea and Williams (2000) found more “know” responses to both old and new pseudowords compared to words. And these results hold when familiarity estimates using the IRK method are evaluated. Hence, these three studies provide some support for familiarity-based accounts of the pseudoword effect. One caveat with these data however, is that Whittlesea and Williams did not use typical remember/know instructions. That is, instead of indicating remember and know, subjects were asked to indicate if they could “recall seeing that item” or “feels familiar”. Nonetheless, these data suggest that familiarity was generally greater for pseudowords than words, supporting the notion that the pseudoword effect is a general familiarity increase.

On the other hand, several studies have failed to find clear evidence for a familiarity-based account of the pseudoword effect. Perfect and Dasgupta (1997) found no difference in “know” responses between words and pseudowords. Although this may suggest that words and pseudowords did not differ in terms of familiarity, familiarity estimates calculated using the IRK method show that familiarity was actually greater for words than pseudowords. This finding actually stands in stark disagreement with the predictions of the familiarity-based account of the pseudoword effect.

Similarly, Gardiner and Java (1990) also investigated the subjective memorability of words and pseudowords using the remember/know procedure. Although “know” responses were higher for studied pseudowords compared to words, there was no difference in terms of new items. This finding was consistent with familiarity estimates calculated with the IRK method. Hence, although some increase in familiarity was observed for pseudowords, it was an increase in familiarity-based memorability, not a general increase, as the familiarity account would predict.

The five studies cited represent the entire literature regarding the recollective and familiarity-based memory of pseudowords. Although these studies stand in some disagreement with regards to the familiarity-based differences between words and pseudowords, all five of these studies do agree in one aspect: pseudowords are less recollectable than words. That is, in all studies, "remember" responses were found to be significantly lower for studied pseudowords compared to studied words (Gardiner & Java, 1990; Greene, 2004; Perfect & Dasgupta, 1997; Rajaram et al., 2002; Whittlesea & Williams, 2000). Although the familiarity-based account of the pseudoword effect is agnostic regarding possible changes in recollection, and therefore, this finding is not inconsistent with the account, the account is also unable to explain these data. And as this finding is the only consistent finding across all five studies examining recollection and familiarity of pseudowords, these preliminary studies seemingly show no overall support for the familiarity-based account of the pseudoword effect.

One confound in many of the previous studies of the pseudoword effect is that they did not control for levels of overall performance. This is particularly relevant due to the fact that recognition accuracy is often worse for pseudowords than words. For example, in an analysis of 35 experiments from 13 previous published recognition memory studies examining words and pseudowords, although not explicitly reported at the time, Ozubko and Joordens (2011) found that in 72% of experiments pseudowords were less memorable than words. Indeed, pseudowords were less memorable than words in the data reported by Greene (2004), Perfect and Dasgupta (1997), Rajaram et al. (2002), and Whittlesea and Williams (2000). Given overall performance differences, it is relatively unsurprising that the less memorable stimulus set would also show a recollection disadvantage. The recollective disadvantage that is often observed for pseudowords then may actually not be an inherent characteristic of pseudowords at all, but rather just a reflection of overall performance differences.

Complicating the interpretation of all of these findings however, is the fact that the inconsistent remember/know results occur across studies, regardless of stimulus consistency. That is, both Greene (2004) and Whittlesea and Williams (2000) used highly-word-like, multi-syllable pseudowords. Although Greene and Whittlesea and Williams found similar recollective and familiarity-based influences for pseudowords, Perfect and Dasgupta (1997) used a similar set of pseudowords and found very different results from Greene or Whittlesea and Williams. Similarly, the results of Gardiner and Java (1990) disagree with those from Rajaram et al. (2002) despite the fact that Rajaram et al. used the exact same pseudowords as Gardiner and Java in their experiments. Thus, the conflicting estimates of recollection and familiarity appear to occur regardless of whether stimuli are help constant across studies. What then, are the important methodological differences that could have lead to such conflicting results? One possibility here is that studies using the same stimuli have produced inconsistent results due to instructional differences in the remember/know procedure.

The remember/know paradigm is currently a somewhat controversial procedure. Although remember/know responses have been argued to represent a qualitative distinction and do not map onto simple measures of confidence (e.g., see Eichenbaum, Yonelinas, & Ranganath, 2007; Perfect, Mayes, Downes, & Van Eijk, 1996; Rajaram, 1993; Skinner & Fernandes, 2007; Yonelinas, 2002) criticisms have been raised as to whether these responses provide proper estimates of recollection and familiarity respectively (e.g., Donaldson, 1996; Hirshman & Master, 1997; Inoue & Bellezza, 1998; Rotello & Zeng, 2008; Wixted, 2007; Wixted & Stretch, 2004). Interestingly, the instructions given to subjects regarding remember and know responses may play a pivotal role here, explaining why some researchers have found that remember/know responses reflect merely confidence whereas others have found remember/know responses provide measures of recollection and familiarity.

Specifically, Rotello, Macmillan, Reeder, and Wong (2005) have demonstrated that the detail of the instructions given to subjects in the remember/know paradigm can affect the reliability of subsequent "remember" responses. When standard remember/know instructions are given, "remember" responses do not converge with estimates of recollection gathered from receiver operating characteristic (ROC) curves, and "remember" responses are relatively prevalent for new items (a finding which should not occur in most paradigms if "remember" responses truly index conscious recollection). However, when more conservative instructions are given to subjects, "remember" responses do provide good measures of recollection that converge with ROC estimates. Other work has similarly shown that conservative remember/know instructions lead to "remember" responses that converge with estimates of recollection as measured by source discrimination and ROC estimates (Yonelinas, 2001; Yonelinas, Dobbins, Szymanski, Dhaliwal, & King, 1996; and see Yonelinas (2002) for a review).

Given the evidence that standard remember/know instructions are inadequate to properly measure recollection via "remember" responses, the results of all previous remember/know studies using pseudowords may be suspect. And if "remember" responses are inaccurate, because the IRK method uses the proportion of "remember" responses to arrive at estimates of familiarity, these familiarity estimates are also suspect. Indeed, even though Gardiner and Java (1990) did examine words and pseudowords with similar levels of overall memorability, the remember/know instructions used here were probably insufficient to accurately gauge recollection and familiarity. All of this is not to say that none of the previous work examining pseudowords in the remember/know paradigm is valid, but more that it is unknown how much of the previous work may have been distorted due to insufficiently detailed instructions.

Given the somewhat convoluted literature, and possibly suspect findings, on the recollective and familiarity-based influences of the pseudoword effect to date, the current work seeks to investigate these influences in a controlled setting. First, the overall recognition accuracy for words and pseudowords was intentionally equated in our experiments. With overall performance equated, we can more precisely compare the recollective and familiarity-based memorability of words and pseudowords, and differences between the two stimuli cannot be attributed to overall performance differences. The familiarity-based account of the pseudoword effect does not rely on any performance differences between words and pseudowords, and hence, we still expect to see a familiarity-based increase to both hits and false alarms for pseudowords when overall performance is equated. Furthermore, since this account also makes no predictions regarding recollection, and we have no other basis to expect pseudowords to be less recollectable than words, we do not expect to see a recollective deficit for pseudowords when overall performance is equated.

Second, conservative remember/know instructions are given to subjects in all experiments reported here. The goal is for "remember" responses to more accurately index recollection, than may have been the case in past studies where standard remember/know instructions were used. Once again, Rotello et al. (2005) as well as Yonelinas and colleagues (Yonelinas, 2001; Yonelinas et al., 1996) have demonstrated that when more conservative instructions are given to subjects, "remember" responses do provide good measures of recollection that converge with ROC estimates. Although we focus on using the remember/know procedure with conservative instructions, it is prudent to validate our approach and findings, using an alternative means to measure recollection and familiarity. Thus, in Experiment 2 we will use the ROC procedure to estimate the influences of recollection and familiarity for words and pseudowords to compare to our other experiments. In this manner, Experiment 2 should demonstrate that our estimates of recollection and familiarity, as obtained using the remember/know paradigm, are indeed similar to those obtained using the ROC procedure, and hence, provide converging evidence as to their validity.

