Bilingual Minds

Ellen Bialystok¹, Fergus I.M. Craik², David W. Green³, and Tamar H. Gollan⁴

¹Department of Psychology, York University, ²Rotman Research Institute, Baycrest Centre, ³Department of Cognitive, Perceptual, and Brain Sciences, University College London, ⁴University of California, San Diego

Summary

The regular use of two languages by bilingual individuals has been shown to have a broad impact on language and cognitive functioning. In this monograph, we consider four aspects of this influence.

In the first section, we examine differences between monolinguals and bilinguals in children’s acquisition of language and adults’ linguistic processing, particularly in terms of lexical retrieval. Children learning two languages from birth follow the same milestones for language acquisition as monolinguals do (first words, first use of grammar) but may use different strategies for language acquisition, and they generally have a smaller vocabulary in each language than do monolingual children learning only a single language. Adult bilinguals typically take longer to retrieve individual words than monolinguals do, and they generate fewer words when asked to satisfy a constraint such as category membership or initial letter.

In the second section, we consider the impact of bilingualism on nonverbal cognitive processing in both children and adults. The primary effect in this case is the enhancement of executive control functions in bilinguals. On tasks that require inhibition of distracting information, switching between tasks, or holding information in mind while performing a task, bilinguals of all ages outperform comparable monolinguals. A plausible reason is that bilinguals recruit control processes to manage their ongoing linguistic performance and that these control processes become enhanced for other unrelated aspects of cognitive processing. Preliminary evidence also suggests that the executive control advantage may even mitigate cognitive decline in older age and contribute to cognitive reserve, which in turn may postpone Alzheimer’s disease.

In the third section, we describe the brain networks that are responsible for language processing in bilinguals and demonstrate their involvement in nonverbal executive control for bilinguals. We begin by reviewing neuroimaging research that identifies the networks used for various nonverbal executive control tasks in the literature. These networks are used as a reference point to interpret the way in which bilinguals perform both verbal and nonverbal control tasks. The results show that bilinguals manage attention to their two language systems using the same networks that are used by monolinguals performing nonverbal tasks.

In the fourth section, we discuss the special circumstances that surround the referral of bilingual children (e.g., language delays) and adults (e.g., stroke) for clinical intervention. These referrals are typically based on standardized assessments that use normative data from monolingual populations, such as vocabulary size and lexical retrieval. As we have seen, however, these measures are often different for bilinguals, both for children and adults. We discuss the implications of these linguistic differences for standardized test performance and clinical approaches.

We conclude by considering some questions that have important public policy implications. What are the pros and cons of French or Spanish immersion educational programs, for example? Also, if bilingualism confers advantages in certain respects, how about three languages—do the benefits increase? In the healthcare field, how can current knowledge help in the treatment of bilingual aphasia patients following stroke? Given the recent increase in bilingualism as a research topic, answers to these and other related questions should be available in the near future.

Introduction

As the world becomes more interconnected, it is increasingly apparent that bilingualism is the rule and not the exception. Not only do some countries support bilingual populations because of cultural and linguistic diversity within its citizenry, but also increased global mobility has enlarged the number of people who have become bilingual at all levels of society. For example, a recent survey of language use in the United States obtained from the American Community Survey in 2007 reported that approximately 20% of the population spoke a non-English language at home, a proportion that has increased by 140% since 1980 (Shin & Kominski, 2010). These numbers are higher when considering world figures: Crystal (1997) estimates bilingualism that includes English and another language.

Corresponding Author:
Ellen Bialystok, Department of Psychology, York University, 4700 Keele Street, Toronto, Ontario M3J 1P3, Canada
E-mail: ellenb@yorku.ca
represents about 235 million people worldwide and that two thirds of the children in the world are raised in bilingual environments.

Recently, evidence indicating that this common experience has a systematic and significant impact on cognitive functioning has accumulated. In this review, we examine the nature of that impact across the lifespan and consider what these effects contribute to our understanding of cognition in general. We begin by examining the linguistic dimensions of bilingualism in terms of children’s language acquisition and adult language processing. In the second section, we investigate the consequences of bilingualism on nonverbal cognitive functioning. The third section describes research documenting how the brain supports bilingual functioning and how it changes in response to it. In the fourth section, we review the clinical implications of bilingualism for diagnosis and intervention. We conclude by identifying and discussing some specific issues for bilinguals in society. By adopting this cognitive perspective, there are a number of topics we do not cover, such as reading, lexical and syntactic processing, and linguistic consequences of brain damage, all of which are beyond the scope of the present review.

There are many ways to be bilingual: Some people are born bilingual, some aspire to bilingualism, and others have bilingualism thrust upon them later in life. Underlying these differences, a myriad of factors make the bilingual experience deeply heterogeneous and potentially alter its consequences. Some of the reasons for bilingualism include immigration, a family that speaks a heritage language, formal education in another language, temporary residence in another country, or a national situation in which the official language is different from the community language. Each of these circumstances is associated with a different set of social, cognitive, and personal factors, and these factors undoubtedly intervene in as well as determine any potential effect of bilingualism. Each of the situations associated with multiple language use also carries different assumptions about expectations for education, values around literacy, standards for language proficiency, the purposes for which one or both of the languages are used, the level of community support for the home language, and the identity of the individual as a member of a majority or minority culture. Therefore, there can be no single outcome and no definitive consequence that follows from incorporating more than one language into daily life. And yet the consequences of bilingualism affect educational policy, social organization, and conceptions of mind.

I. Language Learning and Language Use in Bilinguals

Language acquisition in bilingual children

The most striking feature of a young child’s acquisition of language is the extraordinary ease with which the process appears to progress. Perhaps more remarkable than this achievement, therefore, is that this facility for learning a complex symbolic system is not diminished when the child faces the task of learning two of them. Bilingual language acquisition is as effortless, efficient, and successful as monolingual acquisition. It is now clear that language acquisition is not a simple matter of biological unfolding, as some had previously believed, but rather a process that is finely tuned to features of the environmental input, the child’s attentional and perceptual abilities, and the development of cognitive and conceptual competencies. All of these factors conspire as well to shape the process of acquiring two languages. Moreover, as we describe later, the major milestones concerning competence in sounds, words, and sentences that are the foundation of acquiring language are passed at equivalent times for children growing up with one language in the home and those growing up in a multilingual home.

The acquisition of the phonological system by infants has been well documented for the case of monolingual acquisition: Infants can detect the contrasts that define the phonological system for all human languages almost from birth (e.g., /pa/ vs. /ba/; Eimas, Siqueland, Jusczyk, & Vigorito, 1971), but their ability to perceive these contrasts in languages that are not heard in the environment (e.g., /r/ vs. /l/ for children being raised in Japanese homes) begins to decline at about 6 months of age (Werker & Tees, 1984; see also Kuhl et al., 2006). Thus, until about 6 months old, there is no detectable difference in the perception of phonetic contrasts by infants in monolingual and bilingual environments but diverging patterns appear as bilingual babies maintain and develop the categorical distinctions for the phonetic system in both languages and monolingual infants lose the ability to detect contrasts that are not part of the language they are about to learn (Borns, Yoshida, Hill, & Werker, 2007; Sebastian-Galles & Bosch, 2005). By about 14 months old, infants being raised in bilingual environments have established a clearly demarcated phonological representation for both languages. Therefore, bilingual infants develop the phonological basis for both languages on roughly the same schedule as monolingual children do for their only language. It may be that it is this very early experience that leaves its lifelong trace as a foreign accent when childhood monolinguals attempt to learn new languages later in life.

Beyond the phonetic constituents, infants also need to learn the more general phonological structure of language. Recently, Kovacs and Mehler (2009a) presented auditory stimuli to 12-month-old infants who were being raised in a monolingual or bilingual environment. The stimuli were three-syllable combinations that had the syllabic structure of either ABA or AAB. These stimuli were artificial creations and were not words in any language. The crucial manipulation was that each structure was associated with a different response—namely, look either to the right or to the left to see an interesting toy. The experimental results showed that the monolingual babies could learn only one of the responses but that the bilingual babies learned both, a difference the researchers interpreted as demonstrating more flexible learning in bilinguals. They offer their results as part of the explanation for how bilingual children can learn twice as much about language as monolingual children in the same amount of time (although it is not clear that they do, as will be discussed below), but the task was
only marginally linguistic. If anything, it is more similar to word learning than to speech perception, a process that rests on different perceptual and cognitive processes than phonological development (Burns et al., 2007). In fact, bilingual babies apply their developing phonological system to the learning of new words later than monolingual children do (Fennell, Byers-Heinlein, & Werker, 2007), although a recent study testing 17-month-old infants raised with French and English did not replicate this finding and attributed the difference between studies to details of the phonetic input (Mattock, Polka, Rvachew, & Krehm, 2010). Nonetheless, the results reported by Kovacs and Mehler provide compelling evidence for different levels of performance in a phonological task in the first year of life that can be traced to the experience of building up two linguistic systems.

Undoubtedly the most salient evidence for children’s progress in language acquisition is word learning, particularly the appearance of the child’s first word. As with the developing phonological system, the basic milestones associated with this achievement are similar for children learning one or more languages. The child’s first word appears on average at about 1 year old, regardless of how many languages are in the environment (Pearson, Fernandez, & Oller, 1993) and, more dramatically, regardless of whether the languages are both spoken or one is spoken and one is signed (Petitto et al., 2001). However, two factors may be different for monolingual and bilingual children: the strategies for word learning and the rate and extent of vocabulary acquisition.

One strategy that allows children to rapidly learn new words is to assume that novel words signify unfamiliar objects, presenting a simple pairing of word and concept. This strategy of word-meaning assignment follows from what Markman and Wachtel (1988) posit as the mutual exclusivity constraint—the assumption that a thing can only have one name—although this assumption need not be innately determined. The evidence for mutual exclusivity is that children appear to create mappings between new words and new objects—for example, if a child hears the word “bik” while looking at a cup and an unknown object, the child will assume that the novel item is called a bik. But bilingual children already know that things can have more than one name—the known object could be “a cup” or “une tasse.” Do bilingual children follow the strategy of mapping unknown words to unknown objects? The evidence is mixed, with some studies reporting less reliance on this strategy for bilingual children (Bialystok, Barac, Blaye, & Poulin-Dubois, 2010; Davidson & Tell, 2005) but others reporting no difference between monolingual and bilingual children (Au & Glusman, 1990; Merriman & Kutlesic, 1993). More convincing, however, is evidence from a study by Byers-Heinlein and Werker (2009) in which they compared the adherence to this strategy by children learning one, two, or three languages. Their results showed a systematic decline in the reliance on this heuristic with the number of languages being learned. These results, in conjunction with those reported by Kovacs and Mehler (2009a) suggesting that phonological word structures are perceived differently by monolingual and bilingual children, are consistent with a view in which the actual mechanisms of word learning used by monolingual children differ from those used by bilingual children. Importantly, however, the essential cognitive landmark that guides these mechanisms, namely, the time at which the child is able to produce the first meaningful word, is comparable for all children.

