Published in final edited form as:

Am J Psychol. 2015 ; 128(2): 241–252.

Desirable Difficulties in Vocabulary Learning

ROBERT A. BJORK

University of California, Los Angeles

JUDITH F. KROLL

Pennsylvania State University

Abstract

In this article we discuss the role of desirable difficulties in vocabulary learning from two perspectives, one having to do with identifying conditions of learning that impose initial challenges to the learner but then benefit later retention and transfer, and the other having to do with the role of certain difficulties that are intrinsic to language processes, are engaged during word learning, and reflect how language is understood and produced. From each perspective we discuss evidence that supports the notion that difficulties in learning and imposed costs to language processing may produce benefits because they are likely to increase conceptual understanding. We then consider the consequences of these processes for actual second-language learning and suggest that some of the domain-general cognitive advantages that have been reported for proficient bilinguals may reflect difficulties imposed by the learning process, and by the requirement to negotiate cross-language competition, that are broadly desirable. As Alice Healy and her collaborators were perhaps the first to demonstrate, research on desirable difficulties in vocabulary and language learning holds the promise of bringing together research traditions on memory and language that have much to offer each other.

Vocabulary learning, viewed broadly, is fundamental to our initial and continued learning in almost every domain. We need to know the language, so to speak, not simply in the sense of learning a first or second language but also in the sense of learning the vocabulary that characterizes some field of study, such as biology or the law. Our primary concern as teachers, for example, may be increasing students' higher-level understanding of concepts in some domain and increasing their ability to generalize those concepts to new situations where they are relevant, but achieving those goals rests on students having acquired the basic vocabulary of terms and labels in that domain.

Perhaps understandably, then, vocabulary learning has often been examined in different ways by memory researchers and by language researchers. Memory researchers have frequently examined vocabulary learning using materials that are selected to ensure that the participants know little, typically nothing, of the to-be-learned language, such as having college students learn English translations of Swahili words, before the experiment. The goals of such experiments are to understand more about processes such as response learning,

Address correspondence about this article to Robert A. Bjork, Department of Psychology, University of California, Los Angeles, CA 90095 (; Email: rabjork@psych.ucla.edu).

stimulus learning, how forward and backward associations are formed, how materials should be sequenced and tested in order to optimize long-term retention, and so forth (see Bjork, Dunlosky, & Kornell, 2013). On the other hand, research on language processing has tended to examine the earliest processes during comprehension and in planning speech but typically not the later consequences of those processes (Kroll, Gullifer, & Rossi, 2013). Language researchers have also embraced and tried to understand the complexities of cross-language processes, complexities that have typically been avoided by memory researchers.

Alice Healy has been a counterexample to the preceding generalization about the differing approaches of memory researchers and language researchers. One of her distinct contributions is that she, together with her students and colleagues, has attempted to bring together these differing research traditions (Healy & Bourne, 1998, 2013). In this article we attempt to pursue that effort by examining whether and how certain difficulties that have been shown to increase long-term retention and transfer can be incorporated into vocabulary and language learning.

Learning, Performance, and Introducing Difficulties

A problem teachers and trainers confront—and a problem we all confront in managing our own learning—is that conditions of instruction or practice that make performance improve rapidly often fail to support long-term retention and transfer, whereas conditions of instruction that appear to create difficulties for the learner, slowing the rate of *apparent* learning, often optimize long-term retention and transfer. To the extent that we assume that current performance is a valid index of learning, we become susceptible to choosing less effective conditions of learning or practice over more effective conditions. Alice Healy, together with her students and colleagues, was among the first to emphasize that creating certain types of difficulties can improve learning and slow forgetting, especially in the domain of learning foreign-language vocabulary (Schneider, Healy, & Bourne, 2002).

Examples of instructional manipulations that create “desirable difficulties” (Bjork, 1994) include varying the conditions of learning rather than keeping conditions constant and predictable, distributing or spacing study or practice sessions rather than massing or blocking such sessions, using tests (rather than presentations) as learning events, reducing feedback to the learner, and providing “contextual interference” during learning (e.g., *interleaving* rather than *blocking* practice). It is important to emphasize that creating difficulties for ourselves or for those we are responsible for instructing is hardly a universal good. In real-world contexts, many, perhaps most, of the difficulties we create are undesirable. The difficulties introduced by variation, spacing, interleaving, and so forth are desirable because responding to those difficulties (successfully) engages the very processes that support learning, comprehension, and remembering. Even those difficulties, aside from difficulties that are universally undesirable, become undesirable if the learner, by virtue of prior knowledge and current cues, is not equipped to respond to them successfully.

There is much to be said about the role of creating difficulties, desirable and undesirable, in vocabulary learning, whether in one's own primary language or in learning foreign-language vocabulary. In this article we focus on the role of testing and test-induced errors in

vocabulary learning and on the role of creating "contextual interference" (Battig, 1978), which refers to arranging the conditions of learning so that, during the learning process, the possible sources of interference between separate to-be-learned materials and from external sources are maximized, not minimized. We conclude by speculating that bilingualism imposes difficulties that are desirable in quite general ways.

Testing Versus Restudying

It is common to think that the function of testing is assessment, but in recent years the role of testing as a pedagogical tool has received increasing emphasis (for a review, see Roediger & Karpicke, 2006a). Information and procedures that are retrieved in response to a test of some kind not only become more retrievable in the future than they would have been without such a test, but they even become more retrievable than if they had been presented again for restudy. In that sense, the act of retrieval is a potent learning event, and retrieval is a "memory modifier" (Bjork, 1975) in another sense as well: Retrieving targeted information can create retrieval-induced forgetting (Anderson, Bjork, & Bjork, 1994), that is, decreased subsequent access to information that is in competition with the retrieved information.

The advantages of testing over restudying tend not to be apparent in the short term, however. Typically, restudying appears more productive than testing during the learning process or shortly thereafter, whereas the advantages of testing over restudying become apparent only after a delay (see Hogan & Kintsch, 1971; Roediger & Karpicke, 2006b; Thompson, Wenger, & Bartling, 1978; Wheeler, Ewers, & Buonanno, 2003). It is important to add that the advantage of restudying over initial testing when retention is measured after a short retention interval is, in a sense, artificial: It arises not because restudying is ever more effective than retrieving, even at a short delay, but because restudying strengthens *all* of the to-be-remembered items, whereas initial testing strengthens only the items successfully retrieved.