Finally, after carrying out three experiments with overall performance controlled, in Experiment 4 we will examine recollection and familiarity in the more typical case where pseudowords are less well recognized than words (cf. Ozubko & Joordens, 2011). This final experiment will allow us to assess which, if any, of the findings from Gardiner and Java (1990), Greene (2004), Perfect and Dasgupta (1997), Rajaram et al. (2002), or Whittlesea and Williams (2000) were attributable to the fact that words and pseudowords were not explicitly matched for overall performance. Turning now to Experiment 1, we seek to investigate recollection and familiarity for words and pseudowords with similar overall recognition accuracies, in the remember/know paradigm.

Experiment 1

The goal of Experiment 1 was to investigate recollective and familiarity-based memory for words and pseudowords using the remember/know procedure. Although past work has found a recollective deficit for pseudowords compared to words (Gardiner & Java, 1990; Greene, 2004; Perfect & Dasgupta, 1997; Rajaram et al., 2002; Whittlesea & Williams, 2000), few studies have actually investigated this issue and most of this work comes from cases where pseudowords and words differ in overall levels of performance. Thus, the ambiguous evidence regarding the familiarity of pseudowords and the recollective deficit seen for pseudowords may have been a result of overall performance differences, and nothing more. Furthermore, standard remember/know instructions do not always lead to reliable estimates of recollection from "remember" responses. Because most past studies have used standard remember/know instructions, estimates of recollection from this past work may be distorted, to some degree. In our experiments, conservative remember/know instructions, that have been shown to lead "remember" responses that reliably index recollection (Rotello et al., 2005; Yonelinas, 2001; and see Yonelinas (2002) for a review), were used. Hence, as the familiarity-based account makes no predictions about recollection being impaired for pseudowords, we did not expect to find such a deficit in our experiments.

Before proceeding however, some discussion of the IRK method is warranted. Specifically, because the remember/know procedure involves having subjects respond "know" only when recollection has failed, "know" responses themselves do not provide a direct measure of familiarity. That is, the proportion of "know" responses observed are limited by the proportion of "remember" responses made. In situations where "remember" responses increase but the level of familiarity is constant, "know" responses must necessarily decline. If "know" responses are taken as an index of familiarity, they will indicate that familiarity is declining in this situation, when in fact it is not.

As a result of this dependency, various researchers have called for the adoption of the independent remember/know (IRK) method (e.g., Jacoby et al., 1997; Mangels et al., 2001; Ochsner, 2000; Yonelinas & Jacoby, 1995). In IRK, "remember" responses are assumed to estimate recollection whereas familiarity is estimated as the proportion of "know" responses divided by the proportion of non-"remember" responses. It has been demonstrated that estimates of familiarity from this paradigm are more in agreement with estimates of familiarity from other paradigms, such as ROC analysis and the inclusion/exclusion paradigm, than raw "know" responses are (Yonelinas, 2001; Yonelinas & Jacoby, 1995; Yonelinas, Kroll, Dobbins, Lazzara, & Knight, 1998). Hence, in Experiment 1 we used the IRK method for assessing recollection and familiarity.

Words and pseudowords that produced equivalent overall recognition accuracy were selected to be used in Experiment 1 from Joordens et al. (2008). In this way, any recollective or familiarity-based differences observed in Experiment 1 could not be attributable to overall performance differences. Finally, once again, the familiarity-based account predicts that pseudowords should have a higher hit and false alarm rate than words, due to the fact that pseudowords are generally more familiar than words. Furthermore, because this account makes no explicit predictions about recollection, and we have no other reason to expect a recollective effect, we expect no recollective difference between words and pseudowords.

Methods

Participants

Forty three subjects from participated in the experiment online for 0.5 credit towards a psychology course. Subjects were drawn from the subject pools at both the University of California, Davis and the University of Waterloo.

Materials

A set of 120 pseudowords and 120 words were selected from Joordens et al. (2008) on the basis of overall recognition accuracy (see Appendix A). Words and pseudowords were selected from the high-experimental/high-language based feature overlap (high–high) condition and the low-experimental/high-language based feature overlap (low–high) condition. Stimuli from the high–high condition had a relatively high degree of orthographic overlap with other stimuli of the same class and with a large corpus of dictionary words. Stimuli from the low–high condition had a low degree of orthographic overlap with other stimuli of the same class but a large degree of overlap with a large corpus of dictionary words.

Words from the high–high condition had a mean hit and false alarm rate of .57 (*SD* = .22) and .27 (*SD* = .22) respectively. Pseudowords from the high–high condition had a mean hit and false alarm rate of .72 (*SD* = .23) and .41 (*SD* = .20) respectively. Overall recognition accuracy for the words and pseudowords of these two conditions was very close (i.e., d' for words was 0.79 and for pseudowords was 0.81). Words from the low–high condition had a mean hit and false alarm rate of .68 (*SD* = .17) and .30 (*SD* = .17) respectively. Pseudowords from the low–high condition had a mean hit and false alarm rate of .71 (*SD* = .21) and .35 (*SD* = .14) respectively. Overall recognition accuracy for the words and pseudowords of these two conditions was also very close (i.e., d' for words was 0.99 and for pseudowords was 0.94). Hence, words and pseudowords selected for the stimulus set were expected to be very close in terms of overall recognition accuracy.

The study set was comprised of 60 words and 60 pseudowords. The test set was comprised of the study set and the remaining 60 words and 60 pseudowords. The study set was randomly selected once, and was counterbalanced across subjects.

Procedure

Pilot testing revealed that subjects had difficulty following conservative remember/know instructions over a long test list of 240 items. Because maintaining the validity of "remember" responses was of paramount importance, instead of presenting a study phase of 120 items and a test phase of 240 items, Experiment 1 was presented to subjects in four study-test blocks. Each block consistent of a study phase which included 30 items (15 words and 15 pseudowords). Words and pseudowords were randomly intermixed, and appeared individually in the center of a computer screen for 2.5 s, with a 0.25 s inter-stimulus interval. Subjects were instructed to remember the words and pseudowords for an upcoming test. In the test phase of each block, the 15 studied words and 15 studied pseudowords were randomly intermixed with 15 new words and 15 new pseudowords, resulting in a test with 60 items. Such a short study-test sessions may seem unusual however previous studies examining pseudowords in the remember/know paradigm have often used very short lists (i.e., often using only 15–20 words and 15–20 pseudowords at study; Gardiner & Java, 1990; Perfect & Dasgupta, 1997; Rajaram et al., 2002; Whittlesea & Williams, 2000; or 30 words and 30 pseudowords at study; Greene, 2004).

Before the test phase began, subjects were informed that items would appear individually in the center of the computer screen and that they would need to recognize those items. The remember/know procedure was explained to subjects in detail, and subjects verbally confirmed their understanding of the procedure by explaining it back to the researcher, in their own words. The researcher corrected any mistakes the subject made and provided further instructions/examples if necessary. Subjects were informed that for each "remember" response given they may need to justify those responses at the end of the experiment by telling the researcher what exactly they recollected although these post-test responses were never actually collected.

During the test phase of each block, individual words appeared in the center of the screen with the labels "remember", "know", and "new" at the bottom of the screen. Subjects pressed "m" to indicate "remember", "b" to indicate "know", and "c" to indicate "new". Between each test trial was a 0.5 s inter-stimulus interval. No words or pseudowords repeated across blocks.

After completing a study-test block, subjects watched a 5–7 min clip of an episode of the NBC television program *The Office*. The clips were all taken from the same episode and were shown in sequential order. These clips were provided to prevent the buildup of proactive interference across the study-test blocks, and to keep subjects motivated. After completing the fourth block, subjects were thanked for their participation and debriefed.

Results and discussion

An alpha level of .05 was used as our criterion for significance in all significance tests. Effect size estimates were computed using partial η^2 ($p\eta^2$) or Cohen's *d* where appropriate. All experiments reported in this paper collected data from four study-test block sessions for each subject. The results of the first block were always in agreement with the collapsed results of the four study-test blocks. Due to this result, and because block was not an interest of study, data from the four study-test blocks were collapsed for all subjects in this and all subsequent experiments.

First, overall hit and false alarm rates were obtained by collapsing "remember" and "know" responses into "old" responses for each subject. These hit and false alarm rates were used to calculate d' as an overall measure of memorability. Mean hit rates, false alarm rates, and d' scores can be seen in Table 1. Overall recognition accuracy (d') was successfully equated between words and pseudowords, $t(42) = 1.43$, $p = .16$, $d = 0.22$.