The second difference in word learning between monolingual and bilingual children is in the size of their developing vocabularies. As in phonological discrimination and first word production, the timetable for the critical milestone is similar for children with both types of experience. In this case, the crucial landmark is the establishment of a vocabulary of 50 words, which is achieved by both monolingual and bilingual children at about 1½ years old (Pearson et al., 1993; Pettito, 1987; Pettito et al., 2001), at least for total vocabulary across the two languages. Beyond that, however, the evidence is compelling that, on average, bilingual children know significantly fewer words in each language than comparable monolingual children. A careful investigation examining how many words children between 8 and 30 months old knew in each language confirmed that, on average, this number was smaller in each language for bilingual children than for monolingual learners of that language (Pearson et al., 1993). The number of words in the total vocabulary of a bilingual child, however, is difficult to estimate: Do proper names count for one language or two? Do cognates count once or twice, especially if the pronunciation is unclear? Do childish sounds that are not quite words count as words if they have a consistent meaning?

A clearer illustration of the relative vocabulary size of monolinguals and bilinguals comes from a study of children who were older than those in the Pearson et al. (1993) analysis. Bialystok, Luk, Peets, and Yang (2010) measured the receptive vocabulary of over 1,700 children between the ages of 3 and 10 years old. All the bilingual children spoke English and another language, with English being the language of the community and school for all children. Across the sample and at every age studied, the mean standard score on the English Peabody Picture Vocabulary Test (PPVT) of receptive vocabulary (Dunn & Dunn, 1997) was reliably higher for monolinguals than for bilinguals. These results are shown in Figure 1. At least in one of the two languages and, importantly, the language of schooling, monolingual children had an average receptive vocabulary score that was consistently higher than that of their bilingual peers. It is important to note, however, that the disparities were not equivalent for all words. In a subset of 6-year-olds in the sample, all children achieved comparable scores on words associated with schooling (e.g., astronaut, rectangle, writing) but bilinguals obtained significantly lower scores for words associated with home (e.g., squash, canoe, pitcher). Therefore, the nature of the smaller vocabulary of bilingual speakers of each language than that of monolingual speakers is in fact somewhat complex (Bialystok, Luk, et al., 2010).

The hallmark of human language, however, is not sounds or words, but the grammatically constrained combinations of units to form utterances or sentences. Again, the transition into this stage of language acquisition occurs on the same timetable
for children learning one or more languages: The first word combinations for all children appear at about 1½ years old (Pearson et al., 1993; Petitto et al., 2001), with utterances becoming incrementally more complex on a similar trajectory (de Houwer, 1995). The details of children’s increasing grammatical sophistication appear to be tied to the specific language, with examples for this point coming from children learning English and Spanish (Gathercole, 1997), English and French (Paradis & Genesee, 1996), and French and German (Meisel, 1990).

Current theories of language acquisition are based on the idea that there is a deep connection between words and structure: Grammar is part of the linguistic system and emerges seamlessly when the lexicon has reached a critical mass. The first evidence for structure occurs when the child knows about 50 words, a relationship demonstrated for both monolingual (Bates & Goodman, 1997) and bilingual (Conboy & Thal, 2006) children. In this sense, discussion of children’s early grammar is not different in kind from the discussion of their early lexicon, but the issues in their development present themselves in different ways. And if language acquisition is not guided by dedicated modules equipped to detect and record grammatical structure, then what directs this process? From the cognitive perspective, the linguistic and cognitive systems are intimately interconnected, each guiding the other and profiting from the symbiotic relationship. What happens when a child is learning two languages?

Across the major linguistic features—sounds, words, grammar—the acquisition of language by monolingual and bilingual children follows a similar timetable for milestones that largely reflect cognitive ability, but the linguistic competence that is developing is different. Partly because linguistic knowledge for bilingual children is divided across two languages, the organization and richness of the representational system in each language is different from that acquired by a monolingual speaker of one of the languages. Similarities in developing cognitive abilities keep the process of language acquisition on a common time course, but variation in input and use make the developing linguistic systems quite different both qualitatively and quantitatively. Understanding bilingual language ability and the bilingual mind more broadly requires understanding these interfaces between the linguistic and cognitive systems.

Two languages in the mind

The bilingual mind presents an intriguing set of puzzles. Are the two languages represented in separate or in overlapping systems? Are concepts duplicated or shared across languages? Do interactions between languages facilitate or impede language production? How are the selection of the target language and avoidance of the nontarget language achieved? How does the bilingual switch between languages, both intentionally and unintentionally? None of these questions applies to monolingual language use, so from the outset, the presence of two languages in mind changes fundamental aspects of language processing. Moreover, these questions are all inherently about cognitive systems at least as much as they are about linguistic ones; switching between representational systems and avoiding interference are processes routinely handled by the general executive control system. Therefore, bilingual language use must be intimately tied to a cognitive system in a way that is less essential for monolingual speech. It is those
relations between language and cognition that will be examined in this section: How is language processing different when there are two fully elaborated linguistic systems available? How does that situation change the cognitive processes whose responsibility it is to manage those language systems? There is an active body of research examining these questions, comparing how bilinguals can carry out these tasks in their two languages (for excellent reviews of this literature, see Kroll & de Groot, 2005). However, the present question is not to compare processing of the two languages of bilingual speakers but to compare monolinguals and bilinguals as they perform similar tasks.

To understand how the simple act of speaking may be different for monolinguals and bilinguals, it is necessary to acknowledge two crucial differences between these groups. First, the knowledge base from which all language processing proceeds is less rich or less interconnected for a bilingual in each language than it is for a monolingual speaker of one of those languages. The most salient difference in the language competence of monolingual and bilingual children is in the vocabulary scores obtained in a given language, as described earlier (Bialystok, Luk, et al., 2010)—a pattern that may persist into adulthood. Although it is more difficult to attribute reliable differences in adults’ vocabulary size to bilingualism versus monolingualism than it is for that of children because of the enormous variation in adults’ knowledge of words, there is nonetheless evidence that such systematic differences exist (e.g., Bialystok, Craik, & Luk, 2008a; Portocarrero, Burright, & Donovick, 2007). Gollan and colleagues argue that the essential feature of bilingual representations is the “weaker links” that are established within the network because of less frequent use of each language (Gollan, Montoya, Cera, & Sandoval, 2008); simply using each language less often produces weaker connections in the network than would emerge from greater use. In this view, the knowledge resources underlying language performance for monolinguals and bilinguals who are comparable on many other cognitive abilities are not equivalent.

Second, it is now well documented that both languages of a bilingual are jointly activated even in contexts that strongly bias towards only one of them. Evidence for this claim comes from both behavioral (Beauvillain & Grainger, 1987; Colomé, 2001; Grainger, 1993; Hernandez, Bates, & Avila, 1996; Francis, 1999; Kroll & de Groot, 1997) and imaging studies (Marian, Spivey, & Hirsch, 2003; Martin, Dering, Thomas, & Thierry, 2009; Rodriguez-Fornells, Rotte, Heinze, Nosselt, & Munte, 2002). One of the first pieces of evidence for this conclusion comes from an ingenious experiment by Gutten tag, Haith, Goodman, and Hauch (1984, Experiment 2). On each trial, bilingual participants viewed a word drawn from one of four semantic categories (e.g., metals, clothing, furniture, and trees); two categories were assigned to one response key and the other two categories to a second key. The participant’s task was to press the designated key to indicate the category membership of the target word as rapidly as possible. Each stimulus word also had copies of a further word above and below it as flanker items. These flankers were always in the participant’s other language and belonged to one of four categories: translations of the target word, a different word drawn from the same semantic category as the target, a word from a different category but requiring the same response, or a word from a category requiring a different response. The crucial result is that response times were significantly longer in the second two conditions, showing that participants were unable to ignore the flankers and that some analysis of the flankers’ categories (and possibly responses) took place despite the fact that the flankers were in the nonused language.

This joint activation of the two languages creates a unique need for selection in bilinguals in which language processing must resolve competition not only from within-language alternatives as monolinguals do to select among close semantic neighbors (words that share semantic features, e.g., cup vs. mug; Luce & Large, 2001; Mirman & Magnuson, 2008; Vitevitch, 2002) but also from between-language alternatives for the same concepts (e.g., cup vs. tasse). The predominant view is that language selection does not normally occur prior to speech, making this selection part of bilingual speech production (Kroll, Bobb, & Wodniecka, 2006). For this reason, a somewhat different set of attention and control procedures is necessary for speech production in bilinguals than is necessary for monolinguals (Green, 1998). However, there is less agreement on what those special processes might be. Some studies have shown that the nontarget language is actually inhibited while using the other language (e.g., Levy, McVeigh, Marful, & Anderson, 2007; Philipp & Koch, 2009), but others indicate that correct selection can be achieved by increasing the activation of the preferred response (Costa, Santedeban, & Ivanova, 2006; see Costa, 2005, for a discussion of these views). As we describe later, these alternatives need not be mutually exclusive: Selection depends on the activation level of both the target item to be selected and that of the competing items, incorporating as well views that reject the role of competition and instead focus on selection (Caramazza, 1997; La Heij, 1988; Mahon, Costa, Peterson, Vargas, & Caramazza, 2007; Roelofs, 2003). Therefore, selection is facilitated by either preferentially enhanced activation of the target, inhibition of the competitor, or both. Whatever the mechanism, selection of appropriate lexical items for bilinguals involves either different or additional processes than does the same activity for monolinguals. Taken together, the differences in the linguistic representations and differences in the selection mechanisms lead to sustained differences between monolinguals and bilinguals in fluent speech production.