In the context of vocabulary learning, for example, Storm, Friedman, Murayama, and Bjork (2014) had participants learn English translations of Swahili words (e.g., *mashua*, "boat") and again found an advantage of testing over restudying when the participants were tested after a week's delay. Karpicke and Roediger (2008) also found a sizeable long-term advantage of testing over restudying Swahili–English pairs, but in an interesting and unusual way. All participants went through alternating cycles during which to-be-learned pairs were either studied or tested, but depending on the condition a participant was in, individual pairs were dropped from further study cycles, dropped from subsequent test cycles, dropped from both, or not dropped from either. In each of the four conditions, every one of the 40 to-be-learned translations was gotten correct at least once by every participant, and the participants in each condition, when asked at the end of the study–test cycles, predicted that they would be able to recall about half of the translations when tested in a week. In fact, though, performance a week later was about 80% in the two conditions in which pairs remained in the test cycles but only about 35% when pairs were dropped from subsequent test cycles. Amazingly, once a translation was recalled correctly, that pair could be dropped from subsequent study cycles with impunity, so to speak, but dropping that Swahili word from subsequent test cycles decreased long-term recall drastically.

Errorless Versus Errorful Learning

Given the power of retrieval as a learning event, one legitimate concern about using testing to improve vocabulary learning is that when we do so the errors produced during testing will themselves be learned. Recent evidence suggests, however, that generating errors, especially errors made in attempting to predict or anticipate a to-be-remembered answer or association, can promote rather than impair learning.

That such generation attempts, even when they are assured of being incorrect, can improve subsequent memory for the correct response has drawn a great deal of recent interest (Grimaldi & Karpicke, 2012; Hays, Kornell, & Bjork, 2013; Huelser & Metcalfe, 2012; Knight, Ball, Brewer, DeWitt, & Marsh, 2012; Kornell, Hays, & Bjork, 2009; Potts & Shanks, 2014; Yan, Yu, Garcia, & Bjork, 2014). An experimental paradigm introduced by Kornell et al. (2009) has triggered much of this interest. The participants' task in this paradigm is to learn weakly associated paired associates, such as *whale–mammal*, for purposes of a final cued recall test (*whale:* ????). On some trials, such pairs are studied intact, whereas on other trials participants are asked to try to predict the to-be-remembered response given only the cue word. The basic result is that trying to predict the to-be-remembered target improves later cued recall of the target, even though the pairs are selected so that the participants' predictions are almost always wrong, and even though the time taken to try to anticipate the correct response is time taken away from the time to study the correct response (i.e., the total time on prediction trials is the same as the total time on trials when a pair is presented intact for study). Participants say that studying a pair intact helps them learn the pair better than does first trying to anticipate the to-be-learned response (Huelser & Metcalfe, 2012), however, a finding that is consistent with many other findings in demonstrating how easy it is to get fooled as to the conditions that do and do not optimize retention.

One interpretation of these findings is that the generation task leads to semantic activation of the cue word—for example, that trying to predict the to-be-remembered associate of *whale* activates the semantic network associated with *whale*, which then facilitates linking *mammal* to *whale*. With unrelated word pairs, for example, there is no benefit of generating a prediction error (Grimaldi & Karpicke, 2012; Huelser & Metcalfe, 2012). One might expect, then, that anticipatory error generation would not facilitate vocabulary learning, either of words one does not know in one's native language or of cross-language vocabulary learning. However, an interesting and provocative series of experiments by Potts and Shanks (2014) suggests that error generation might indeed facilitate such vocabulary learning.

In two experiments, Potts and Shanks (2014) had English-speaking participants learn words they did not know. In one experiment the words were English words but very unusual ones, such as *bistoury*, which means “knife,” or *gaberlunzie*, which means “beggar.” In the other experiment the words to be learned were words in Euskara, a Basque language unknown to the participants, words such as *hodei* (“cloud”) or *bidaia* (“journey”). During the learning phase, a to-be-learned word and its translation were shown intact for 13 s, or the to-be-learned word was shown by itself for 8 s before the correct translation was shown for 5 s. During the 8 s a word was shown without the translation, participants, depending on the

experimental condition, had to try to predict (guess at) the correct translation or choose from four alternatives. In the latter case, for example, if the word was *valinch*, the participant might have to choose from the alternatives “maid,” “lane,” “tube,” and “horn.” Whether on a given trial the translation was to be studied, generated (guessed at), or chosen from four alternatives, participants were ask to predict the likelihood they would be able to recall the correct translation on the final test, during which the participants also had to choose from four alternatives.

The results from the experiment that involved learning Euskara words are shown in Figure 1; the same pattern of results was found in the experiment that involved learning unusual English words. As is apparent in Figure 1, generating an erroneous prediction led to the best later recall of a given translation but was accompanied by participants predicting that it would lead to the worst later recall.

The results shown in Figure 1 are especially difficult to explain, given that no preexisting associations to the Euskara words existed in the participants' memories. According to Potts and Shanks (2014), the results suggest that there may well be “something about the active process of generating a response, rather than merely selecting one, which facilitated encoding of corrective feedback, even when the generated response was incorrect” (p. 17). Their proposal is that errorful generation contributes to a focusing of attention that is not as present in the study-only condition, but a more complete answer is likely to come from ongoing research. In the meantime, the results in Figure 1, together with the results of Potts and Shanks's experiment using unknown English words, adds to the evidence that difficulties and errors can contribute to vocabulary learning, whether in one's native language or in learning a new language.

Comparing the Two Approaches to Vocabulary Learning

As noted earlier, two different research traditions have examined vocabulary learning, one as a vehicle to reveal general properties of learning and memory and the other to examine the way in which the lexicon develops in conjunction with other language processes. Adult speakers know many thousands of words, and the first task for second-language (L2) learners is typically to acquire vocabulary in the new language. The approach described in the first part of this article provides a framework for understanding how initial conditions of study that are more difficult, induce errors, and require greater elaborative processing may benefit learning in the long term (Bjork et al., 2013). In contrast, language processing studies tend to catch the earliest moments of comprehension on the fly and to examine the way in which that initial understanding maps onto spoken production. But it is rare that language processing studies consider the enduring consequences of processing for later memory. Studies of vocabulary learning from a language processing perspective also tend to examine the performance of actual learners who are attempting to acquire a second language to some degree of proficiency. Where these two approaches come together is in training studies of word learning. In the next section we compare the evidence from these two approaches. A unique contribution Alice Healy has made to the field is to suggest that they may be providing converging evidence about the same underlying learning mechanisms. Here we illustrate how that may be the case.