Next, the presence of the pseudoword effect was assessed by examining overall hit and false alarm rates. Examining these data in a 2 (old vs. new) × 2 (word vs.

Table 1
Mean hit rates, false alarm rates, and d' for words and pseudowords in Experiments 1 through 4. Note that standard errors are shown in brackets below the means.

Condition	p("old"|old)		p("old"|new)		d'	
	Words	Pseudowords	Words	Pseudowords	Words	Pseudowords
Experiment 1	.69 (.02)	.80 (.02)	.18 (.02)	.33 (.02)	1.55 (0.09)	1.43 (0.06)
Experiment 2	.71 (.02)	.81 (.02)	.22 (.02)	.33 (.02)	1.42 (0.08)	1.43 (0.08)
Experiment 3	.81 (.02)	.89 (.02)	.17 (.03)	.25 (.03)	2.14 (0.03)	2.17 (0.14)
Experiment 4	.87 (.02)	.86 (.02)	.13 (.02)	.27 (.03)	2.61 (0.18)	1.99 (0.13)

pseudoword) within-subjects ANOVA, we find that hit rates were higher than false alarm rates, $F(1,42) = 718.27$, $MSe = 0.01$, $p\eta^2 = .95$, that both hit and false alarm rates were higher for pseudowords than words, $F(1,42) = 66.18$, $MSe = 0.01$, $p\eta^2 = .61$, and that there was no interaction, $F(1,42) = 2.30$, $MSe = 0.01$, $p = .14$, $p\eta^2 = .05$. Thus, the pseudoword effect was observed in Experiment 1, with more "old" responses to pseudowords than words in general.

Although a pseudoword effect was present in Experiment 1, the critical question to ask is why this effect arose. That is, how did the recollection and familiarity of pseudowords differ from that of words, to give rise to the pseudoword effect? Based on the "remember" and "know" responses, measures of recollection and familiarity were calculated using the IRK method (Jacoby et al., 1997; Mangels et al., 2001; Ochsner, 2000; Yonelinas & Jacoby, 1995). Recollection (R) was measured as the proportion of "remember" responses; familiarity (F) was measured as the proportion of "know" responses divided by the proportion of non-"remember" responses [i.e., F = p("know")/(1 − p("remember"))]. These estimates were calculated separately for each subject for studied and unstudied words and pseudowords. These estimates of recollection and familiarity are shown in Fig. 1A.

According to a familiarity-based account of the pseudoword effect, hits and false alarms are higher for pseudowords than words because pseudowords are generally more familiar than words. Examining familiarity responses in a 2 (old vs. new) × 2 (words vs. pseudowords) within-subjects ANOVA, we find that familiarity responses were greater for old vs. new items, $F(1,42) = 323.07$, $MSe = 0.02$, $p\eta^2 = .89$. Furthermore, consistent with the familiarity-based account of the pseudoword effect, there were more familiarity responses to pseudowords than words, $F(1,42) = 50.75$, $MSe = 0.02$, $p\eta^2 = .55$. There was no

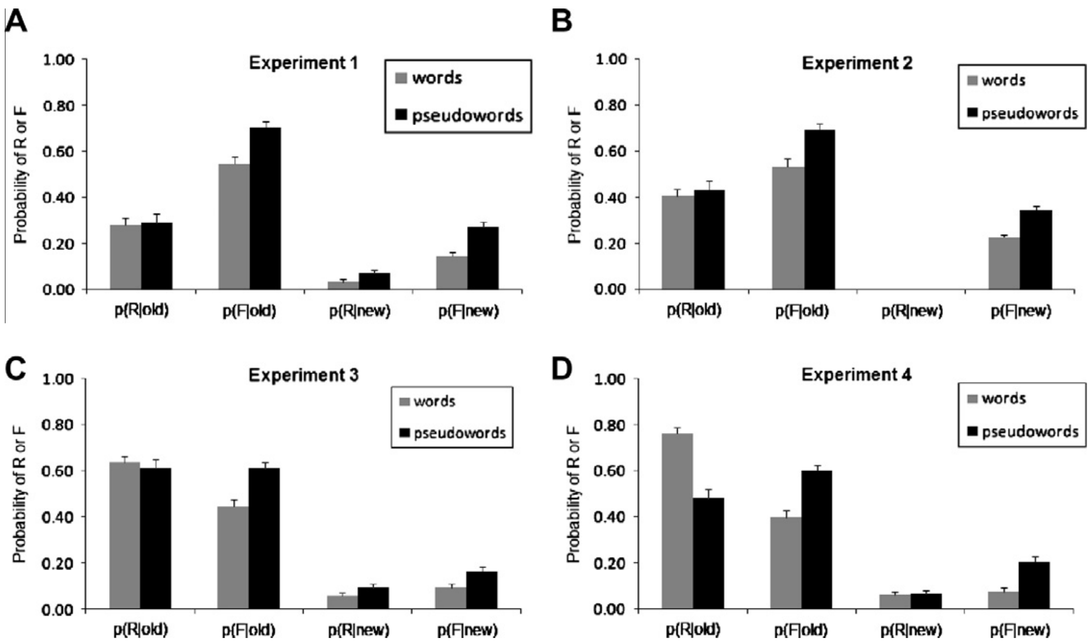

Fig. 1. Mean estimates of recollection (R) and familiarity (F) from Experiments 1 through 4. Estimates in Experiments 1, 3, and 4 were calculated using the independent remember/know method. Estimates in Experiment 2 were calculated using the dual-process signal detection model. Error bars are represented with standard error of the means.

interaction, $F(1,42) = 1.77$, $MSe = 0.01$, $p = .19$, $p\eta^2 = .04$. Hence, Experiment 1 is consistent with the notion that pseudowords are generally more familiar than words, and that this effect is giving rise to the pseudoword effect. However, before concluding on this fact, we examined the recollective responses to determine their contribution to overall performance.

Recollection responses were examined in a 2 (old vs. new) × 2 (words vs. pseudowords) within-subjects ANOVA. Although recollection responses were more frequent for old than new items, $F(1,42) = 80.33$, $MSe = 0.03$, $p\eta^2 = .66$, there was no main effect of stimulus type and no interaction, both F's < 1. Hence, although past work has found a recollective deficit for pseudowords compared to words (Gardiner & Java, 1990; Greene, 2004; Perfect & Dasgupta, 1997; Rajaram et al., 2002; Whittlesea & Williams, 2000), we find no difference here. Given that the familiarity-based account makes no predictions that recollection should differ between words and pseudowords and that past studies did not equate words and pseudowords based on overall performance or use conservative remember/know instructions, we argue that these findings suggest that pseudowords are not inherently less recollectable than words. Past work demonstrating a recollective difference were likely demonstrating this result due to differences in overall recognition accuracy or perhaps as a result of standard remember/know instructions.

Experiment 2

The results of Experiment 1 demonstrate that, when overall recognition accuracy between words and pseudowords is controlled, a pseudoword effect arises because of inflated familiarity to pseudowords at test. To estimate the influences of familiarity and recollection, Experiment 1 used the remember/know procedure. Although this provides a connection with past studies examining recollection and familiarity of pseudowords (Gardiner & Java, 1990; Greene, 2004; Perfect & Dasgupta, 1997; Rajaram et al., 2002; Whittlesea & Williams, 2000), it does mean that the remember/know procedure is the only procedure that has ever been used to investigate recollection and familiarity of pseudowords. The goal of Experiment 2 was therefore to provide converging evidence with the results of Experiment 1 by using a completely different methodology to assess recollection and familiarity than the remember/know procedure. Thus, in Experiment 2 we use the ROC procedure and the dual-process signal detection model to estimate the influences of recollection and familiarity for words and pseudowords (Yonelinas, 1994, 1997).

The remember/know procedure assumes that subjects can consciously differentiate between recollection and familiarity, and this assumption has been questioned (e.g., Donaldson, 1996; Hirshman & Master, 1997; Inoue & Bellezza, 1998; although also see Perfect et al., 1996; Skinner & Fernandes, 2007; Yonelinas, 2002; Yonelinas & Jacoby, 1995). For this reason we repeated Experiment 1, but used an ROC method to estimate recollection and familiarity, in order to determine if the results of experiment would generalize across methods. Based on the results of Experiment 1, and the fact that remember/know instructions that are known to lead to responses that provide good measures of recollection and familiarity were used, we expect that Experiment 2 should confirm the findings of Experiment 1. Namely, the pseudoword effect should result due to a general difference in familiarity, and no difference in recollection for words and pseudowords should be observed.