Although ordinary conversation does not generally signal observable deficits in bilingual language processing, controlled experimental procedures can reveal more subtle differences between these two groups. Two such features are the speed with which target words can be retrieved in response to a cue and the number of words that can be generated to satisfy a criterion. Evidence for the first comes primarily from studies of picture naming or semantic classification, and evidence for the second comes from studies of verbal fluency.

**Lexical retrieval in bilinguals.** Much of the research in lexical retrieval compares the relative ability of multilingual speakers
to perform such tasks as naming the pictures in their two (or more) languages (Costa & Santesteban, 2004; Hernandez, Martinez, & Kohnert, 2000), making semantic classifications for words in the two languages (Dufour & Kroll, 1995), or translating between languages (Kroll & Stewart, 1994). The purpose is to compare lexical access to the two languages and, in some cases, as in the study by Dufour and Kroll (1995), to compare bilinguals who are more or less fluent in the two languages. The issue we are discussing here is different: to compare monolingual and bilingual speakers naming pictures in the same language. The comparison is inherently fraught with difficulty: If we assume that bilinguals never have identical proficiency in their two languages and, moreover, that even their ability in their stronger language may not fully resemble the language competence of a monolingual speaker of that language, then any comparison of monolinguals and bilinguals seems unfair. And yet, proficient bilinguals manage to function perfectly well, belying the notion of an underlying handicap. Thus it may be that the task of rapidly accessing target lexical items is carried out differently by monolinguals and bilinguals, an outcome that would be important in understanding the relation between language and cognitive systems in the bilingual mind.

Research shows that bilingual participants take longer and make more errors than monolinguals on naming tasks. Using the Boston Naming Task (Kaplan, Goodglass, & Weintraub, 1983), bilinguals produced fewer correct responses (Roberts, Garcia, Desrochers, & Hernandez, 2002; Gollan, Fennema-Notestine, Montoya, & Jernigan, 2007) and made more errors on a speeded version of the task (Bialystok et al., 2008a) than did monolinguals. On timed picture naming, bilinguals performed more slowly than did monolinguals (Gollan, Montoya, Fennema-Notestine, & Morris, 2005). Similar results (slower responses in bilinguals) are found in both comprehending (Ransdell & Fischler, 1987) and producing words (Ivanova & Costa, 2008), even when bilinguals respond in their first and dominant language. The simple act of retrieving a common word seems to be more effortful for bilinguals.

Healthy aging is frequently accompanied by a reduction in productive language abilities—searching for words and names becomes a more salient part of every conversation. Consistent with this trend, picture naming is carried out more slowly by older adults than by younger adults, even for monolinguals (e.g., Albert, Heller, & Milberg, 1988). Therefore, older bilinguals should find lexical access particularly difficult, since both age and language status are associated with poorer performance. The situation is even more problematic for older bilinguals who may have spent the majority of their adult lives using one of their two languages, usually the second language (L2), and have been removed from a daily context that supports the first language (L1). The outcome of this situation can be attrition of the L1. Therefore, difficulties in performance on tests of lexical access such as picture naming can be attributable to normal aging, L1 attrition, or both. These possibilities were evaluated in a study by Goral, Libben, Obler, Jarema, and Ohayon (2008) comparing younger and older Hebrew-English bilinguals who lived in an English-speaking or Hebrew-speaking society. Their conclusion was that the slower retrieval time for older bilingual adults in their L1 was caused primarily by attrition of that language and not by aging. These results point to the importance of gauging proficiency level, such as vocabulary knowledge, in linguistic processing and in performance on psycholinguistic tasks.

Linguistic differences between monolinguals and bilinguals go beyond vocabulary size. The consistent result showing longer picture-naming times for bilinguals suggests that word retrieval is carried out differently for bilinguals than for monolinguals. To explore a possible explanation for this effect, Hernandez and Meschyan (2006) conducted a functional magnetic resonance imaging (fMRI) study in which Spanish-English bilinguals who learned the L2 in adolescence named pictures in both languages. The results showed that naming the pictures in the weaker second language produced greater activity in the executive control network, a system that will be described in more detail in Sections 2 and 3. Extrapolating to monolingual performance, where naming is always carried out in a strong language, it appears that this executive control network is involved in word retrieval for bilinguals in a way not required by monolingual language production. We will return to this idea later.

Studies of verbal fluency. The second experimental paradigm in which reliable differences between monolinguals and bilinguals have been reported is the verbal fluency task. The basic procedure is to ask participants to generate as many words as possible in 60 seconds that satisfy a criterion determined either by the category (semantic fluency) or the initial letter of the word (phonological fluency). There are standardized versions of the task, such as in the Delis-Kaplan Executive Function Battery (DKEFS; Delis, Kaplan, & Kramer, 2001) and the Controlled Oral Word Association Test (COWAT; Strauss, Sherman, & Spreen, 2006), that allow performance to be interpreted in terms of normalized tables and used as an instrument for neuropsychological assessment. The clinical applications of this test are explained in Section 4, but in the present discussion we consider the task as an experimental tool. The semantic and letter versions assess different aspects of competence and engage different processes. The demands of category fluency are congruent with normal procedures for word retrieval in that the meaning is cued and words associated with that meaning are primed and available. Thus, when asked to generate names of fruits, the inherent associations among various fruits in semantic memory facilitate recall. In contrast, the letter fluency condition imposes an arbitrary criterion on word generation: Conversation does not normally require the generation of words by virtue of their initial letter. Moreover, the letter fluency task additionally imposes a set of restrictions that exclude repetitions of words in different forms and therefore requires more intensive monitoring and working memory. Thus, category fluency is strongly indicative of vocabulary size (how many types of fruit can you name?) and letter fluency requires additional and effortful procedures for monitoring and controlling attention (how well can you keep
A more detailed understanding of performance on the verbal fluency task comes from examining the function showing the production of words in real time across the 1 minute allotted to each trial. Following the logic explained by Rohrer, Wixted, Salmon, and Butters (1995), a deficit in vocabulary size should manifest itself in a function that shows very few words being produced toward the end of the time period because the potential set of items has been exhausted. In this case, monolinguals would continue producing words later into the time course than would bilinguals. In contrast, slower time to produce each item, possibly because of the need to resolve competition from the nontarget language, would produce a function that continues longer into the time period than one representing faster retrieval of the same total number of words. In this case, bilinguals would produce words later in the time course than monolinguals.

These predictions were tested in two studies using time-course analysis to compare monolinguals and bilinguals performing a verbal fluency task. A study by Sandoval, Gollan, Ferreira, and Salmon (2010) compared monolinguals and Spanish-English bilinguals who reported high proficiency in both languages for their performance on several category and letter fluency conditions in English, and in a second experiment also compared the time course of retrieval from bilinguals’ two languages (English vs. Spanish). In another study by Luo, Luk, and Bialystok (2010), a standardized version of the category and letter fluency tasks in English was administered to monolinguals and bilinguals who were either matched on English vocabulary or had a lower English receptive vocabulary. In both studies, the bilinguals produced words later into the allotted time, indicating slower and more effortful retrieval for each word produced, likely due to interference from the nontarget language (Sandoval et al., 2010). In addition, the comparison between the two English proficiency groups in the study by Luo et al. indicated a second effect attributable to vocabulary size. Once vocabulary was matched, the bilinguals with English proficiency comparable to that of monolinguals performed as well as the monolinguals on the category fluency task and better than monolinguals on letter fluency. Having equated for differences in vocabulary resources, the bilinguals were able to display better control than the monolinguals in the condition that required monitoring and working memory. Figure 2a displays the results for category fluency in which monolinguals and high-vocabulary bilinguals show identical retrieval patterns because performance is driven primarily by vocabulary size, which in this case is matched. Figure 2b displays the results for letter fluency; in this case, the high-vocabulary bilinguals maintain a higher production rate throughout the time course than do the other two groups because the task additionally requires high levels of executive control.

These results point to the need to guarantee that participants who are performing a language task have linguistic resources adequate to carry out the task. Without explicitly controlling for language proficiency, it is impossible to localize the effects of bilingualism as opposed to the effects of weaker proficiency in the language of testing. Moreover, when proficiency in the
two languages had been controlled by using receptive vocabulary as a matching variable, a bilingual advantage emerged in the letter fluency task. This pattern was replicated in a comparison between monolinguals, bilinguals with matched vocabulary, and bilinguals with lower vocabulary on a simple behavioral comparison of the number of words produced in each of these fluency tasks (Bialystok, Craik, & Luk, 2008b). Clearly, not all tasks requiring processing of linguistic material are performed more poorly by bilinguals.

Control over linguistic resources. To this point, the studies described have generally found more effortful (longer response

![Figure 2](https://example.com/figure2.png)

*Fig. 2.* Number of items produced as a function of time in (A) category task and (B) letter task for monolinguals, high-vocabulary (HV) bilinguals, and low-vocabulary (LV) bilinguals. Best fit lines are logarithmic functions. From Luo, Luk, and Bialystok (2010).
time, RT) or poorer (more errors) performance by bilinguals than by monolinguals when rapid retrieval of specific lexical items is required. When language proficiency is matched, however, bilinguals perform as well as monolinguals in category fluency (which depends on vocabulary) and better than monolinguals on letter fluency (which depends more extensively on cognitive control). Therefore, at least some of the differences observed between monolinguals and bilinguals on language production tasks reflect a simple difference in linguistic resources and may mask a potential advantage in control over those resources once proficiency has been equated.

If bilinguals do have better control over linguistic resources than do monolinguals, then it should be possible to demonstrate this processing difference in tasks that require monitoring or manipulation of verbal stimuli. Two tasks meet these criteria. The first is a paradigm developed by Jacoby (1991), called the process dissociation procedure (PDP), that is designed to distinguish between automatic (familiarity) and controlled (recollection) aspects of memory. The second is a paradigm called release from proactive inhibition (PI) that assesses the ability to monitor items for their source (e.g., Kane & Engle, 2000). Both paradigms have been widely used in studies of cognitive processes involved in memory performance. Although substantially different from each other, they share the feature that participants are asked to remember words for later recall when an intervening event has made it difficult to keep track of the source of the target words. In the case of PDP, words are presented in two lists, or two formats (for example, visually or orally), and the crucial recall test requires responding only to the words presented in one of them (for example, visually) and ignoring the others. In the case of PI, lists of different words from the same semantic category are presented successively and participants are asked to report the words on the list just heard without reporting words from the previous lists. “Release” from PI is observed when words from a different category are presented. Both tasks, therefore, require monitoring and control to attend to the target words and inhibition to avoid making errors on the distractor words. As predicted, bilinguals obtained lower scores than monolinguals on tests of receptive vocabulary but performed better than monolinguals on both PDP (Wodniecka, Craik, Luo, & Bialystok, 2010) and release from PI tasks (Bialystok & Feng, 2009). Again, separating verbal ability from control over verbal processing produces a more complex picture in which bilinguals demonstrate better processing in the context of poorer verbal performance.