The Role of Contextual Interference

Schneider et al. (2002) reported a vocabulary learning experiment in which native English speakers with no preexisting knowledge of French studied translation pairs from English to French or from French to English. When tested, the participants were required to type the response word. At the end of the Day 1 training, participants were tested in the practiced direction; that is, they had to produce the English translation if they had been given French cues during Day 1 practice, or they had to give the French translation if they had been given English cues during Day 1 practice. Not surprisingly, as shown in Figure 2, the French-to-English direction produced better performance on the test administered immediately after the Day 1 practice trials. From the perspective of paired-associate learning, the effect of study direction reveals the relative importance of encoding new information and attaching it to existing knowledge, compared with the relative difficulty of producing a novel word in the new language. That is, it was harder for participants to produce the novel French word when cued with the English translation than to recognize it and produce the translation in English when cued with the French word.

When, however, the participants returned a week later and were tested again, either in the direction they practiced a week earlier or in the opposite direction, Schneider et al. (2002) found benefits of having learned in the more difficult English-to-French direction. As shown in Figure 2, performance on the delayed test, plotted as a function of the translation direction during Day 1 training and averaged over the direction tested a week later, was better given the more difficult practice a week earlier. When performance on the delayed test was broken down by whether the test direction matched or mismatched the trained direction a week earlier, Schneider et al. found that participants who learned the vocabulary in the more difficult English-to-French direction not only performed better on the delayed test administered in that direction than did participants who learned the pairs in the easier French-to-English direction but also performed as well (slightly better, numerically) when the delayed test required producing the English translation of the French words. The pattern of results is consistent with an interpretation that the harder-to-learn condition, from English to French, imposed a desirable difficulty that resulted in long-term benefits to memory and retention.

Schneider et al. (2002) related the observed effects of cue direction to claims about the direction of translation in language processing. Kroll and Stewart (1994) reported a study in which highly proficient Dutch–English bilinguals translated words from one language to the other and also named words in each language. They reported a translation asymmetry, with shorter translation latencies when bilinguals translated into the first language, the L1, than into the second language, the L2. The result resembles the cuing effect in accuracy reported by Schneider et al.

Critically, Kroll and Stewart (1994) demonstrated that only translation into the L2, the forward direction of translation, was influenced by the semantic composition of the list of words to be translated. Translation from the L2 into the L1, the backward direction of translation, appeared to be immune from the influence of the meaning of the words that were

translated (Figure 3). When bilinguals named words in each language, simply speaking the name of the word without translating, they were slower to name words in the L2 than in L1.

Despite the greater difficulty in speaking the L2, the word naming data showed no differential effects of the semantics, suggesting that the effect observed in translation reflected the mapping between word forms, not the greater difficulty per se in producing the L2. The differential semantic effect in translation may be a processing analog to the sort of contextual interference described earlier in regard to memory experiments that have shown that interleaving, a more difficult study condition than blocking, produces benefits to learning (for reviews, see Bjork et al., 2013; Lee, 2012). Kroll and Stewart (1994) demonstrated that translation from L1 to L2 was slower when the words to be translated were semantically blocked than randomly mixed. The same result has been reported in other speeded processing tasks in which a single word is required for production in contexts in which semantic blocking increases the competition between the lexical candidates from which a selection must be made (Belke, Meyer, & Damian, 2005). Importantly, an incidental recall task at the end of the translation experiment in Kroll and Stewart revealed a significant effect of semantic blocking for recall in the difficult forward direction of translation (from L1 to L2) but no effect in the easier backward direction of translation (from L2 to L1). In effect, recall was best for the condition that produced the greatest processing costs in translation.

Like Kroll and Stewart (1994), Schneider et al. (2002) also included a manipulation of semantic blocking to determine whether contextual interference might produce a desirable difficulty in learning. Unlike Kroll and Stewart, they did not find more of an effect of blocking in the more difficult cuing condition but instead in the easier condition, where the cue in French required a response in English. At later relearning, that difference was absent. Schneider et al. noted that the difference in the results for the two experiments might have been attributable to the fact that the participants in the Kroll and Stewart study were real bilinguals who actively used the two languages, Dutch and English, and for whom the blocking manipulation represented a meaningful context. In Schneider et al.'s study, the French words were used only as a context in which to train new vocabulary, but the participants were neither actual language learners nor bilingual. Differences in design and in the timing of study and test in the two studies may also account for the observed effects and the small contribution of semantic blocking in the Schneider et al. study.

Also, viewed more broadly from a contextual interference standpoint, blocking by semantic category may involve a fundamental tradeoff: On one hand, the pool of possible responses is reduced by blocking; on the other hand, correct responses must be selected from among more potent competitors. The relative roles of those two factors may well have differed in the two studies. Critically, for present purposes there were effects of language direction in both studies, with the more difficult conditions producing better performance in a later test.

Accounting for Asymmetries in Bilingual Translation

Kroll and Stewart (1994) proposed the revised hierarchical model (RHM) to account for the asymmetries obtained in bilingual translation. The model, shown in Figure 4, assumes that

when adults initially learn new L2 words, they attach those words to the language system via associative links to their respective L1 translations. Because adult learners have an existing L1 lexicon and know the meanings of the L1 words, they can exploit the mappings between words and concepts for the L1 in learning the new L2 words. Therefore, the L2-to-L1 direction of translation is hypothesized to be lexically mediated. The process of lexical mediation is another manifestation of the more general principle of transfer of new information to old information that is seen not only in learning experiments in the lab but also in studies of actual second-language learning for all levels of language processing, including lexicon, grammar, and phonology (MacWhinney, 2005). When words are presented in the L1 to be translated into the L2, the semantics will be rapidly activated. However, the route from meaning to the L2 word, the process that is called lexicalization in models of language production, is hypothesized to be slow and error prone because there will be some concepts that are not known in the L2 and because those that are known will be only weakly associated to their respective meanings until the speaker is highly proficient in both languages. In this account, only translation from L1 to L2, in the forward direction, will be influenced by a semantic manipulation such as the category of words to be translated, whereas lexical mediation, from the L2 to L1, in the backward direction, can be achieved without semantic access.