Methods

Participants

Thirty four subjects from participated in the experiment online for 0.5 credit towards a psychology course. Subjects were drawn from the subject pools at both the University of California, Davis and the University of Waterloo.

Materials

The stimuli used were identical to those in Experiment 1.

Procedure

Experiment 2 was run identically to Experiment 1 except that instead of recognizing items on the remember/know/new scale, subjects were asked to indicate recognition confidence from 1 to 6 for each item. Subjects were instructed to press 4, 5, or 6 if they believed a word to have been studied—6 if they were "absolutely sure," 5 if they were "very sure," and 4 if they were "somewhat sure." Similarly, subjects were instructed to press 1, 2, or 3 if they believed a word to be new. Subjects pressed 1 if they were "absolutely sure," 2 if they were "very sure," and 3 if they were "somewhat sure." Subjects were asked to think carefully about their confidence ratings and to try to use the entire scale over the course of the recognition test.

Results and discussion

To examine the influence of familiarity and recollection, ROC curves were plotted and analyzed for words and pseudowords. ROC analysis involves plotting hit rates against false alarm rates across various response criteria. ROC curves were plotted separately for words and pseudowords; these are shown in Fig. 2. In ROC analyses, the first data point is obtained by considering only the strictest response criterion (i.e., only "6" responses to be hits or false alarms). For both words and pseudowords then, the hit rates at this level of confidence were plotted at the level of false alarms. These points are plotted as the two far left points in Fig. 2 (i.e., the first empty and first filled circle). The next point is obtained by considering a slightly less strict criterion (i.e., either "5" or "6" responses to be hits or false alarms). For both words and pseudowords then, the hit rates at this level of confidence were plotted at the level of false alarms and can be seen as the second two points from the left in Fig. 2. This process is repeated until the most lax criterion has been plotted (i.e., when a response from "2" to "6" is considered a hit or false alarm). ROC analysis does not plot a point where responses from "1" to "6" are considered hit or false alarms because these points would always add to 1.0, and thus would provide no

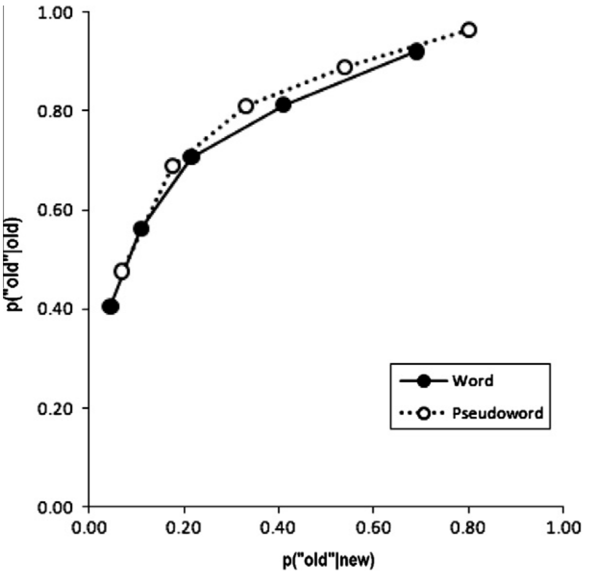

Fig. 2. The recognition receiver operating characteristics (ROCs) for words and pseudowords from Experiment 2.

useful information. Hence, from a 6-point confidence scale, 5-point ROC curves can be plotted.

Overall hit and false alarm rates were obtained by collapsing 6 (''absolutely sure studied''), 5 (''very sure studied''), and 4 (''somewhat sure studied''). Parallel to Experiment 1, these hit and false alarm rates were used to calculate d' as an overall measure of memorability. Mean hit rates, false alarm rates, and d' scores can be seen in Table 1. Overall memorability was successfully equated between words and pseudowords, $t(42) = 0.07$, $p = .95$, $d = 0.03$.

To test whether a pseudoword effect was present, a 2 (old vs. new) × 2 (words vs. pseudowords) within-subjects ANOVA was conducted on hit rates and false alarm rates separately. Overall there were more hits than false alarms, $F(1,33) = 504.30$, $MSe = 0.02$, $p\eta^2 = .94$. More importantly however, both hit and false alarms were greater for pseudowords than for words, $F(1,33) = 34.91$, $MSe = 0.01$, $p\eta^2 = .51$, and these factors did not interact, $F < 1$. Hence, a pseudoword effect was observed in the hit and false alarm rate data. Moreover, an examination of the ROC curves in Fig. 2 shows that the ROC points in the pseudoword conditions were shifted up and to the right compared to the word ROC. This suggests that for each level of confidence, hits and false alarms were greater for pseudowords than for words.[1]

Given that a pseudoword effect was present, we now turn to analyses of recollection and familiarity to see

Table 2
Mean recollective disriminability (R), familiarity d' estimated from ROC curves, and familiarity d' estimated from F values for words and pseudowords in Experiment 2. Note that standard errors are shown in parentheses below the means.

Measure	Words	Pseudowords
R	.39 (.03)	.41 (.04)
Familiarity d' (from ROC)	0.84 (0.09)	0.92 (0.11)
Familiarity d' (from F)	0.89 (0.09)	0.93 (0.10)

how these may have contributed to the effect. The dual-process signal-detection model can be used to fit ROC curves and yield estimates of both recollection and familiarity (Yonelinas, 1994, 1997) and has been used in the past to obtain measures of recollection and familiarity that can be compared to the output of remember/know procedures (Yonelinas & Jacoby, 1995; Yonelinas et al., 1998). This model assumes that recollection is a threshold process, and therefore estimates of recollection (R) are provided on a probability scale, the same as in the remember/know procedure. However, the model further assumes that familiarity is best described as a signal-detection process, not a threshold process. In other words, the probability of accepting an item as ''old'' based on familiarity is a function of how much more familiar studied items are relative to new items. Hence, the model produces estimates of familiarity in terms of d', not raw probabilities. The dual-process signal-detection model was fit to the ROC curves for words and pseudowords separately, to yield separate estimates of recollection and familiarity discriminability for the two stimulus types. These estimates are shown in Table 2.

Overall, neither recollective nor familiarity-based discriminability differed between words and pseudowords, $t(33) = 0.81$, $d = 0.13$, $p = .42$ and $t(33) = 1.06$, $d = 0.19$, $p = .30$ respectively. Hence, Experiment 2 finds no evidence that words or pseudowords differ in terms of recollective or familiarity-based *discriminability*.[2] Why then did the pseudoword effect arise? Recall that the ROC curve for pseudowords was shifted up and to the right of the ROC curve for words (see Fig. 2). This means that at each level of confidence, hit and false alarm rates were greater for pseudowords than for words. As well, both ROCs appear to fall on the same function, and so no memorability difference in either R or familiarity d' would be expected. Hence, this shift does indicate a general familiarity difference. In essence, pseudowords may have been more familiar to subjects than words despite the fact that discriminability (i.e., the difference between familiarity for old and new items) was equivalent.

[1] Note that formal analyses of the ROC curve at these different confidence levels generally does confirm that hits and false alarms were always greater than words, $F(1,132) = 24.35$, $MSe = 0.03$, $p\eta^2 = .43$. These analyses are not reported in detail however, because the hit and false alarm rates at these different confidence levels are not independent. Hence, testing hit and false alarm rates at each confidence level is statistically inappropriate as well as somewhat redundant with the overall hit and false alarm rate comparison that is reported.

[2] Note that we do not mean to say that there was no difference in recollection or familiarity between words and pseudowords at all, merely that the discriminability (i.e., the difference between old and new items) of words or pseudowords based on recollection and familiarity was equivalent.