Cognitive control and bilingual language processing

All the illustrations of language acquisition and use described in this section have demonstrated the importance of considering the interaction between language and cognitive systems in explaining outcomes for bilinguals. Bilingual children acquire language on the same timetable as monolingual children, largely because this timetable is determined by the process of cognitive development. As acquisition proceeds, however, bilingual children develop different types of competence (e.g., smaller vocabulary in each language) and probably use different strategies (e.g., phonemic cues and mutual exclusivity for word learning). In adulthood, the ability of bilinguals to effectively use language in such tasks as word retrieval and word generation depends on both linguistic competence and cognitive procedures for access and monitoring. Thus, levels of vocabulary determine how many words can be associated with a meaningful category but levels of control determine how many words can be selected to fit an arbitrary restrictive criterion.

What is the source of these interactions? One possibility is that the interacting systems are set in motion because the joint activation of the two languages for a bilingual creates a problem not experienced by monolinguals—namely, the need to select from the target system in the context of compelling and active alternatives. There is substantial evidence, described in Sections 2 and 3, that the response to this conflict is to recruit the executive control system that has evolved to resolve conflict across all domains of perceptual and cognitive processing. The constant use of this executive control system for bilingual language management opens the possibility that the system itself is modified, changing its valence or efficiency for all tasks. That is, the use of a set of executive control procedures to manage attention to language, to avoid interference from the nontarget language, and to monitor two simultaneously active languages may alter the nature or efficiency of those executive control processes more generally. This possibility is examined in the next section. To anticipate, the evidence suggests that whereas bilingual children and adults have somewhat lower vocabulary levels than their monolingual counterparts, the bilinguals possess an advantage in cognitive control that generalizes beyond language processing to other aspects of cognitive functioning.

2. How Bilingualism Affects Cognitive Control

For many years it was assumed that while bilingualism might be an asset for adults—in terms of culture, travel, and trade, for example—it was a handicap for children in the educational system. The idea was that learning in two languages imposed an additional burden on schoolchildren who must learn two vocabularies, two sets of grammar, and probably two sets of cultural habits and expectations. This negative view of bilingualism was at least questioned by the results of a study by Peal and Lambert (1962). They gave a battery of intelligence tests to French-speaking children in Montreal who were also fluent English speakers. They expected to find that monolingual and bilingual children would be equivalent on measures of nonverbal intelligence but that bilinguals would obtain lower scores on verbal measures. To their surprise, however, bilingual children outperformed their monolingual peers on virtually all of the tests, including tests of nonverbal intelligence. Further analysis revealed that there was little difference between the groups on spatial-perceptual tests but that the bilingual children showed an advantage on tests requiring symbol manipulation and reorganization. This latter finding
has the interesting implication that extra effort and more extensive learning in the area of language apparently confers benefits to nonverbal mental abilities, refuting the idea that language is a separate module of mind and brain that relies on dedicated processes (e.g., Fodor, 1983); instead, language must be viewed as recruiting processes from the general cognitive system. On the basis of their unexpected findings, Peal and Lambert suggested that bilingual children may show enhanced mental flexibility, perhaps as a consequence of having to switch between their two languages.

The study by Peal and Lambert (1962) may be criticized on the grounds that francophone children in Montreal in 1960 who spoke English were likely of higher than average social class, or at least were the children of intelligent and ambitious parents, and were therefore less representative than their monolingual counterparts (Bialystok, 2001). Nevertheless, the study was important in showing both that bilingualism in children might help rather than hinder the development of other abilities and also that language learning may influence nonverbal cognitive processes supporting the view that language is not a separate and independent module of mind.

Some decades following the Peal and Lambert study, supporting evidence for a bilingual advantage in general cognitive functioning for children was found in studies using a variety of experimental paradigms. For example, Bialystok (1992) reported that bilingual children performed better than their monolingual counterparts on the Embedded Figures Test. In this test, participants must find a simple visual pattern concealed in a larger complex figure. Bialystok suggested that the better performance of bilingual children might reflect their superior ability to focus on wanted information and ignore misleading information. That is, the advantage might be one of enhanced selective attention, involving the ability to inhibit irrelevant or unwanted information and the complementary ability to concentrate on relevant aspects.

This interpretation was in line with another demonstration in which children were asked to judge whether phrases were grammatically correct, regardless of meaning. Bilingual children were better than their monolingual age-mates at ignoring the misleading meaning in sentences such as “Apples grow on noses” or “Why is the cat barking so loudly?” and stating that the grammar was correct (Bialystok, 1988). More generally, research demonstrated enhanced metalinguistic awareness in bilingual children compared to their monolingual peers (Ben-Zeev, 1977; Cummins, 1978; Galambos & Hakuta, 1988; Ricciardelli, 1992).

Why might bilingual children show an advantage in the ability to inhibit attending to unwanted information and select relevant aspects? The answer may follow from the surprising finding described earlier: that when bilingual speakers use one language, the other language is still active. However, this does not mean that a full analysis of incoming stimuli in the nonused language inevitably takes place, nor that formulating speech in the other language during speech. Though such intrusions are rare (Poulisse, 1997; Poulisse & Bongaerts, 1994; Sandoval et al., 2010), these instances reflect occasions in which the appropriate word in the language being used is difficult to locate or the word or phrase in the nonused language is made particularly likely because of the context or its salience. Bialystok (2001) commented that such intrusions are more common in bilingual children than in adults and are also more common (anecdotally at least) in older than in younger adults (Sandoval, 2010). In turn, this age-related pattern suggests that the brain mechanisms responsible for maintaining attentional set (in this case maintaining attention on the selected language) are less effective in childhood and in older adulthood. One candidate for such mechanisms is integrity of frontal lobe functioning, since it is well established that the frontal lobes develop slowly in childhood and are among the first parts of the brain to decline in efficiency in older adulthood (Craik & Grady, 2002; Diamond, 2002; Raz, 2000).

Our suggestion is that bilingual speakers must develop an unusually strong ability to temporarily inhibit access to the nonrelevant language while maintaining attentional set (“maintaining concentration”) on the language in current use. This ability may be mediated by the frontal lobes and may therefore exhibit a lifespan developmental trend that peaks in young adulthood. The further suggestion is that the constant necessity to exercise this inhibitory control leads to the development of particularly effective attentional functions that are then drawn on to mediate good performance on a variety of nonverbal tasks requiring inhibition of unwanted or misleading material and concurrent selection of relevant aspects.

Inhibition or selection?

What would it mean to have enhanced control over attentional functions? When Bialystok (2001) surveyed studies of the effects of bilingualism on children’s cognitive processes, she concluded that “the most consistent empirical finding about the cognition of bilingual children is their advantage in selective attention and inhibition” (Bialystok, 2001, p. 246). This conclusion was based on some of her own work (e.g., Bialystok, 1988, 1992) as well as on a growing number of studies from other laboratories. An example that illustrates how these processes are used by children is the dimensional change card sort task (DCCS) developed by Zelazo, Frye, and Rapus (1996). This is a game in which images that vary on two dimensions, usually shape and color, are sorted according to one of them. For example, cards containing either red or blue circles or squares are sorted into containers marked by an image of either a red square or a blue circle. Children are asked to first sort the cards by one dimension—blues in this box and reds in this box—and then to switch to the other—circles in this box and squares in this box. The dramatic finding is that young
children can easily state the new rule but continue to sort by the first rule; they have great difficulty overriding the habit set up in the first phase. When this experiment was repeated with bilingual and monolingual children aged between 4 and 5 years, the bilingual children were markedly better at switching to the new rule (Bialystok, 1999; Bialystok & Martin, 2004). This result was obtained despite there being no difference in pre-switch performance. The researchers thus concluded that the constant need to inhibit the nonused language generalized to more effective inhibition of nonverbal information.

These demonstrations were followed by studies that extended the investigation to adults and used other paradigms in which a prepotent response tendency must be inhibited. One such situation is embodied in the Simon task. The participant views a screen on which either a red or green square appears; there are two response keys, one for red squares and the other for green squares. The keys are positioned below the sides of the screen, and the squares can appear either immediately above their relevant response key (congruent condition) or above the other key (incongruent condition). Response latencies are longer in the incongruent case, and the difference between incongruent and congruent latencies is termed the Simon effect. If participants are able to resist the misleading information carried by spatial position in the incongruent situation, the Simon effect will be smaller, and we may conclude that they have well-developed inhibitory control mechanisms. Using this logic, Bialystok, Craik, Klein, and Viswanathan (2004) tested groups of younger and older adults who were either monolingual or bilingual on a version of the Simon task. When the colored squares are presented centrally, there is no conflict between the position of the stimulus and side of the appropriate response, and in this case there were no differences in reaction time between monolinguals and bilinguals, although older participants took longer to respond (Fig. 3a). When the colored squares appeared laterally, however, Simon effects were found, and these were larger for monolinguals—especially older monolinguals (Fig. 3b). This evidence for a bilingual advantage in inhibitory control in adults extended the results of previous studies on children. Moreover, the bilingual advantage was especially strong in older adults, suggesting that bilingualism may afford some protection against at least some forms of cognitive aging.

Two other unexpected results emerged from this study. The first is that the bilingual advantage in response time was found for congruent as well as incongruent stimuli. This result was obtained in all three experiments and has been consistently observed in subsequent studies (e.g., Costa, Hernández, & Sebastian-Galles, 2008). Why should there be a bilingual advantage for congruent stimuli when there is no misleading information to inhibit? Most experiments of this sort are run under mixed conditions in that experimental runs contain both congruent and incongruent stimuli, so participants must keep the rule in mind throughout the experimental run and monitor each trial for the type of processing needed (conflict or no conflict). It may be that bilinguals are also better at these aspects of executive control. The test of this conjecture is to check what happens in experiments containing pure runs of all congruent or all incongruent stimuli, and the finding there is that the bilingual advantage disappears (Bialystok, Craik, & Ryan, 2006).