The predictions of the RHM have been tested and debated, particularly with respect to the role of lexical mediation (for a recent review, see Kroll, Van Hell, Tokowicz, & Green, 2010). Critically, for present purposes the model provides an account of language processing that converges closely with the associative learning explanation for the Schneider et al. (2002) results. In each case, the condition that was more difficult for learners and for proficient bilinguals, requiring production of the weaker new language or L2 word, produced greater benefits to later memory and more evidence of conceptual processing than the condition that was easier, requiring only recognition of the L2 word and production in L1. In effect, production into the L2 created a desirable difficulty with respect to learning and memory.

Desirable difficulties can reflect the imposition of encoding strategies that require longer processing times, greater conceptual elaboration, or an increased presence of erroneous responses during learning. But they can also arise from strategies that are imposed by learners themselves, as a function of their experience, what Bjork et al. (2013) call self-regulated learning. In the next section we describe another experiment on vocabulary learning that we believe provides a second illustration of how the mechanisms that account for learning and memory may converge with evidence from language processing.

Domain-General Effects of Bilingualism

The recent literature is filled with accounts of how bilingual language experience may produce consequences for domain-general cognition (see Bialystok, Craik, & Luk, 2012, for a review). The evidence on language processing suggests that the bilingual's two languages are always active and competing. Most bilinguals are more proficient in one language than the other, typically the native language. The claim is that to become a proficient bilingual, it is necessary to learn to regulate the control of the two languages so that the weaker of the

two languages can be used without intrusion from the stronger language (see Kroll, Bobb, and Hoshino, 2014, for a recent review of the evidence on cross-language activation).

Of key relevance to the present issues, the constant requirement for bilinguals to control the two languages has been hypothesized to produce a range of consequences to cognition that extend beyond language use. Much of the recent work on these consequences of bilingualism is focused on executive function and on the ways in which the brain networks that support executive function are tuned in response to the ways in which the two languages are used (Green & Abutalebi, 2013; Kroll & Bialystok, 2013). We return to this issue at the end of the article to consider how multiple language use and bilingualism itself may produce desirable difficulties. However, it is important to note that very little of this research has addressed the issue of new learning or the ways in which the use of a second language may change the sort of regulatory strategies people bring to new learning contexts. The few experiments that have investigated the consequences of bilingualism for vocabulary learning have produced results that are largely positive, with evidence that bilinguals are better new language learners than are monolinguals when confronted with new foreign language vocabulary (Kaushanskaya & Marian, 2009). What is unclear is whether the reported advantage is simply another reflection of the more general consequences of bilingualism for cognition or whether bilingualism produces specific consequences for learning that reflect the life experience bilinguals have in regulating the use of the two languages.

Bogulski and Kroll (in preparation) conducted a study that provides evidence for the role of learner-imposed strategies that may be responsible for the observed bilingual advantages in word learning. The study asked whether the bilingual word learning advantage depended on learning the new words via the native or dominant L1. If bilingualism confers advantages to executive function that extend to new learning, then bilinguals might be expected to reveal learning advantages that are broad in scope and not dependent on the specific way that they themselves learned the L2. However, if the new learning engages the mechanisms that were active during initial learning of the L2 or that are used to enable regulation of the two languages, then the advantage might be specifically tied to that aspect of language experience.

Bogulski and Kroll (in preparation) speculated that learning via the L1 might underlie the bilingual word learning advantage because there is abundant evidence that bilinguals exercise inhibitory control to enable them to use the weaker of the two languages. Studies of language comprehension and language production show that bilinguals appear to inhibit alternatives in the more dominant L1 to enable them to process information in the less dominant L2 (Green, 1998; Martín, Macizo, & Bajo, 2010; Meuter & Allport, 1999; Misra, Guo, Bobb, & Kroll, 2012). Critically, it is the L1 that is inhibited. The weaker L2 does not require inhibition to the same degree. The hypothesis that Bogulski and Kroll tested is that bilinguals will be advantaged when learning new foreign language vocabulary but only when learning via the L1, the language with which they have regulatory experience. Of interest is the fact that almost all past studies of the bilingual advantage in word learning have examined learning via the L1, so it is impossible from the past literature to determine whether the advantage is a general consequence of bilingualism or a more specific effect reflecting the fact that bilinguals learn to regulate their L1 to enable proficient use of the L2.

Bogulski and Kroll (in preparation) compared the performance of three bilingual groups. One group was native English speakers highly proficient in Spanish as the L2 (English–Spanish bilinguals), another was native Spanish speakers highly proficiency in English as the L2 (Spanish–English bilinguals), and a third group was native Chinese speakers highly proficient in English as the L2 (Chinese–English bilinguals). The three bilingual groups were compared with monolingual speakers of English on a set of word learning tasks using Dutch as the foreign language that was equally unfamiliar for monolinguals and bilinguals alike. At study, all participants were shown a Dutch word followed by its English translation. The task was to name the word in English as soon as they judged that they had studied the word adequately. The task was not speeded, but they were required to name the English translation within 5 s. They were then tested on a translation recognition task in which they had to judge whether a Dutch and English word were translation equivalents and then returned to the lab for a separate testing session in which they restudied the Dutch words and then performed a Dutch lexical decision task in which they judged whether a letter string was a word in Dutch. Across the experiment, all participants studied the words three times, twice in an initial session and once more in a second session.