In order to directly test this notion, we used the hit and false alarm rates and R estimates from Experiment 2 to estimate F responses to old and new items. Specifically, the dual-process signal detection model assumes that hit rates are equal to the influence of R plus $(1 - R)F_{old}$. Furthermore, false alarms are equal to F_{new}. Using the overall hit and false alarm rates, as well as R, from Experiment 2, we can solve for F_{old} and F_{new} (and these estimates should be conceptually similar to those of Experiment 1). F_{old} and F_{new} estimates were thus calculated, and can be seen in Fig. 1B, along with R estimates. Note that although this method assumes that there are no R values for new items, the F estimates here did provide an excellent fit to the observed data. Specifically, to check if the F_{old} and F_{new} estimates obtained were reasonable, we used these F values for words and pseudowords to recalculate familiarity-based d'. For words and pseudowords the familiarity d' values calculated from F estimates can be seen in Table 2. These values did not significantly differ from the familiarity d' values estimated from the ROC curves themselves, $t(33) = 1.61$, $d = 0.28$, $p = .12$ and $t(33) = 0.64$, $d = 0.06$, $p = .53$ respectively. Furthermore, the recalculated familiarity d' values for words and pseudowords were highly correlated with the familiarity d' values for words and pseudowords from the ROC curves, $r(33) = .94$ and $r(33) = .96$ respectively. Hence, F_{old} and F_{new} estimates appear to provide reasonable estimates of the familiarity of old and new words and pseudowords.

Examining F responses in a 2 (old vs. new) × 2 (words vs. pseudowords) within-subjects ANOVA, we find that familiarity responses were greater for old vs. new items, $F(1,33) = 120.72$, $MSe = 0.03$, $p\eta^2 = .79$. Furthermore, consistent with the familiarity-based account of the pseudoword effect and Experiment 1, there were more familiarity responses to pseudowords than words, $F(1,33) = 42.55$, $MSe = 0.01$, $p\eta^2 = .56$. There was no interaction, $F(1,33) = 2.60$, $MSe = 0.004$, $p = .12$, $p\eta^2 = .07$. Hence, like Experiment 1, Experiment 2 provides good support for the notion that pseudowords are generally more familiar than words, and that this effect is giving rise to the pseudoword effect.

Experiment 3

Across two studies we have found evidence in favor of a familiarity-based explanation of the pseudoword effect. Furthermore, we have yet to observe a recollective deficit for pseudowords that was cited in earlier studies (i.e., Gardiner & Java, 1990; Greene, 2004; Perfect & Dasgupta, 1997; Rajaram et al., 2002; Whittlesea & Williams, 2000). In the previous two experiments however, the encoding of words and pseudowords was left uncontrolled. It remains possible that the differences in memorability observed between words and pseudowords rests with differential encoding strategies. Therefore, Experiment 3 enforces similar encoding on both words and pseudowords. Namely, subjects are given rote rehearsal instructions, which should be equally applicable to words and pseudowords. If natural encoding strategies do differ between words and pseudowords, we should not expect to replicate the findings of the previous experiments in Experiment 3, where the encoding of words and pseudowords is encouraged to be similar.

Methods

Participants

Twenty seven subjects from participated in the experiment online for 0.5 credit towards a psychology course. Subjects were drawn from the subject pools at both the University of California, Davis and the University of Waterloo.

Materials

The stimuli in Experiment 3 were the same as those used in Experiments 1 and 2.

Procedure

The procedure of Experiment 3 was identical to that of Experiment 1 except that subjects were encouraged to rote rehearse both words and pseudowords. Specifically, subjects were told to read each item as it appears on the screen, but then to repeat that item several times.

Results and discussion

Overall hit and false alarm rates were obtained by collapsing "remember" and "know" responses into "old" responses for each subject. These hit and false alarm rates were used to calculate d' as an overall measure of memorability. Mean hit rates, false alarm rates, and d' scores can be seen in Table 1. Overall memorability was successfully equated between words and pseudowords, $t(26) = 0.23$, $p = .82$, $d = 0.04$.

The presence of the pseudoword effect was assessed by examining overall hit and false alarm rates. Examining these data in a 2 (old vs. new) × 2 (word vs. pseudoword) within-subjects ANOVA, we find that hit rates were higher than false alarm rates, $F(1,26) = 606.12$, $MSe = 0.02$, $p\eta^2 = .96$, that both hit and false alarm rates were higher for pseudowords than words, $F(1,26) = 23.21$, $MSe = 0.01$, $p\eta^2 = .47$, and that there was no interaction, $F < 1$. Thus, the pseudoword effect was observed in Experiment 3, with more "old" responses to pseudowords than words in general.

Based on the "remember" and "know" responses, measures of recollection (R) and familiarity (F) were calculated using the IRK method (Jacoby et al., 1997; Mangels et al., 2001; Ochsner, 2000; Yonelinas & Jacoby, 1995). These estimates can be seen in Fig. 1C. Examining familiarity responses in a 2 (old vs. new) × 2 (words vs. pseudowords) within-subjects ANOVA, we find that familiarity responses were greater for old vs. new items, $F(1,26) = 87.74$, $MSe = 0.05$, $p\eta^2 = .77$. Furthermore, consistent with the familiarity-based account of the pseudoword effect, there were more familiarity responses to pseudowords than words, $F(1,26) = 18.98$, $MSe = 0.02$, $p\eta^2 = .42$. As well, although there was a interaction, $F(1,26) = 6.51$, $MSe = 0.01$, $p\eta^2 = .20$, follow-up analyses confirmed that despite the interaction there were more familiarity responses to pseudowords than words for both old and new items, $t(26) = 4.04$, $d = 0.79$ and $t(26) = 3.08$, $d = 0.63$

respectively. Hence, Experiment 3 supports the results of Experiment 1 and 2: Pseudowords were more generally more familiarity than words during the recognition memory test.

Further confirming our previous findings, recollection responses were examined in a 2 (old vs. new) × 2 (words vs. pseudowords) within-subjects ANOVA. Although recollection responses were more frequent for old than new items, $F(1,26) = 156.70$, $MSe = 0.05$, $p\eta^2 = .86$, there was no main effect of stimulus type and no interaction, $F < 1$ and $F(1,26) = 1.32$, $MSe = 0.02$, $p = .26$, $p\eta^2 = .05$ respectively. Hence, as with Experiments 1 and 2, we once again see no recollective advantage for words over pseudowords when overall recognition accuracy is matched. Experiment 3 therefore provides no evidence that subjects employ differential encoding strategies when studying words and pseudowords. When subjects were required to encode words and pseudowords using the same encoding method, as in Experiment 3, the results replicated our previous findings.

Experiment 4

Across three experiments so far, when overall recognition accuracy of words and pseudowords is equated, we find no evidence of a recollective deficit for pseudowords. This suggests that previous studies that have found a recollective deficit for pseudowords may have found this effect because of general performance differences between words and pseudowords. Although this seems like a reasonable conclusion, an alternative explanation is that insufficiently strict remember/know instructions, or some other element of previous studies were the culprit for the recollective deficit for pseudowords. The goal of Experiment 4 was to directly test the notion that the recollective deficit may have arisen due to the fact that words and pseudowords were not equated for overall memorability in much past work.

If the recollective deficit for pseudowords is the result of general performance differences between words and pseudowords, then we should be able to see this deficit using the exact same stimulus set and methods that we have used thus far, providing the overall memorability of words is selectively enhanced. Thus, in order to maintain identical stimulus sets and methods with the previous experiments, but to selectively enhance the memory of words over that of pseudowords, we opted for an instructional manipulation: Namely, semantic vs. non-semantic encoding.[3]

In Experiment 4 we encourage subjects to encode words according to their semantic meanings at study, and pseudowords through simple rote rehearsal, as in Experiment 3. Past studies on elaboration have shown that focusing on encoding the semantic meaning of stimuli can lead to superior memory compared to focusing on encoding surface characteristics of stimuli (i.e., levels of processing; Craik & Lockhart, 1972; Craik & Tulving, 1975). In fact, not only is general memorability enhanced due to semantic encoding, but recollective memory is often enhanced more than familiarity-based memory (see Yonelinas, 2002).

Given past work with elaboration, we expect that the encoding manipulation in Experiment 4 should make words more memorable than pseudowords, specifically by enhancing the recollection advantage of words over pseudowords. Hence, in Experiment 4 we expect to replicate the recollective advantage of words over pseudowords that past studies have noted (Gardiner & Java, 1990; Greene, 2004; Perfect & Dasgupta, 1997; Rajaram et al., 2002; Whittlesea & Williams, 2000). Furthermore, because pseudowords are being rehearsed in Experiment 4 just as in Experiment 3, we expect to observe similar results for pseudowords in Experiment 4 as in Experiment 3.