The second unexpected result found by Bialystok et al. (2004) was that prolonged practice on the Simon task reduced the difference between monolinguals and bilinguals. In Experiment 3, participants performed the Simon task for 10 consecutive blocks of 24 trials; by the end of the session the monolingual disadvantage had disappeared and both groups showed minimal differences between congruent and incongruent stimuli. It is interesting to speculate that everyone may be able to inhibit the effects of misleading information in specific situations with sufficient practice but that bilinguals can learn this type of inhibition more rapidly.

The Stroop effect may be considered the “gold standard” of tests of inhibition. In this paradigm, participants name colors as rapidly as possible, both when the colors are presented as colored squares on a screen and when the stimuli are color names (e.g., “red,” “green,” “blue”) but presented in a different colored font (e.g., the word “red” printed in green ink). The difference in speed between naming colored squares and the color of words is the Stroop effect; again, a smaller Stroop effect indicates a strong ability to inhibit the misleading tendency to name the word rather than its color. Bialystok et al. (2008a) tested groups of 24 younger and older adults who were monolingual or bilingual on this paradigm. In four different conditions, participants named the color of displays of Xs, named a color word presented in black font, named the font color of words printed in their own color (congruent condition), and named the font color of words printed in a different color (the incongruent Stroop condition). For the control conditions (naming words and colored Xs), naming times were faster for words and for younger participants but there were no language-group differences. Response times for the congruent and incongruent colored-word conditions are shown in Figure 4 as differences (positive or negative) from the time taken to name colored Xs. The figure shows that congruent stimuli are associated with relatively faster response times (a facilitation effect) and are indicated by positive RT differences in the figure, whereas incongruent stimuli show the classic Stroop pattern in which slower response times are indicated by negative RT differences. Statistical analysis revealed a significant three-way interaction of age, language, and congruence; both younger and older bilinguals sustained smaller costs than their monolingual peers, but only the older bilinguals showed greater facilitation. We may thus conclude that the older bilinguals exhibited greater degrees of cognitive control than their monolingual counterparts, in that they both took greater advantage of congruent conditions and at the same time were less impaired by incongruent conditions. Younger bilinguals showed the latter effect but not the former.

Other results from the Bialystok et al. (2008a) study included a bilingual advantage for the older participants in a version of the Simon task using directional arrows, but no bilingual advantage for either age group in a condition in which
participants were instructed to respond in the direction opposite to that indicated by a single arrow. There was also no bilingual advantage on the Sustained Attention to Response Task (SART; Robertson, Manly, Andrade, Baddeley, & Yiend, 1997), which involves withholding a response to the number 3 while responding rapidly to all other digits. In both these latter tasks, the participant can encode a simple rule (e.g., “press in the opposite direction”) and then follow that rule; there is essentially no need to select one aspect of the stimulus and suppress other aspects, as with the Simon, Stroop or flanker tasks. This account claiming no need for control in these tasks is reinforced by other results showing no bilingual advantage in children who were instructed to respond “day” when shown a picture of a dark night, and “night” when shown a sunny day (Martin-Rhee & Bialystok, 2008). These investigators also replicated the finding of no bilingual advantage in children given the reverse arrow task, even though the same children demonstrated a bilingual advantage when the arrows were placed in side positions on the display that created conflict.

This pattern of presence and absence of advantages is in line with the distinction between interference suppression and

Fig. 3. Mean reaction time (RT) on Simon task by decade for monolinguals and bilinguals. Graph a shows mean RT for the control condition; Graph b shows mean RT cost as the difference between congruent and incongruent trials (Simon effect). From Bialystok, Craik, Klein, and Viswanathan (2004).

Fig. 4. Mean reaction time (RT) and standard error for facilitation and cost for young monolinguals and bilinguals and older monolinguals and bilinguals in the Stroop task. The values are mean differences from baseline (0 milliseconds) calculated as the average difference in the time taken to name colors from the time taken to name neutral stimuli (Xs). From Bialystok, Craik, and Luk (2008a).
response inhibition proposed by Bunge, Hazeltine, Scanlon, Rosen, and Gabrieli (2002). Interference suppression refers to situations in which misleading information evokes a faulty response and must therefore be ignored or suppressed; this appears to be the type of situation that bilinguals can deal with particularly well. Response inhibition is the ability to avoid responding in error to a habitual or highly salient cue, and bilinguals show no advantage under these circumstances. In other words, the bilingual advantage appears when there is conflict between two potential responses, but not when there is a need to withhold a single primed response.

As a final converging point, Kimberg, D’Esposito, and Farah (1997) have commented that patients with lesions in the prefrontal cortex are impaired on tasks in which the most salient cue evokes the wrong response and must therefore be suppressed to select the cue associated with the correct response. If one effect of bilingualism is to boost frontal lobe functions, it follows that bilingual children and adults should be adept at tasks involving interference suppression.

Converging evidence from other studies has supported the conclusion that bilinguals show strong abilities to inhibit irrelevant or interfering information. Zied and colleagues (2004) found that balanced bilingual adults of various ages responded more rapidly than unbalanced bilinguals on the Stroop task. In an ingenious series of studies, Philipp and colleagues (Philipp, Gade, & Koch, 2007; Philipp & Koch, 2009) asked participants who were fluent in three languages (English, French, & German) to switch among their languages in a number naming task; thus “2” was named either “two,” “deux,” or “zwei” depending on a concurrent instruction. The main finding was that naming in language A was slower on the third trial of a sequence ABA than in a sequence CBA. That is, A (e.g., French naming) was slower on the third trial of a sequence French, German, French than it was in the third trial of a sequence English, German, French, suggesting that in the first sequence French was subjected to a temporary global inhibitory effect to permit access to German. When French was needed immediately after that, negative priming slowed access to the target name. Although there is no bilingual advantage in this study—monolinguals were not tested, and the study was not designed to test for bilingual advantages—the results demonstrate the role of a general inhibitory process applied to the nonused language in order to avoid interference effects in the selected language.

Negative priming was also used in an experiment by Treccani, Argyri, Sorace, and Della Sala (2009). Targets could appear at one of four positions on a screen and participants responded by pressing one of four keys. When a target was accompanied by a distractor stimulus in another location, bilingual adults were better able to ignore it (interference suppression) and so made fewer errors than did their monolingual counterparts. However, the bilinguals were more negatively affected (making more errors than monolinguals) when a target appeared in the position previously occupied by a distractor item. In this situation, the better inhibition of the distractor carried over to the next trial, providing more negative priming to the bilingual participants. The authors concluded that whether bilinguals show an advantage or a disadvantage relative to monolinguals depends on task characteristics.

The studies reviewed so far have endorsed the notion that the bilingual advantage found in these studies is due to an advantage in inhibition or suppression of interfering material, but there remains the possibility that bilinguals show an advantage in the positive selection of wanted information. The latter interpretation is favored by a number of investigators. Costa, Miozzo, and Caramazza (1999) argue that although lexical candidates in both languages are active during the planning of an utterance, the intention to speak in one language rather than another effectively restricts selection to words in the target language. Colzato and colleagues (Colzato et al., 2008) set out to compare what they termed “active inhibition” with “reactive inhibition.” By active inhibition they mean general global suppression of the nonrelevant language (cf. inhibition in the study by Philipp & Koch, 2009) and by reactive inhibition they mean lack of suppression of specific interfering stimuli. Evidence for the latter was found in the attentional blink paradigm in which detection of a target stimulus is impaired if the same stimulus was presented earlier in a rapid sequence of events. The authors predicted that if bilinguals show more reactive inhibition, then they will process the first presentation of the target to a greater extent and therefore show less suppression of intervening items. Without suppression, these items would then interfere more with the second presentation of the target, creating a larger attentional blink effect. This is what they found, and so they suggested that the bilingual advantage is not due to constant exercise of inhibition of the nonused language but rather to prolonged practice at maintaining the relevant attentional set, though they grant that such selection may involve strong inhibition of competing items.

The debate over inhibition versus selection may rest on a false dichotomy: inhibition may not be an all-or-none phenomenon but may rather be found to different degrees under some circumstances. One such factor that might influence the degree of inhibition required to perform a task is the effect of context. Kroll, Bobb, Misra, and Guo (2008) describe work showing that cross-language cognates were activated (that is, naming a word in one language activated its cognate in bilinguals’ second language) when a word was named out of context, but this cognate facilitation was eliminated in contexts that were semantically constrained in that the required word was more clearly determined from the context (see also Schwartz & Kroll, 2006; van Hell & de Groot, 2008). One possibility, then, is that the degree to which both languages are active may not be constant but may vary probabilistically with the contextual constraints provided by language, topic, and the external environment.

Another possible effect of context was suggested by Costa, Hernández, Costa-Faidella, and Sebastian-Galles (2009). They tested monolinguals and bilinguals on versions of a flanker task in which different conditions contained varying proportions of incongruent trials: 8%, 25%, 50%, or 92% (therefore mixed with 92%, 75%, 50%, or 8% congruent trials, respectively). The bilingual advantage was strongly present in the 50%/50% version,
reduced in the 75%/25% version, and entirely absent in the 92%/8% version. The authors conclude that the bilingual advantage is related to their greater ability to monitor the environment when the probability of change is high, as in the 50%/50% condition. Under low-monitoring conditions, when most of the trials are of one type, there is little need to monitor and thus no bilingual advantage is found. The notion of monitoring is similar to the idea of set maintenance described previously by Colzato et al. (2008). Costa and his colleagues also make the interesting prediction that bilinguals who live in situations in which their two languages are used in different contexts (e.g., Italian at home, English at work) rarely need to monitor language changes and so may not develop strong monitoring abilities and thus show no bilingual advantage.