Bogulski and Kroll (in preparation) found that only the English–Spanish bilinguals who had studied the new Dutch words via English, their L1, were advantaged in the later test of lexical decision relative to monolingual participants. Neither the native Spanish or native Chinese speakers who studied the new words via English, their L2, produced an advantage. Critically, the performance at study revealed an unexpected difference across the groups of learners. The English–Spanish bilinguals who later revealed the word learning advantage appeared to have adopted a strategy during initial study of the new words that was slow and strategic. These bilinguals were slower than the monolingual learners by hundreds of milliseconds, although both groups were native English speakers and closely matched on other dimensions. These data are shown in Figure 5. One hypothesis may simply be that this group of English–Spanish bilinguals was particularly slow. However, a comparison of the same groups on a picture naming task, used for the purpose of assessing English proficiency, revealed identical naming latencies, suggesting that the English–Spanish bilinguals were not slow in processing overall but only selectively slow when learning new words. The hypothesis is that the bilingual learners have experience in learning how to inhibit their L1 even when required to use the L1 to respond. Although English is the L1 for the monolingual speakers, they presumably have little experience in having to inhibit the L1 because it is their only functional language. Likewise, the two other bilingual groups were producing English at study as their L2 and have little reason to inhibit English. It appears that in the absence of this inhibitory pattern, there is no bilingual advantage in word learning, suggesting that the effect is specific to the conditions of learning rather than a more general cognitive consequence of bilingualism. In the context of the present discussion, these data can be interpreted as revealing a desirable difficulty but one imposed by the nature of self-regulation rather than by the conditions of study.

Does Bilingualism Impose Desirable Difficulties?

In the media, there has been widespread recent coverage attesting to the benefits of multiple language use for the mind and the brain. A particularly provocative claim is that a lifetime of

bilingualism protects aging brains from both normal and pathological declines associated with cognitive aging and with disease (Bak, Nissan, Allerhand, & Deary, 2014; Bialystok, Craik, Green, & Gollan, 2009). Perhaps the most dramatic finding is that bilinguals diagnosed with Alzheimer's-type dementia appear to present with symptoms 4 to 5 years later than their monolingual counterparts (Bialystok, Craik, & Freedman, 2007). The hypothesis is that bilingualism gives the brain a workout, tuning the brain networks that are responsible for cognitive control and conflict resolution (Abutalebi et al., 2012; Gold, Kim, Johnson, Kriscio, & Smith, 2013). Experience in resolving the competition that normally occurs in regulating the use of two languages to enable fluent performance in each is thought to provide compensatory protection in the presence of aging or disease. In the case of Alzheimer's, there is striking evidence that when bilinguals present with symptoms, not only are they older than monolinguals, but their brains are also more diseased than monolinguals presenting with the same symptoms (Schweizer, Ware, Fischer, Craik, & Bialystok, 2012), suggesting a longer compensatory period during which the bilinguals were apparently able to cope with their symptoms.

How do these remarkable benefits to the mind and brain arise from a life of bilingual language experience? The general account is the one mentioned earlier. Studies of bilingual language processing show that when bilinguals listen to speech, read, and plan speech in each of their two languages, the language not in use is active and competing for selection. The bilingual has been described as a mental juggler, constantly negotiating competing demands across the two languages (Kroll, Dussias, Bogulski, & Valdes-Kroff, 2012). What is notable is that the observed effects of cross-language activation and their competitive consequences appear to be present at every level of language processing, from the words that are spoken to the grammar that is selected. Yet bilinguals rarely make errors of language, suggesting that the regulatory mechanisms to which we have alluded provide an elegant means of cognitive control. In this sense, the bilingual experience is one of negotiating a set of difficulties that may not be desirable in any sort of obvious way at the point of learning or using language but that provide deep benefits to cognition across the life span.

In reviews of the cognitive consequences of bilingualism (Bialystok et al., 2009), consideration is given to outcomes that appear to be positive and beneficial, such as more efficient resolution of cognitive conflict (Abutalebi et al., 2012). But not all consequences of bilingualism are positive, and in the context of discussing desirable difficulties it is useful to consider whether the documented costs associated with multiple language use might really be desirable difficulties in disguise. Bilinguals are often slower to speak, even in their native language, than monolingual speakers of the same native language (Gollan, Montoya, Cera, & Sandoval, 2008). They also produce fewer exemplars in a verbal fluency task than monolinguals (Sandoval, Gollan, Ferreira, & Salmon, 2010) and have a larger number of tip-of-the-tongue experiences than monolinguals (Gollan & Brown, 2006). Bilingual children also have smaller vocabularies than their monolingual peers (Bialystok et al., 2009).

The evidence for a bilingual deficit in verbal processing has been interpreted in two different ways in the past literature. Gollan and colleagues (2008) have attributed the bilingual disadvantage to the idea that bilinguals have fewer opportunities to use each language than monolinguals. If all humans have only 24 hours each day but some use more than one

language, then each of the languages will have lower frequency relative to monolingual speakers. From this perspective, bilinguals will always be running a deficit, even in their native language. The alternative is that the two languages are always competing and that there are costs engaged by the mechanisms needed to resolve that competition (see Kroll & Gollan, 2014, for a detailed comparison of these two accounts). The competition-for-selection account has the advantage of providing a basis on which costs may translate into benefits. Each opportunity to resolve cross-language competition may extract costs that draw on executive function and working memory resources, but the need to engage in the process of conflict resolution may itself confer benefits to other processes. Bilinguals may learn something more general about resolving competition during learning and problem solving that monolinguals simply do not face to the same degree. Recent studies of both behavior (Blumenfeld & Marian, 2011) and brain function (Abutalebi et al., 2012; Gold et al., 2013) show that bilinguals are more efficient in resolving conflict than monolinguals. They need less brain activation to get the same job done, and they are faster to inhibit distracting alternatives. Although a great deal of research remains to be performed to determine precisely how problems in language processing map onto their respective consequences, it is appealing to think that the bilingual disadvantages that have been mentioned are the other side of the same coin. Only when we challenge language processing do the subsequent benefits to cognition later appear. Those benefits may be subtle in young adulthood, but they appear to be robust as people age and under conditions of reduced cognitive resources.

Beyond Vocabulary Learning: Some Concluding Comments

We end by paying tribute to Alice Healy, whose work has shown us from the start that we need to attend to all these processes: those that can be examined on the fly as they unfold in real time and those that extend over longer periods of time and perhaps over the entire life span. The research on vocabulary learning that Alice and her students first used as a laboratory paradigm to investigate learning and memory has created a set of deep questions that are broad in scope and suggest an exciting new research agenda. At the heart of that agenda is the recognition that learning, memory, and language are part of an integrated network. Identifying desirable difficulties in learning contexts and in learners themselves provides a rich approach to investigating that network.