Methods

Participants

Thirty one subjects from participated in the experiment online for 0.5 credit towards a psychology course. Subjects were drawn from the subject pools at both the University of California, Davis and the University of Waterloo.

Materials

The stimuli in Experiment 4 were the same as those used in Experiments 1 through 3.

Procedure

The procedure of Experiment 4 was identical to that of Experiment 3 except that subjects were encouraged to rehearse words and pseudowords differently. Specifically, although subjects were still told to read each pseudoword as it appears and repeat in several times, subjects were instructed to think of the meanings of words and to visualize them. Hence, subjects were instructed to encode words more deeply and semantically than pseudowords.

Results and discussion

Overall hit and false alarm rates were obtained by collapsing "remember" and "know" responses into "old" responses for each subject. These hit and false alarm rates were used to calculate d' as an overall measure of memorability. Mean hit rates, false alarm rates, and d' scores can be seen in Table 1. As was intended, overall recognition accuracy (d') was greater for words than pseudowords, $t(29) = 3.62$, $d = 0.69$.

The presence of the pseudoword effect was assessed by examining overall hit and false alarm rates. Examining these data in a 2 (old vs. new) × 2 (word vs. pseudoword) within-subjects ANOVA, we find that hit rates were higher

[3] Note that an alternative method here would be to select new words and pseudowords that differed in terms of overall memorability and then test those in Experiment 4. One concern with this method however, is that by selecting new stimuli, the results of Experiment 4 wouldn't be comparable to the previous experiments reported here because the new pseudoword set would could easily differ from the prior set on other important dimensions beyond memorability. Furthermore, whether words are rendered more memorable than pseudowords via selecting a new set or via an encoding manipulation on the existing set, the end result is the same: the words in Experiment 4 are subjectively more memorable to subjects. Because this could be accomplished using the exact same words as in the previous experiments, it was the preferred method over selecting new words and/or pseudowords.

than false alarm rates, $F(1,29) = 753.53$, $MSe = 0.02$, $p\eta^2 = .96$. Furthermore, although there was a main effect of stimulus type, $F(1,29) = 13.48$, $MSe = 0.02$, $p\eta^2 = .31$, this effect interacted with old/new status, $F(1,29) = 22.90$, $MSe = 0.01$, $p\eta^2 = .43$. Specifically, although false alarms were greater for pseudowords than words, $t(29) = 5.13$, $d = 0.95$, hit rates did not differ, $t(29) = 0.51$, $p = .62$, $d = 0.09$. Thus, a full pseudoword effect was not observed in Experiment 4.

Although it should not be surprising that a full pseudoword effect was not found considering the memorability of words was intentionally increased above that of pseudowords, the familiarity-based account makes specific predictions about recollection and familiarity in this scenario. Specifically, the familiarity-based account suggests that a full pseudoword effect is observed in situations where pseudowords are more familiar than words, and this familiarity acts to increase both hits and false alarms above that of words. Given that false alarms were still inflated for pseudowords in Experiment 4, this account would still predict that familiarity was generally inflated for pseudowords over words. Hence, familiarity should still be greater for old pseudowords than words. However, because hit rates did not differ between words and pseudowords, the familiarity-based account would require that words be more recollectable than pseudowords. The question then, is do any of these predictions hold?

Based on the "remember" and "know" responses, measures of recollection (R) and familiarity (F) were calculated using the IRK method (Jacoby et al., 1997; Mangels et al., 2001; Ochsner, 2000; Yonelinas & Jacoby, 1995). These estimates can be seen in Fig. 1D. Examining familiarity responses in a 2 (old vs. new) × 2 (words vs. pseudowords) within-subjects ANOVA, we find that familiarity responses were greater for old vs. new items, $F(1,29) = 87.65$, $MSe = 0.04$, $p\eta^2 = .75$. Furthermore, there were more familiarity responses to pseudowords than words, $F(1,29) = 18.27$, $MSe = 0.05$, $p\eta^2 = .39$. There was no interaction, $F(1,29) = 1.46$, $MSe = 0.03$, $p = .24$, $p\eta^2 = .05$.

Recollection estimates were examined in a 2 (old vs. new) × 2 (words vs. pseudowords) within-subjects ANOVA. Recollection responses were more frequent for old than new items, $F(1,29) = 340.66$, $MSe = 0.03$, $p\eta^2 = .92$. Furthermore, although there was a main effect of stimulus type, $F(1,29) = 14.44$, $MSe = 0.04$, $p\eta^2 = .33$, this interacted with old/new status, $F(1,29) = 18.85$, $MSe = 0.03$, $p\eta^2 = .39$. Follow-up analyses confirmed that, while there was no difference in recollection responses to new items, $t(29) = 0.12$, $p = .90$, $d = 0.02$, there was a recollection impairment for pseudowords compared to words, $t(29) = 4.13$, $d = 0.79$.

Thus, as was predicted by the familiarity-based account of the pseudoword effect, due to the fact that overall hit rates did not differ between words and pseudowords, although there was still a familiarity advantage for pseudowords over words there was a recollection deficit for pseudowords compared to words. Considering that the same pseudowords that did not give rise to a recollective deficit in Experiment 1 through 3 did give rise to a recollective deficit in Experiment 4, it is unlikely that any inherent characteristic of pseudowords themselves makes them more or less recollectable than words. The current results therefore demonstrate that the recollective deficit often observed for pseudowords may simply be the result of overall memorability differences between words and pseudowords, and the tendency in the literature to select pseudowords which are less memorable than words.

General discussion

According to a familiarity-based account of the pseudoword effect, the effect arises because pseudowords are generally more familiar than words during recognition memory tests. This familiarity differential acts to increase both hits and false alarms of pseudowords above that of words, thus leading to the pseudoword effect. Little work has directly tested the recollective and familiarity-based influences of pseudowords, and what work has been done, provides mixed support for the familiarity-based account (Gardiner & Java, 1990; Greene, 2004; Perfect & Dasgupta, 1997; Rajaram et al., 2002; Whittlesea & Williams, 2000). However, these past studies did not explicitly match words and pseudowords based on overall recognition accuracy or use conservative remember/know instructions (something that is arguably crucial in order to obtain good estimates of recollection and familiarity from the remember/know paradigm). Across 3 experiments where the overall recognition performance of words and pseudowords was equated and either conservative remember/know instructions or the ROC procedure were used, we found evidence for inflated familiarity for pseudowords compared to words.

Furthermore, past studies have consistently found a recollective advantage for words over pseudowords (Gardiner & Java, 1990; Greene, 2004; Perfect & Dasgupta, 1997; Rajaram et al., 2002; Whittlesea & Williams, 2000). However, none of our experiments where overall performance was matched between words and pseudowords and where conservative remember/know instructions were used found such an effect. As Ozubko and Joordens (2011) note, pseudowords that are used in recognition memory studies are often less memorable than words. When words were more memorable than pseudowords (Experiment 4), we did indeed see the recollective deficit that was found in previous work. This result suggests that recollective impairments seen for pseudowords in the past may have simply been the result of overall performance differences, rather than due to any inherent property of pseudowords themselves. Although it should be added that Gardiner and Java (1990) found a recollective impairment for pseudowords that were of roughly equivalent overall memorability as words, it is possible that the different outcome in that study was related to the use of the traditional remember/know instructions used in that study.

Consistency with past work on the pseudoword effect

In reviewing the literature regarding the pseudoword effect, Ozubko and Joordens (2011) noted that the hit rate portion of the pseudoword effect was often smaller than the false alarm rate portion. That is, pseudowords tended to produce a hit rate about 3–5% higher than words, and a false alarm rate about 11% higher than words. Despite the fact that this is the typical finding, in Experiments 1 through 3 we saw an equivalent difference for hits and false alarms between words and pseudowords and in Experiment 4 we observed no difference in hit rates despite a pseudoword effect still in false alarms. Thus, how do these results fit with the more typical patterns observed?

In actuality, our results are not inconsistent with the previously mentioned trends. Firstly, in Experiments 1 through 3 pseudowords were intentionally selected to be of comparable memorability as a set of words. In most studies examining the pseudoword effect, pseudowords are not controlled in this manner and are less memorable than words (Ozubko & Joordens, 2011). Thus, given the controls around memorability in our experiments, we would not necessarily expect to replicate the same nuances as the previous literature. More important however, are the results of Experiment 4.