Finally, the distinction between selection and inhibition was examined in a study by Hernández, Costa, Fuentes, Vivas, and Sebastian-Gallés (2010), in which participants rapidly judged how many items (letters or numerals) appeared on a screen. The items appeared either in a congruent form (1, 22, 333), an incongruent form in which the displayed numerals did not match the required response (e.g. 3, 11, 222), or a neutral form (Z, GGG, MM). Relative to the neutral baseline, congruent stimuli were associated with faster response times (facilitation) and incongruent stimuli with slower response times (interference). Bilingual participants showed smaller interference effects but larger facilitation effects than their monolingual counterparts (cf. Bialystok, Craik, & Luk, 2008a), so their advantage may be described as one of better executive control of perception/action processing. The conclusion of Costa and colleagues is that the bilingual advantage is reasonably high level, involving top-down working memory processes, and is manifested as enhanced set maintenance or monitoring. This description suggests that the advantage may stem from enhanced frontal lobe effectiveness, as suggested by Bialystok (2001).

**Selective attention and executive control**

We have seen in the previous section that research aimed at assessing inhibitory abilities in bilinguals evolved to consider such concepts as selection, set maintenance, and monitoring. However, the distinction between these concepts and notions of attention and executive control is difficult to discern. In many ways, all these concepts are simply aspects of attention and executive control. Therefore, in this section we consider work that assesses group differences in attention and control more directly.

Costa et al., (2008) examined the performance of monolingual and bilingual participants on the attentional network task (ANT) developed by Fan, McCandliss, Sommer, Raz, and Posner (2002). The bilinguals were young adults who spoke Catalan and Spanish; the monolinguals were young adults who spoke Spanish only. The ANT task assesses abilities on three different attentional networks: alerting, orienting, and executive control. The test is a flanker task in which the participant responds to the direction of a central arrow that is flanked by two arrows on each side pointing in the same (congruent) or different (incongruent) direction as the central target arrow. Alerting is studied by presenting a cue before the target stimulus, and orienting is assessed by the presence or absence of a cue signaling the future spatial position of the target. The results supported the hypothesis of greater attentional control by bilinguals in the alerting and executive control networks. The bilingual participants responded faster than the monolinguals on all conditions and showed a smaller cost for the incongruent trials, indicating better conflict resolution. Two final results from this study were that this bilingual advantage disappeared by the third block of trials (cf. Bialystok et al., 2004, Study 3) and that bilinguals had smaller switching costs between congruent and incongruent trials, a point to which we will return.

Similar results were obtained by Carlson and Meltzoff (2008) with much younger participants. They administered a battery of executive function tests to 50 kindergarten children who were English-speaking monolinguals, English-Spanish bilinguals, or children who were in a language immersion elementary school. The major finding was that the native bilingual children performed better on the executive function battery than did both other groups, once differences in age, vocabulary, and parents’ education and income levels were statistically controlled (recent work extends this finding that bilingualism can offset the negative effects of lower socioeconomic status on task switching to young adults; Prior & Gollan, 2010). The effects were specific to only some aspects of control: There were no bilingual advantages in suppressing a motor response on delay-of-gratification tasks (response inhibition) but significant advantages on conditions requiring memory and inhibition of attention to a prepotent response (interference suppression; cf. Martin-Rhee & Bialystok, 2008). The authors conclude by endorsing the notion that “language experiences can influence further development of frontal lobe functions such as inhibition and the control of attention” (p. 293).

**Task switching**

The features of executive control discussed to this point are somewhat invisible in ordinary cognitive performance. The interference suppression that allows us to perform a Stroop task or ignore misleading flankers in the ANT seems to have little role in everyday cognition. A more noticeable aspect of executive control might be task switching—the ability to move easily between two tasks, keeping two protocols simultaneously active. Task switching might come closest to the special processes bilinguals engage in as they switch between languages.

In one of the first studies to find positive things to say about bilingualism, Peal and Lambert (1962) suggested, as we noted earlier, that bilingual children may show an advantage in mental flexibility—an idea presumably stemming from the fact that bilinguals must switch easily from one language to another. A large body of research investigates task switching, typically by asking participants to classify a long series of two-dimensional stimuli by one criterion or the other as rapidly as possible. Such sorting times are relatively short when successive trials
continue with the same criterion (e.g., continue sorting by shape), but local switching costs are incurred when instructions change to sort by the other dimension (e.g., switch and sort by color). Some runs of trials involve only one dimension (e.g., all trials require sorting by color), so it is also possible to measure mixing costs, defined as the difference in time taken to classify a set of trials under single- and dual-criterion conditions (Meiran & Gotler, 2001; Pashler, 2000). Typically, sorting times are longer when it is necessary to bear in mind the requirement to switch when the instruction changes.

Several studies have now explored monolingual—bilingual differences in such paradigms, with the prediction that bilinguals should show reduced costs, owing perhaps to their prolonged practice in switching languages and monitoring which language may be spoken in which context. The prediction with regard to which type of cost might be affected by bilingualism is less clear. To take an analogous difference between individuals—development and aging over the lifespan—the typical finding is that younger adults have smaller mixing costs than children or older adults do, whereas the age groups do not differ markedly on local switch cost (Reimers & Maylor, 2005; for review, see Mayr & Liebscher, 2001). The relatively large value for mixing costs in young children and older adults was speculatively attributed to their greater difficulty in simultaneously maintaining two task sets. Given bilinguals’ apparent advantage in maintaining task set (Colzato et al., 2008), it should follow that they should also show reduced mixing costs. This result was indeed reported by Bialystok et al. (2006) in an experiment in which participants needed to respond on the same or opposite side as a target depending on a cue. Participants performed single-task runs in which only one cue was used and mixed runs in which either cue might appear. Response times to the target were slower under mixed conditions, and mixing costs were greater for monolingual participants.

Three other task-switching studies investigating monolingual and bilingual college students have yielded mixed results. First, Prior and MacWhinney (2010) asked participants to classify stimuli by color (red/green) or shape (circle/triangle). They found no mixing-cost advantage to bilinguals and no speed differences between the two groups on non-switch trials, but the bilinguals were faster than monolinguals on switch trials when instructions changed to sort on the alternate dimension. Thus, their study found a local switch-cost advantage to bilinguals with no mixing-cost advantage. Subsequent experiments replicated the switching advantage in bilinguals who reported that they frequently switched languages and no switching advantage in a less balanced group, although this less-balanced group exhibited significant associations between fluency in a nondominant language and switching and mixing costs (Prior & Gollan, 2010). These results suggest dissociations of switching and mixing costs with respect to group differences and imply that multiple aspects of bilingualism may influence task shifting. Frequent language switching may lead to task-switching advantages, whereas close monitoring of which language may be spoken when (and avoiding switching) may lead to task-mixing advantages. A third study provides clues with respect to the origin of the mixing advantage. In this study, Hernández, Martin, Barcelo, and Costa (2010) also used a color–shape switching task to test young adult Spanish-Catalan bilinguals and Spanish-speaking monolinguals. A rule was set at the beginning of a run (e.g., classify by shape), then trials continued for an unpredictable number without further cues until a second cue was presented. The second cue was either explicit (e.g., classify by color) or implicit (e.g., switch to the other rule or repeat the previous rule). It was found that switching was slower than repeating the same criterion but that this effect did not interact with group. Implicit cues were associated with slower response times than were explicit cues, and this effect did interact with language group; bilinguals were faster in the implicit version but not in the explicit version. The researchers also measured “restart costs”—slower RTs for the first trial than for the second trial after a repeat cue. Bilinguals had smaller costs than monolinguals on this measure too, but again only with implicit cues. These results suggest that the bilingual participants were better at maintaining the current set, monitoring the changing situation, and updating when necessary. Although the task was similar in many respects to that used by Prior and MacWhinney, the instructions were presented differently, and the bilinguals’ use of two very similar languages might account for the differences in results. If that is the case, one would need to be cautious about generalizing about differences in local and global task switching between monolinguals and bilinguals without considering further details of the participants and task situation.

There are still too few studies to conclude much that is definitive on the effect of bilingualism on task switching. Better bilingual performance for mixing costs (Bialystok et al., 2006) and dealing with implicit cues (Hernández et al., 2010) suggests that the advantage is in monitoring or set maintenance, but the results of the Prior and MacWhinney (2010) study speak more to the notion of greater mental flexibility or greater inhibitory control. In addition, bilingual language use may require different underlying control processes and may therefore lead to different processing advantages (Prior & Gollan, 2010). The few current studies involve many differences in methods and in participants, so the traditional cry of “more research is needed!” is very much the case before decisive conclusions can be drawn.

Bilingualism and memory

Since being bilingual necessarily entails the management and appropriate development of two language systems, it makes sense that these special skills of mental management should also apply to aspects of attention, conflict resolution, and cognitive control. But should bilingualism confer benefits on other cognitive functions—on memory, for example? The answer may depend substantially on the type of memory being investigated. Working memory (the manipulation of small amounts of material held briefly in mind) is generally considered to be either part of, or closely related to, executive processes, so bilingual advantages might be expected with such paradigms.
However, performance on semantic memory tasks (tapping stores of acquired knowledge) is likely to reflect experience with the type of information tested. Given that we have seen that bilingual vocabulary levels are typically lower than those of comparable monolinguals, we might expect that retrieval of verbal information would be poorer in bilingual participants, and, as described in the first section, performance on naming tasks and other tasks of lexical retrieval do in fact show this pattern. Moreover, performance on episodic memory tasks may again depend on the material in question.

For both working memory and episodic memory, the evidence is mixed. In one condition of the Simon task reported by Bialystok et al. (2004), color patches were presented centrally and so required no cognitive control, and participants responded to the color by pressing one of two response keys. In one version, two possible colors mapped to the two keys, and in the second version, four possible colors mapped to the two keys, with two colors associated with each key. The four-color version has greater demands on working memory, so working memory costs were taken as the difference between the two-color and the four-color versions. Bilingual participants aged 30 to 80 years showed smaller costs than did their monolingual counterparts, and were therefore deemed to show a bilingual advantage in working memory. This advantage has obvious similarities to the bilingual advantage in mixing costs found in some studies using the task-switching paradigm.