Acknowledgments

This article reflects our admiration for Alice Healy and discussions we had over the last year when last year when J.K. was a sabbatical visitor at UCLA and attended the Cogfog lab meetings. J.K. thanks Bob and Elizabeth Bjork and the Cogfog group for stimulating new insights and research directions in memory and language; R.B. thanks J.K. for opening his eyes to the potential research on language and bilingualism has to address broadly important questions about the functional architecture of human memory. The writing of this article was supported in part by grant 29192G from the McDonnell Foundation to E. L. and R. A. Bjork and by National Institutes of Health grant HD053146 and National Science Foundation grants BCS-0955090 and OISE-0968369 to J. F. Kroll.

REFERENCES

Abutalebi J, Della Rosa PA, Green DW, Hernandez M, Scifo P, Keim R, Costa A. Bilingualism tunes the anterior cingulate cortex for conflict monitoring. Cerebral Cortex. 2012; 22:2076–2086. [PubMed: 22038906]

Anderson MC, Bjork RA, Bjork EL. Remembering can cause forgetting: Retrieval dynamics in long-term memory. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1994; 20:1063–1087.

Bak TH, Nissan JJ, Allerhand MM, Deary IJ. Does bilingualism influence cognitive aging? Annals of Neurology. 2014; 75(6):959–963. [PubMed: 24890334]

Battig, WF. The flexibility of human memory. In: Cermak, LS.; Craik, FIM., editors. Levels of processing and human memory. Erlbaum; Hillsdale, NJ: 1978.

Belke E, Meyer AS, Damian MF. Refractory effects in picture naming as assessed in a semantic blocking paradigm. Quarterly Journal of Experimental Psychology. 2005; 58:667–692. [PubMed: 16104101]

Bialystok E, Craik FIM, Freedman M. Bilingualism as a protection against the onset of symptoms of dementia. Neuropsychologia. 2007; 45:459–464. [PubMed: 17125807]

Bialystok E, Craik FIM, Green DW, Gollan TH. Bilingual minds. Psychological Science in the Public Interest. 2009; 10:89–129. [PubMed: 26168404]

Bialystok E, Craik FI, Luk G. Bilingualism: Consequences for mind and brain. Trends in Cognitive Science. 2012; 16:240–250.

Bjork, RA. Retrieval as a memory modifier. In: Solso, R., editor. Information processing and cognition: The Loyola Symposium. Erlbaum; Hillsdale, NJ: 1975. p. 123-144.

Bjork, RA. Memory and metamemory considerations in the training of human beings. In: Metcalfe, J.; Shimamura, A., editors. Metacognition: Knowing about knowing. MIT Press; Cambridge, MA: 1994. p. 185-205.

Bjork RA, Dunlosky J, Kornell N. Self-regulated learning: Beliefs, techniques, and illusions. Annual Review of Psychology. 2013; 64:417–444.

Blumenfeld HK, Marian V. Bilingualism influences inhibitory control in auditory comprehension. Cognition. 2011; 118:245–257. [PubMed: 21159332]

Bogulski, CA.; Kroll, JF. A bilingual advantage in vocabulary acquisition depends on learning via the dominant language. Pennsylvania State University; in preparationUnpublished manuscript

Gold BT, Kim C, Johnson NF, Kriscio RJ, Smith CD. Lifelong bilingualism maintains neural efficiency for cognitive control in aging. Journal of Neuroscience. 2013; 33:387–396. [PubMed: 23303919]

Gollan TH, Brown AS. From tip-of-the-tongue (TOT) data to theoretical implications in two steps: When more TOTs means better retrieval. Journal of Experimental Psychology: General. 2006; 135:462–483. [PubMed: 16846276]

Gollan TH, Montoya RI, Cera C, Sandoval TC. More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. Journal of Memory and Language. 2008; 58:787–814. [PubMed: 19343088]

Green D. Mental control of the bilingual lexico-semantic system. Bilingualism: Language and Cognition. 1998; 1:67–81.

Green DW, Abutalebi J. Language control in bilinguals: The adaptive control hypothesis. Journal of Cognitive Psychology. 2013; 25:515–530. [PubMed: 25077013]

Grimaldi PJ, Karpicke JD. When and why do retrieval attempts enhance subsequent encoding? Memory & Cognition. 2012; 40:505–513. [PubMed: 22238214]

Hays MJ, Kornell N, Bjork RA. When and why a failed test potentiates the effectiveness of subsequent study. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2013; 39:290–296. doi:10.1037/a0028468.

Healy, AF.; Bourne, LE., Jr., editors. Foreign language learning: Psycholinguistic studies on training and retention. Psychology Press; Hove, UK: 1998.

Healy AF, Bourne LE Jr. Empirically valid principles for training in the real world. American Journal of Psychology. 2013; 126:389–399. [PubMed: 24455807]

Hogan RM, Kintsch W. Differential effects of study and test trials on long-term recognition and recall. Journal of Verbal Learning and Verbal Behavior. 1971; 10:562–567.

Huelser BJ, Metcalfe J. Making related errors facilitates learning, but learners do not know it. Memory & Cognition. 2012; 40:514–527. [PubMed: 22161209]

Karpicke JD, Roediger HL. The critical importance of retrieval for learning. Science. 2008; 319:966–968. [PubMed: 18276894]

Kaushanskaya M, Marian V. Bilingualism reduces native-language interference during novel-word learning. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2009; 35:829–835.

Knight JB, Ball BH, Brewer GA, DeWitt MR, Marsh RL. Testing unsuccessfully: A specification of the underlying mechanisms supporting its influence on retention. Journal of Memory and Language. 2012; 66:731–746.

Kornell N, Hays MJ, Bjork RA. Unsuccessful retrieval attempts enhance subsequent learning. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2009; 35:989–998.

Kroll JF, Bialystok E. Understanding the consequences of bilingualism for language processing and cognition. Journal of Cognitive Psychology. 2013; 25:497–514.

Kroll JF, Bobb SC, Hoshino N. Two languages in mind: Bilingualism as a tool to investigate language, cognition, and the brain. Current Directions in Psychological Science. 2014; 23:159–163. [PubMed: 25309055]

Kroll, JF.; Dussias, PE.; Bogulski, CA.; Valdes-Kroff, J. Juggling two languages in one mind: What bilinguals tell us about language processing and its consequences for cognition. In: Ross, B., editor. The psychology of learning and motivation. Vol. Vol. 56. Academic Press; San Diego, CA: 2012. p. 229-262.