Experiment 4 was setup to induce the same memory differential that is typically seen between words and pseudowords. Hence, memory for words was selectively enhanced. In this scenario, we would expect to produce the same small hit rate difference but larger false alarm rate difference that is typically observed. Therefore, to remain consistent with the previous literature, in Experiment 4 (compared to Experiments 1 through 3) either the false alarm rate difference would have needed to increase, or the hit rate difference would have needed to decrease. Consistent with this second possibility, in Experiment 4 the hit rate portion of the pseudoword effect did significantly decrease compared to Experiments 1 through 3 (to the point where it was eliminated in fact). It may be somewhat surprising to observe no hit rate difference in Experiment 4 however, in their review, Ozubko and Joordens (2011) found that 17% of studies examining pseudowords found either no hit rate difference or a greater hit rate for words compared to pseudowords. Thus, it is not that unusual that no hit rate difference was found in Experiment 4. Like much past literature then, we reported a smaller (in fact non-existent) hit rate difference despite a significant false alarm rate difference.

Another important point to remember here is that, although we are arguing that pseudowords are not any actually less recollectable than words, in many practical circumstances pseudowords may demonstrate less recollection than words. Specifically, in modeling work examining the pseudoword effect, Ozubko and Joordens (2011) demonstrated that as stimuli are changed from being word-like (i.e., having semantic details) to being more pseudoword-like (i.e., having less semantic details), not only do hits and false alarms increase, but the difference between hits and false alarms decreases. In essence, it may be a relatively natural state of affairs for pseudowords to be less memorable than words. And in this circumstance, we would expect pseudowords would also be less recollectable than words. Thus, pseudowords may often demonstrate less recollection than words, although once again, we would argue that this is likely due to overall recognition accuracy differences rather than an inherently necessary property of pseudowords.

The selection of pseudowords

One issue worth considering in more detail is our selection of pseudoword stimuli. Specifically, pseudowords were selected as pronounceable nonwords from past studies, with the restriction that they match a word sets level of overall memorability. This approach was adopted given our interest in “remember” and “know” responses in our experiments—if words and pseudowords were not matched based on overall memorability, it would become difficult to interpret recollective or familiarity differences since the base-rate of memorability would also be changing. However, selecting pseudowords in this manner likely meant that while the pseudowords matched words in terms of overall memorability, they may have mismatched words on other dimensions, and this may have contributed to the pattern of data observed above and beyond the matching for memorability. For example, some of the pseudowords in this study were drawn from the high-experimental based feature overlap condition of Joordens et al. (2008) whereas others were drawn from the low-experimental based feature overlap condition. As a result, perhaps the pseudowords in our experiments have a wider range of experimental based feature overlap (i.e., orthographic overlap with other words and pseudowords in the experiment) than in most past studies.

Although it remains possible that a wider range of experimental based feature overlap than is usual, or some other important dimension or dimensions were changed by controlling for memorability, it is hard to say for certain. In fact, it is hard to say for certain what dimensions are essential to pseudowords across most of the literature. That is, in almost all past work with pseudowords, the pronounceability/orthographic regularity of pseudowords has been the only dimension consistently controlled for.

For example, Gardiner and Java (1990) created a set of pronounceable nonwords as pseudowords, which were controlled to be 4 letters long and one syllable in length (e.g., JOSP, LORT, KLIB; see also Rajaram et al., 2002); Perfect & Dasgupta, 1997 created a different set of pronounceable nonwords to be used as pseudowords (e.g., PENDON, FRUMSTLE) with little description of how these were constructed; and Whittlesea and Williams (2000) created yet another set of orthographically regular nonwords to be used as pseudowords (e.g., HENSION, FRAMBLE, BARDEN; see also Greene, 2004) with little description as to how those items were constructed. These few examples demonstrate that, very often, no special dimensions are considered or controlled beyond pronounceability/orthographic regularity. This is not to say that other dimensions do not matter. Indeed, our experiments here demonstrate the importance of controlling one dimension, memorability. Instead we would suggest that considering other important dimensions that may influence the memorability of pseudowords is a task that should be continued. In fact, future work may be able to delineate other important dimensions worth considering in regards to pseudowords, that we have not considered here. Nonetheless, we would suggest that our work is a first step in the process of considering the other properties important to pseudowords and memorability.

Response-bias alternative accounts

Earlier we introduced the overcompensation account as an alternative account of the pseudoword effect. By this account, the pseudoword effect arises because subjects believe that pseudowords are less memorable than words, and as a result adopt a more lax recognition criterion for pseudowords. Although the current data was not meant as a critical test of this account, it has some bearing on it. Namely, in Experiments 1 through 3, recognition accuracy was equated for words and pseudowords, yet a pseudoword effect stile arose. If we assume that subjects were able to pick up on the fact that words and pseudowords did not differ in terms of memorability (especially considering that there were multiple study-test blocks to notice the memorability of words and pseudowords), then the overcompensation account would predict no general pseudoword effect. Yet, a pseudoword effect was found. We would not make the strong case against the overcompensation account based solely on this, somewhat speculative, evidence. However, in light of previous evidence demonstrating that the pseudoword effect is still observed even when words are more explicitly made less memorable (Greene, 2004), this work adds to the notion that subjects do not adopt different criteria for words and pseudowords in the face of memorability or perceived memorability differences between these stimuli.

Regardless of the evidence for or against the overcompensation account, the overcompensation account represents merely one instantiation of a response-bias account of the pseudoword. Response-bias accounts, in general, represent any account of the pseudoword effect that suggest it arises due to changes in criterion or responding that do not reflect a true memorability change. Considering our experiments, a simple dual-process response-bias account could explain our results. Namely, if we suppose that a separate criterion is selected for "remember" and "know" responses (or that "remember" responses have no criterion), then it becomes possible for subjects to have a lower "know"-response criterion for pseudowords compared to words, and this would explain all of our data. It is therefore, difficult to ignore the possibility that the familiarity increases we observed for pseudowords represented a reduction of a familiarity criterion compared to words, as opposed to a true increase in familiarity.

Thus, the data reported here do not rule out a response-bias approach to the pseudoword effect. However, now that the basic recollective and familiarity-based properties of the pseudoword effect have been laid out, future work may be able to focus on the familiarity differences between words and pseudowords, to examine whether this difference represents a true change in familiarity, as opposed to a criterion shift or some other response bias. Such work could provide a definitive answer as to whether the familiarity differences are best understood as a response bias, or a true difference in familiarity.

The remember/know paradigm

Before concluding we present a brief discussion of the remember/know paradigm itself. Namely, one of the key contributions of our work here is to clarify the relatively muddled remember/know findings regarding pseudowords. Past studies have reported mixed and indeed conflicting reports regarding pseudowords, even when identical pseudoword sets were used across studies. It is our belief that much of this inconsistency is probably due to inconsistent remember/know instructions across laboratories and eras. Indeed, as already discussed, the quality of the remember/know instructions can change the validity of the "remember" and "know" responses (Rotello et al., 2005). Although we would still advocate the remember/know procedure, it is important to understand its limitations. Care must be taken when instructing participants: After all, if subjects do not understand what you are asking them to discriminate between, how can they properly perform the task? Further, gaining converging evidence from other methods such as ROC curves, as we have done here, provide further evidence that your measures of recollection and familiarity are indeed accurate. Through both of these methods, perhaps the remember/know paradigm itself may gradually become less controversial once again.

Conclusion

The clear conclusion from our experiments is that the pseudoword effect does arise from a general increase in familiarity. The preferred interpretation is that this represents a true increase in overall familiarity, as suggested by the familiarity-based account of the pseudoword effect, although it could arise due to a response-bias. Nonetheless, we find no evidence of a recollective deficit between words and pseudowords beyond that which is observed when one stimulus set is generally less memorable than another. These findings suggest that the contradictory findings from past work examining recollection and familiarity of pseudowords may have been, at least partially, due to the fact that overall recognition accuracy was not controlled.

Acknowledgments

This research was supported by an Alexander Graham Bell Canada Graduate Scholarship and a Michael Smith Foreign Study Supplement, both from the Natural Sciences and Engineering Research Council of Canada, as well as funding from the National Institute of Mental Health (MH059352).

Appendix

Words and pseudowords used in Experiments 1 through 4.