The results of other studies are less clear. Bialystok, Craik, and Luk (2008a) gave older and younger adult bilinguals and monolinguals two tests of working memory. The self-ordered pointing task requires participants to remember which of 12 abstract drawings have been selected previously; no language-group differences were found. The Corsi Block task is a test of short-term spatial memory, and in this case there was a bilingual advantage for younger but not older adults. Feng (2008) also presented various working memory tasks to monolingual and bilingual children and young adults. In the latter group, she found no bilingual advantage in either the Corsi Block task or in alpha span—a word-span task in which participants must mentally rearrange a short list of words from a presented order into alphabetic order. However, Feng did find a bilingual advantage for both children (Feng, Diamond, & Bialystok, 2007) and adults (Feng, 2008) in a test of spatial working memory in which items are presented in a random order in a 3 x 3 matrix (for children) or on a 5 x 5 matrix (for adults). The task is to recall the positions of the items in “matrix order”—that is, starting at the top left and progressing through the matrix left to right, line by line.

Whether or not there is a bilingual advantage in working memory may depend on the type of material used and the way in which working memory is tested. Working memory tasks may not be tapping one fixed cognitive mechanism but rather reflect a family of related functions generally concerned with holding and manipulating material that is in the focus of attention (Cowan, 1999) or simply “held in mind.” Tentatively, it seems to us that a bilingual advantage should be found in working memory, given the previously reviewed evidence suggesting that bilinguals have an advantage in set maintenance (e.g., Colzato et al., 2008) and in the related abilities of monitoring (Costa et al., 2009) and updating (Hernández et al., 2010).

The effects of bilingualism on episodic memory are also unclear at present, as only a few studies have been reported. In the studies described earlier, Fernandes, Craik, Bialystok, and Kreuger (2007) found poorer word recall by bilinguals, but Wodniecka et al. (2010) reported that the disadvantage was overcome when monitoring the list was required, as in the assessment of recollection. At present, therefore, there is little clear evidence for a bilingual advantage in episodic memory, some tentative suggestions of an advantage in working memory, and a clear disadvantage for bilinguals in the retrieval of items from semantic memory.

**The bilingual advantage across the lifespan**

Does the bilingual advantage in cognitive control change through the lifespan? It is well established that executive control functions first increase in effectiveness from childhood to young adulthood and then decline in the course of aging (Craik & Bialystok, 2006; Dempster, 1992; Diamond, 2002), so it seems possible that bilingualism might modify such functions and that the bilingual advantage might also show the same lifespan trajectory.

If the bilingual advantage in cognitive performance we have seen in this section is related to the enhancement of the executive control function, how early might we expect these differences to emerge given that the executive function system is late to develop? Similarly, if the cognitive advantage depends on protracted experience with two languages in which attention to systems and switching between them becomes practiced, could such advantages be found in children before they use language productively? A recent study by Kovacs and Mehler (2009b) provides dramatic evidence for the very early appearance of a bilingual advantage in 7-month-old infants. The infants who participated in the experiments were preverbal but were classified as bilingual if they had been exposed to two languages from birth because one parent consistently spoke to them in one language and the other parent used a different language. The researchers reported three experiments in which the infants learned to look for a visually rewarding puppet at one of two squares on a screen in response to either a speech stimulus (a trisyllabic nonsense word) or a visual pattern. After the learning phase, which was performed equally well by monolingual and bilingual infants, a new cue signaled the appearance of the visual reward in the alternate square. Thus, infants had to inhibit their first learned response and switch to a new response. The finding in all three experiments was that the bilingual infants learned to switch to the other square but the monolingual infants did not. The authors suggest that simply perceiving and processing utterances from the two languages during the first few months of life serves to accelerate the development of general executive functions that can then be applied in a variety of cognitive situations. This interesting
result does not negate the notion that some forms of the bilingual advantage are caused by inhibition of the nonused language but rather raises the interesting possibility that the advantage may have more than one causative mechanism.

What happens throughout life once bilingualism has modified these executive control systems? Does the bilingual advantage simply increase as the person accumulates experience dealing with two or more languages? And if bilingualism offers some protection against age-related cognitive decline (Bialystok, Craik, & Freedman, 2007; Kavé, Eyal, Shorek, & Cohen-Mansfield, 2008), does an increase in the bilingual advantage occur simply as a result of monolinguals showing a steeper decline in cognitive functioning than bilinguals do?

One problem with assessing these possibilities is that most studies deal with just one age group, so the opportunity to make lifespan developmental comparisons is limited. One exception is an article by Bialystok, Martin, and Viswanathan (2005) reporting studies on 5-year-olds and young, middle-aged, and older adults performing the same task, the Simon task. This series of studies showed a bilingual advantage (faster RTs) that was substantial in the 5-year-old children, virtually absent in 20-year-old undergraduate participants, but present again in groups of middle-aged (30–59) and older (60–80) adults. The authors suggested that the absence of an advantage in young adults may reflect the fact that cognitive control is most efficient at that time, so bilingualism provides no further boost. The two studies involving middle-aged and older adults were consistent in showing a larger bilingual advantage for the oldest (60–80) group, because the drop in efficiency from the middle-aged to older participants was greater for monolinguals than for bilinguals. This pattern of an especially strong advantage for the oldest bilingual participants was also found in three other studies by Bialystok and collaborators (Bialystok et al., 2004, 2006, 2008a; see Fig. 3b).

In the first two sections of this report, we reviewed behavioral studies of language and cognition, presenting the general finding that bilingual children and adults have smaller vocabularies and slower lexical access times than do their monolingual peers but that they also show enhanced cognitive control on a variety of tasks. What are the neural correlates of these effects? Is it possible to detect these subtle differences through neuroimaging techniques? In the next section, we survey the current evidence for structural and functional changes in the brain that result from bilingual experience.

3. Neural Bases of Language Control in Bilinguals

Whether one speaks just one language or more than one language, everyday use of language involves cognitive control. Bilingual speakers do not develop a separate control system; rather, as we have argued above, the use of two languages imposes on a single control system additional demands beyond those experienced by speakers of just one language. Our central claim is that this control system or network is used by both monolinguals and bilinguals but that the additional role in bilingual language processing modifies it, changing its performance for all tasks. In Section 2 we examined the cognitive consequences of such enhanced control. Here we make explicit the components of the network involved in language control, demonstrate how they also mediate the cognitive advantages shown by bilinguals, and explore the neural bases of control using many of the same tasks discussed in Section 2.

Figure 5 identifies the basic components of the control network, distinguishing it from the bilingual language system that it controls. The bilingual language system refers to a person’s mental representation of their languages; for present purposes, we leave this undifferentiated and focus on the components of the control system. A bilingual can perform different language tasks: He or she can choose to speak one language rather than another, can switch between languages, or can translate between them. Task schemas configure the bilingual language system so as to achieve the intended task, but these schemas are in competition to control the bilingual language system. Their activation must be monitored and, if necessary, adjusted by a higher-order executive process. For example, a bilingual must either maintain the current language in use if the circumstance demands (e.g., when speaking to a monolingual speaker of that language)—and so avoid inadvertently switching into the other language—or, on occasion, deliberately switch to the other language if the circumstance changes—for example, when a monolingual speaker of the other language enters the conversation.
of the task set, its “task schema”) must exceed and continue to exceed that of the competing task. Therefore, the speaker must monitor the speech output, and where marked slowing is detected or an error is noticed (i.e., saying “four” rather than “quatre”) the speaker must make some adjustments. The speaker might increase the activation of the required task (“name in French”) or suppress the activation of the alternative task (“name in English”)—as discussed Section 2 when we examined selection versus inhibition. Executive and monitoring processes are needed to establish new schemas (e.g., in the case of an experimental task) and invoke ones that are already part of a person’s repertoire. In this role, these processes work proactively; in response to performance difficulties, they work reactively (Green, 1998). A person may be conscious of the need to make such adjustments when an overt error is made, but on other occasions control adjustment may occur automatically, as in the way a thermostat adjusts power output in response to a deviation from the desired temperature (Green, 1998; Paradis, 2009; see Fernandez-Duque & Knight, 2008, for work suggesting that only conscious control leads to performance benefits across tasks).

What produces slower responses or overt naming errors? Marked slowing in naming in French, for example, may reflect successful inhibition of a strongly competing name in the other language (i.e., English), whereas naming in the wrong language indicates a failure of control. Activation of the English name may also increase the activation of the task schema for English and lead to increased competition with the task schema for French. Resolving such competition requires suppression of the English task schema. In other words, when a bilingual speaks two languages regularly, speaking in just one of these languages requires use of the control network to limit interference from the other language and to ensure the continued dominance of the intended language.

Would there be a difference in the switch cost if the bilingual were more fluent in French (L1) than in English (L2)? In that case, French would be the easier task and English the more difficult task, and the interesting finding is that it takes longer to switch into the easier task (143 milliseconds, ms) than it does to switch into the more difficult task (85 ms; Meuter & Allport, 1999). A plausible explanation for this seemingly paradoxical asymmetry of switch costs is that in order to name in English (the more difficult task), the easier task (naming in French) must be strongly inhibited, and it takes more time to reestablish the easier task, producing an asymmetry in the switching cost. Similar results were obtained in a study by Misra, Guo, Bobb, and Kroll (2007). Participants were asked to name pictures in L1 or L2 under either mixed conditions, when either L1 or L2 could be required, or in blocked conditions, when only one language was used. Their results showed that naming in L1 was slower under mixed conditions than it was under blocked conditions and that L1 naming was slower than L2 naming in the mixed conditions (an effect of reversed language dominance), supporting the interpretation that L1 was inhibited to permit the possibility of L2 naming. No asymmetry of switch costs is found when bilinguals switch languages voluntarily, yet a complete reversal of language dominance is found—again suggesting some form of inhibition of the L1 (Gollan & Ferriera, 2009).

Not all the research is consistent on this point. Finkbeiner, Almeida, Janssen, and Caramazza (2006) had bilingual participants name digits in either L1 or L2 and then perform a picture-naming task in their dominant language. Following the argument for greater inhibition of the dominant language, the hypothesis is that it should take longer to name pictures in L1 if the digit naming had been performed in L2. However, Finkbeiner et al. found no difference in picture-naming latency and so concluded that no inhibition of the nonused language took place. Their conclusion, though, is difficult to reconcile with evidence of global language inhibition identified in the later study by Philipp and Koch (2009). A more complete review of these issues is presented by Kroll et al. (2006).