Kroll, JF.; Gollan, TH. Speech planning in two languages: What bilinguals tell us about language production. In: Ferreira, V.; Goldrick, M.; Miozzo, M., editors. The Oxford handbook of language production. Oxford University Press; Oxford, England: 2014. p. 165-181.

Kroll JF, Gullifer J, Rossi E. The multilingual lexicon: The cognitive and neural basis of lexical comprehension and production in two languages. Annual Review of Applied Linguistics. 2013; 33:102–127.

Kroll JF, Stewart E. Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. Journal of Memory and Language. 1994; 33:149–174.

Kroll JF, Van Hell JG, Tokowicz N, Green DW. The revised hierarchical model: A critical review and assessment. Bilingualism: Language and Cognition. 2010; 13:373–381.

Lee, TD. Contextual interference: Generalization and limitations. In: Hodges, NJ.; Williams, AM., editors. Skill acquisition in sport: Research, theory, and practice II. Routledge; London, England: 2012. p. 79-93.

MacWhinney, B. A unified model of language acquisition. In: Kroll, JF.; De Groot, AMB., editors. Handbook of bilingualism: Psycholinguistic approaches. Oxford University Press; New York, NY: 2005. p. 49-67.

Martín MC, Macizo P, Bajo T. Time course of inhibitory processes in bilingual language processing. British Journal of Psychology. 2010; 101:679–693. [PubMed: 20184787]

Meuter RFI, Allport A. Bilingual language switching in naming: Asymmetrical costs of language selection. Journal of Memory and Language. 1999; 40:25–40.

Misra M, Guo T, Bobb SC, Kroll JF. When bilinguals choose a single word to speak: Electrophysiological evidence for inhibition of the native language. Journal of Memory and Language. 2012; 67:224–237.

Potts R, Shanks DR. The benefit of generating errors during learning. Journal of Experimental Psychology: General. 2014; 143:644. [PubMed: 23815457]

Roediger HL, Karpicke JD. The power of testing memory: Basic research and implications for educational practice. Perspectives on Psychological Science. 2006a; 1:181–210. [PubMed: 26151629]

Roediger HL, Karpicke JD. Test-enhanced learning: Taking memory tests improves long-term retention. Psychological Science. 2006b; 17:249–255. [PubMed: 16507066]

Sandoval TC, Gollan TH, Ferreira VS, Salmon DP. What causes the bilingual disadvantage in verbal fluency? The dual-task analogy. Bilingualism: Language and Cognition. 2010; 13:231–252.

Schneider VI, Healy AF, Bourne LE Jr. What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention, and transfer. Journal of Memory and Language. 2002; 46:419–440.

Schweizer TA, Ware J, Fischer CE, Craik FI, Bialystok E. Bilingualism as a contributor to cognitive reserve: Evidence from brain atrophy in Alzheimer's disease. Cortex. 2012; 48:991–996. [PubMed: 21596373]

Storm BC, Friedman MC, Murayama K, Bjork RA. On the transfer of prior tests or study events to subsequent study. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2014; 40:115–124.

Thompson CP, Wenger SK, Bartling CA. How recall facilitates subsequent recall. Journal of Experimental Psychology. 1978; 4:210–221. [PubMed: 660096]

Wheeler MA, Ewers M, Buonanno JF. Different rates of forgetting following study versus test trials. Memory. 2003; 11:571–580. [PubMed: 14982124]

Yan VX, Yu Y, Garcia MA, Bjork RA. Why does trying, and failing, to predict to-be-learned responses enhance later recall of those responses? Memory & Cognition. 2014; 42:1373–1383. [PubMed: 25120240]

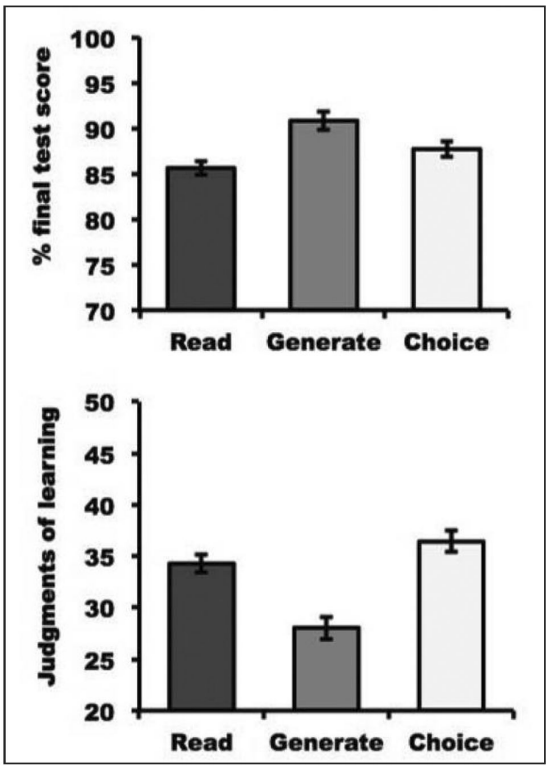

FIGURE 1.

Actual (top panel) and predicted (bottom panel) final test performance as a function of whether the English translation of a Euskara word was studied, was predicted before being studied, or was chosen from 4 alternatives before being studied (data from Potts & Shanks, 2014)

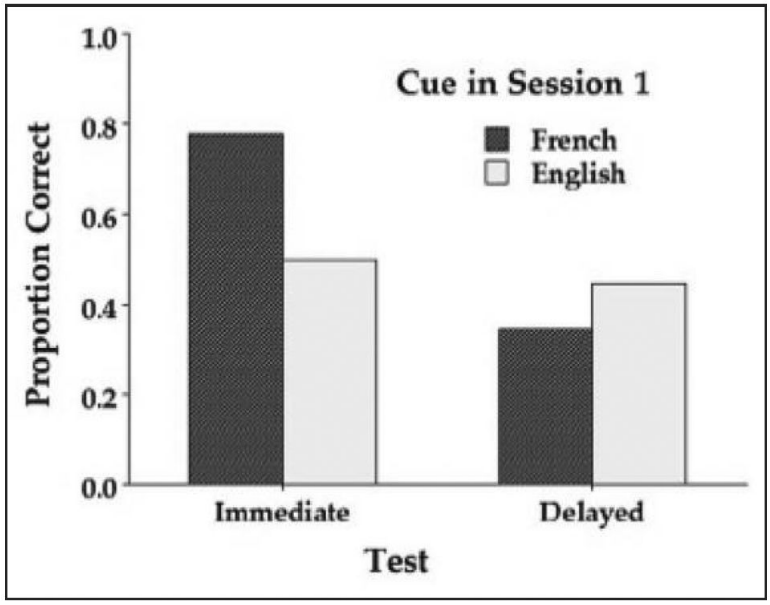

FIGURE 2.