Words				Pseudowords			
ACTION	FLAMINGO	MEMBER	RANGE	ANANRE	DEDETA	INSTON	PLANDER
ADDITION	FLOATING	MILLION	REACTION	ANCANT	DEEDAT	INTEEL	PLENDON
ANSWER	FORMER	MODERN	READING	ANLELA	DEESRE	IODITR	POTIMER
ATTENTION	FRIEND	MOMENT	RECORD	ANTIOU	DELETI	IOESER	PRAMIS
BARNACLE	FRONT	MONTH	REPORT	ARMAN	DELICON	IOROAL	PURDEN
BEGINNING	FUNCTION	MORNING	RESEARCH	ATATEL	DIONRI	ISATIT	RARECH
BETTER	GALLERY	MOTHER	RESPECT	ATATLI	DIROEN	ISHETE	REARLE
BILLION	GENERAL	MOUNTAIN	RESULT	ATEDIN	EDCARO	ISINOR	REDEON
BUSINESS	GREEN	NATION	RIDICULE	ATENLE	ELIOSE	ISTIAN	REESCA
CENTURY	HUMANE	NATURE	RIVER	ATITTE	ELTEST	ITARTI	RIANIC
CERTAIN	INCREASE	NOTHING	SECOND	ATOREL	ERCHEL	LAERED	RINGIC
CHANGE	INDUSTRY	NOTION	SECRETARY	BANDAL	ERICES	LEISER	ROCOEN
CONSCIOUS	INSIDE	OPERATION	SECTION	BARDEN	ERNEED	LERIES	ROGATION
CONSIST	ISLAND	ORDER	SENSE	BELLAND	ESALTA	LIINNG	ROLAES
CONTROL	JUSTICE	PAPER	SERIES	BINGLE	ESNTCA	LILECA	ROTICA
CORNER	KNITTING	PARTICLE	SERPENT	BINICAL	FISSEL	LILEED	SEICED
COUNTRY	LEADER	PATTERN	SERVICE	BLINDEN	FLEMIN	LITECH	SENDAL
DECISION	LENGTH	PERIOD	SPRING	BLISSEN	GARDER	LOMAND	SENEAL
DESERT	LETTER	PERSON	STAND	BRENDER	GRAMEN	MANIPER	SONDER
DIRECTION	LEVEL	POINT	STATEMENT	CADERI	HALBERT	MESSEL	STARTI
DISMOUNT	LIVING	POSTAGE	STREET	CALIDON	HALLID	MESTIC	TACHTI
DISTINCT	LOCATION	PRESENT	STRENGTH	CAMENT	HEANOU	NETEOU	TARRION
DOING	LONGER	PRESIDENT	STUDENT	CHITOU	HEANTR	ORNEDI	TECHED
EDIBLE	MANNER	PRESSURE	SUMMER	CHRICA	HENSION	ORORIS	TEDICH
EDUCATION	MARITIME	PRIVATE	TABLET	CLORAL	HEONDE	ORRADE	TRATIC
EVENING	MATERIAL	PROGRESS	THEATRE	CONDER	HETEHE	OUALIC	TRESPAT
EXTENT	MATURE	PROPERTY	THING	COOURA	ICARIS	OUESIC	TRISER
FASTEN	MEANING	QUESTION	TREATMENT	COOUST	ICATLA	OUNGLA	VERSAL
FATHER	MEANS	RADIO	TRIBUTE	CRABLE	INRENE	PELLIS	WIMBER
FEELING	MEETING	RAINBOW	WRITING	DECHIN	INRIOU	PENDON	WINDON

References

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior, 11*, 671–684.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*, 268–294.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24*, 523–533.

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience, 30*, 123–152.

Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition, 18*, 23–30.

Greene, R. L. (2004). Recognition memory for pseudowords. *Journal of Memory and Language, 50*, 259–267.

Greene, R. L. (2007). Foxes, hedgehogs, and mirror effects: The role of general principles in memory research. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger, III* (pp. 53–66). New York, NY: Psychology Press.

Hirshman, E., & Master, S. (1997). Modeling the conscious correlates of recognition memory: Reflections on the remember-know paradigm. *Memory & Cognition, 25*, 345–351.

Hockley, W. E., & Niewiadomski, M. W. (2001). Interrupting recognition memory: Tests of a criterion-change account of the revelation effect. *Memory & Cognition, 29*, 1176–1184.

Inoue, C., & Bellezza, F. S. (1998). The detection model of recognition using know and remember judgments. *Memory & Cognition, 26*, 299–308.

Jacoby, L. L., Yonelinas, A. P., & Jennings, J. M. (1997). The relation between conscious and unconscious (automatic) influences: A declaration of independence. In E. Jonathan, D. Cohen, & W. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 13–47). Hillsdale, NJ: Erlbaum.

Joordens, S., Ozubko, J. D., & Niewiadomski, M. W. (2008). Featuring old/new recognition: The two faces of the pseudoword effect. *Journal of Memory and Language, 58*, 380–392.

Mangels, J. A., Picton, T. W., & Craik, F. I. M. (2001). Attention and successful episodic encoding: An event-related potential study. *Cognitive Brain Research, 11*, 77–95.

Ochsner, K. N. (2000). Are affective events richly recollected or simply familiar? The experience and process of recognizing feelings past. *Journal of Experimental Psychology: General, 129*, 242–261.

Ozubko, J. D., & Joordens, S. (2008). Super Memory Bros: Going from mirror patterns to concordant patterns via similarity enhancements. *Memory & Cognition, 36*, 1391–1402.

Ozubko, J. D., & Joordens, S. (2011). The similarities (and familiarities) of pseudowords and extremely high frequency words: Examining a familiarity-based explanation of the pseudoword effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 123–139.

Perfect, T. J., & Dasgupta, Z. R. R. (1997). What underlies the deficit in reported recollective experience in old age? *Memory & Cognition, 25*, 849–858.

Perfect, T. J., Mayes, A. R., Downes, J. J., & Van Eijk, R. (1996). Does context discriminate recollection from familiarity in recognition memory? *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 49A*, 797–813.

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition, 21*, 89–102.

Rajaram, S., Hamilton, M., & Bolton, A. (2002). Distinguishing states of awareness from confidence during retrieval: Evidence from amnesia. *Cognitive, Affective & Behavioral Neuroscience, 2*, 227–235.

Rotello, C. M., Macmillan, N. A., Reeder, J. A., & Wong, M. (2005). The remember response: Subject to bias, graded, and not a process-pure indicator of recollection. *Psychonomic Bulletin & Review, 12*, 865–873.

Rotello, C. M., & Zeng, M. (2008). Analysis of RT distributions in the remember-know paradigm. *Psychonomic Bulletin & Review, 15*, 825–832.

Skinner, E. I., & Fernandes, M. A. (2007). Neural correlates of recollection and familiarity: A review of neuroimaging and patient data. *Neuropsychologia, 45*, 2163–2179.

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1379–1396.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26*, 1–12.

Whittlesea, B. W. A., & Williams, L. D. (2000). The source of feelings of familiarity: The discrepancy-attribution hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 547–565.

Wixted, J. T. (1992). Subjective memorability and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 681–690.

Wixted, J. T. (2007). Signal-detection theory and the neuroscience of recognition memory. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger, III* (pp. 67–82). New York, NY, US: Psychology Press.

Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review, 11*, 616–641.

Yonelinas, A. P. (1994). Receiver operating characteristics in recognition memory: Evidence for a dual process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1341–1354.

Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition, 25*, 747–763.

Yonelinas, Andrew P. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology: General, 130*, 361–379.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*, 441–517.

Yonelinas, A. P., Dobbins, I., Szymanski, M. D., Dhaliwal, H. S., & King, L. (1996). Signal-detection, threshold, and dual-process models of recognition memory: ROCs and conscious recollection. *Consciousness and Cognition: An International Journal, 5*, 418–441.

Yonelinas, A. P., & Jacoby, L. L. (1995). The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of Memory and Language, 34*, 622–643.

Yonelinas, A. P., Kroll, N. E. A., Dobbins, I., Lazzara, M., & Knight, R. T. (1998). Recollection and familiarity deficits in amnesia: Convergence of remember-know, process dissociation, and receiver operating characteristic data. *Neuropsychology, 12*, 323–339.