Immediate and 1-week delayed performance as a function of translation direction (English to French or French to English) during Session 1. On the immediate test, participants were tested in the direction practiced during training. On the delayed test half the participants were tested in the practiced direction and half were tested in the other direction, and the results shown for the delayed test average over that manipulation (data from Schneider, Healy, & Bourne, 2002)

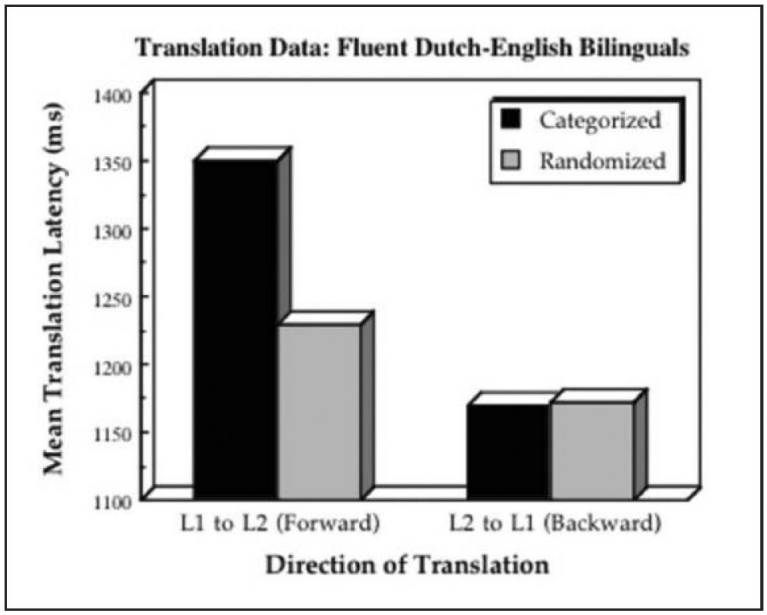

FIGURE 3.

Mean translation latencies for Dutch–English bilinguals to translate words in each direction of translation when word lists were semantically categorized or randomly mixed (data from Kroll & Stewart, 1994)

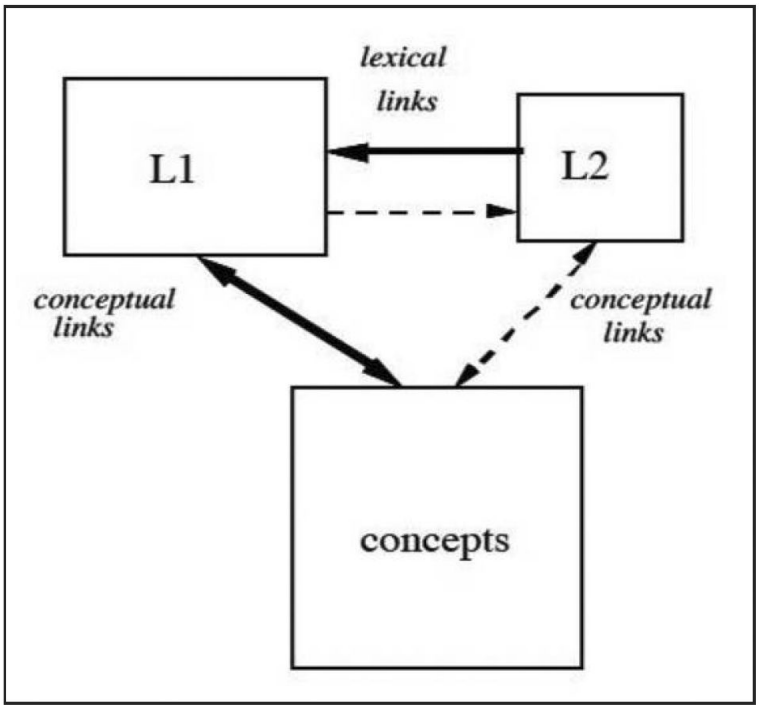

FIGURE 4.
The revised hierarchical model (adapted from Kroll & Stewart, 1994)

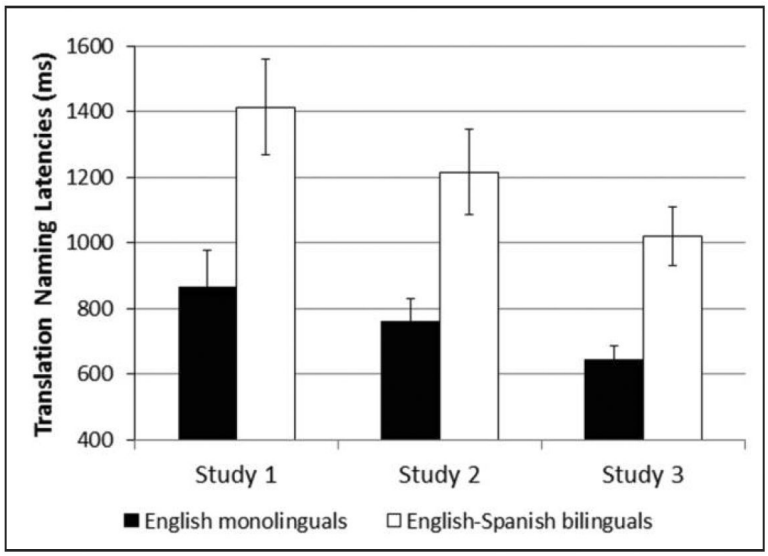

FIGURE 5.

Latencies to name the English translations of Dutch words for monolingual English speakers and English–Spanish bilinguals at 3 times of study (data from Bogulski & Kroll, in preparation).