The experimental research on bilingual task switching generally uses explicit cues to signal the language required on the current trial. Deliberate language switching in real life also requires a speaker to monitor the context for cues as to which language to speak (e.g., this person speaks L1 but not L2) and ensure correct language selection and suppression of any competing responses. Our premise, then, is that the additional demands on bilingual speakers relative to monolingual speakers entail greater use of this control network. The particular tasks that are subject to control are varied (e.g., naming pictures in one language, describing a scene in a second language, translating from one language to another). However, the components involved in monitoring performance and ensuring correct selection of the intended language task are applicable to other nonlanguage tasks, and, as we saw in the previous section, they appear to generalize to nonverbal tasks.

Neural bases of cognitive control

Figure 6 identifies the cortical and subcortical structures that are components of the cognitive control network in Figure 5. We follow others in separating the neural structures mediating control from those that process linguistic or other kinds of sensory or motor data (Posner & Petersen, 1990). The idea is that these cortical and subcortical structures work together to limit the effects of interference and to switch between tasks. For example, they may function as a control loop that continually monitors attention to the required task (e.g., Botvinick, Braver, Barch, Carter, & Cohen, 2001; Kerns et al., 2004). In its monitoring role, the anterior cingulate cortex may detect and help resolve interference (Lau, Rogers, & Passingham, 2006) and signal the prefrontal cortex, with its widespread connections to other regions (Dehaene & Changeux, 1991; Desimone & Duncan, 1995; Miller & Cohen, 2001), to alter the activation of the task schemas. Another region in the medial frontal cortex superior to the anterior cingulate cortex, the pre-supplementary motor area (pre-SMA), is also implicated in the control of action but seems linked more closely to spontaneously chosen actions than to response conflict (Lau et al., 2006).

The parietal cortex is involved in representing the task, through its connection to the prefrontal cortex, and in selecting
among competing responses, through its connection to the basal ganglia (Bunge et al., 2002). The basal ganglia are particularly important in task switching. Whereas traditional views (Alexander & Crutcher, 1990; Mink, 1996) emphasize the role of the basal ganglia in the control of movement, recent work emphasizes their key role in cognitive control too (e.g., Graybiel, 2000; Kotz, Schwartze & Schmidt-Kassow, 2009). Both cortical and subcortical structures are therefore important in understanding how interference is controlled and task switching achieved, so it is necessary to understand their role in language control. We shall examine the involvement of these regions in two broad categories of tasks: those requiring the control of interference and those based on switching between tasks and languages.

**The control of interference**

Using neuroimaging studies, we now consider the neural bases for controlling interference. These studies mostly rely on functional magnetic resonance imaging (fMRI) to assess the response of different neural structures when there is an increased demand to control interference. The basic data are the relative activation of different neural regions as detected by fMRI. A common assumption is that an increase in activation reflects an increase in difficulty. There is more extensive research on the control of interference in monolingual speakers, so our review makes use of meta-analyses of data from a number of studies.

**Interference control in monolinguals.** The argument being developed here is that bilinguals use the cognitive control network shown in Figure 6 to control interference from the competing language. Therefore, it is necessary to establish that these regions are recruited when monolinguals perform tasks involving response conflict. We consider work that has looked at the neural regions involved in controlling interference in three different tasks that, as described in Section 2, show a bilingual advantage: a nonverbal flanker task, a Simon task, and a Stroop task.

Although studies have examined these tasks separately, the strongest evidence for a common set of regions involved in cognitive control comes from studies testing two or more of them in the same individuals (Fan, Flombaum, McCandliss, Thomas, & Posner, 2003; Liu, Banich, Jacobson & Tanabe, 2004; Peterson et al., 2002). Fan et al. (2003) contrasted performance on a flanker task, a Simon task, and a manual version of the Stroop task in which individuals pressed one of four buttons corresponding to the font color of a presented word. In all the tasks, individuals responded faster in congruent trials than in incongruent trials. Fan et al. identified two regions that showed a common effect of conflict: one in the anterior cingulate cortex and one in the left prefrontal cortex (see Roberts & Hall, 2008, for a review).

Nee, Wager, and Jonides (2007) examined data from 47 papers using different interference tasks. Their review confirmed the importance of the left prefrontal cortex (dorsolateral...
region) and the anterior cingulate cortex, along with a region in the left posterior parietal cortex, in overcoming Stroop conflict. Neuropsychological data also support the importance of a frontal region in verbal control. Hamilton and Martin (2005) found that a patient with damage to a left inferior frontal region showed a large interference effect in the Stroop task but interference within the normal range for a spatial-conflict task. The analyses of Nee et al. also showed that different kinds of conflict induce slightly different patterns of neural response. In resolving conflict based on resisting responding to an infrequent stimulus, frontal and parietal regions in the right hemisphere, in addition to the left dorsolateral prefrontal cortex and the anterior cingulate cortex, were activated.

Much research on the control of interference has examined the role of cortical structures but ignored the role of subcortical structures, such as the caudate, that, as indicated in Figure 6, should be involved in selecting among competing responses. However, there is some relevant research on these subcortical structures. The left head of the caudate activates when a previously learned movement has to be inhibited (Shadmehr & Holcomb, 1999; Parsons, Harrington, & Rao, 2005) and when a prepotent response has to be blocked (Li, Yan, Sinha, & Lee, 2008). The caudate is also active in controlling interference in the Stroop task (Ali, Green, Kherif, Devlin, & Price, 2010). A reasonable supposition, then, is that the caudate is involved in the inhibition of plans of action and therefore controls both verbal and nonverbal types of interference.

To summarize, neuroimaging research with monolinguals confirms that a network involving the prefrontal cortex, anterior cingulate cortex, and caudate is recruited in tasks that require resolution of conflict from competing responses.

**Interference control in bilinguals.** We argue that bilinguals use this same network to control conflict from two languages. Therefore, if retrieving the name of a picture is effortful for bilinguals because of the need to overcome interference from the other language, then we would expect to find evidence for the involvement of this control network in a picture naming task. De Bleser et al. (2003) examined covert picture naming in an L1 and a later-acquired L2. (For technical reasons, neuroimaging studies sometimes adopt the expedient of asking participants to mouth picture names or to name pictures only covertly, so they will not move.) Participants were native speakers of Flemish/Dutch who had learned French from the age of 10. In one condition the picture names were cognates (i.e., the translation equivalents were phonological and orthographically similar), and in another condition they were noncognates. For pictures with noncognate names, naming in the L2 showed more activation in regions responsible for linking conceptual information and word form than did naming in the L1. The more important result is that activation increased in two inferior frontal regions associated with more effortful lexical and semantic retrieval. Therefore, data from this study, along with others (e.g., Abutalebi, Cappa, & Perani, 2001; Rodriguez-Fornells et al., 2005), suggest that naming in the L2 is associated with more effortful processing, an idea consistent with the involvement of cognitive control processes. Moreover, as proficiency in the L2 increases, the relative difference in activation between L1 and L2 decreases, again consistent with the idea that there is a decrease in effort (Abutalebi & Green, 2007).

Even early and highly proficient bilingual speakers show evidence of more effortful processing in their L2 and recruitment of control regions, despite demonstrating a processing profile that is similar to that of native speakers. Koveman, Baker, and Petitto (2008) asked Spanish-English bilinguals and English monolingual speakers to judge whether visually presented sentences were plausible or not. For the bilinguals, the sentences were presented in separate experimental blocks for each language. The English sentences (and their Spanish translations) varied in their syntactic complexity, being either subject–object relatives (e.g., “The child spilled the juice that stained the carpet”) or arguably more complex object–subject relatives (e.g., “The juice that the child spilled stained the carpet”). As expected, bilingual speakers showed a differential response to complexity as a function of the presented language. Spanish relies more on morphological marking than word order to signal grammatical relations. Like the English monolinguals, bilingual speakers showed increased left inferior frontal activation for the more complex English sentences. In contrast, they showed no differentiation as a function of complexity when processing the Spanish sentences. However, the study also showed that bilingual speakers processing English showed more activation in the left frontal region than monolingual English speakers did. In other words, processing even in a language in which they are highly fluent is more effortful for bilingual speakers and engages regions associated with cognitive control.

Increased proficiency in the L2 may also alter processing in the L1 precisely because of increased competition. In reading, the mappings between letters and sounds differ between languages, so the same string of letters can give rise to conflicting pronunciations. For example, what happens when native readers of Italian (which has a regular relationship between letters and sounds) read in their L1 after learning English, in which the relationship is irregular? As vocabulary knowledge in English increases, native Italian readers reading Italian show a linear increase in activation in a left frontal region associated with mapping letters to sounds (Nosarti, Mechelli, Green, & Price, 2010). Such an outcome indicates increased competition. More to the point, there is a linear increase in activation in a left frontal region used to resolve irregular pronunciations in monolingual native English readers. Interestingly, this region is also one that helps resolve lexical competition (e.g., de Zubicaray, McMahon, Eastburn, & Pringle, 2006). These data again suggest that bilingual speakers and readers, at least in contexts where both languages are active, experience increased verbal conflict and recruit a left frontal region to resolve it.

Other research allows us to see both cortical and subcortical regions involved in controlling interference. Van Heuven, Schriefers, Dijkstra, and Hagoort (2008) made use of a special relationship that exists between words in two languages such as
English and Dutch. Their participants were highly proficient Dutch-English university students who had learned English at the age of 10 to 12 years. Van Heuven et al. asked participants to decide whether a presented word was a real English word or not—an English lexical decision task. Some English words, termed interlingual homographs, are also real words in Dutch; for example, *room* means “cream” in Dutch. In an English lexical decision task, “room” elicits a competing “No” response because it is a word in Dutch, and in an English lexical decision task Dutch words should receive a “No” response. Relative to control words, therefore, correctly deciding that an interlingual homograph was a real English word elicited increased activation in three regions displayed in Figure 6: the left inferior prefrontal regions, the anterior cingulate cortex (together with another region we have noted previously in the medial frontal cortex, the pre-SMA) and the left caudate. As expected, van Heuven et al. observed no differential activation for interlingual homographs in a group of monolingual English speakers. This experiment left unresolved whether the activated regions were signaling conflict arising from the stimulus itself (i.e., “room” elicits two meanings in Dutch-English bilinguals) or conflict arising from ambiguities associated with the response (i.e., is “room” a word in English?).

To determine which regions responded to stimulus-based rather than response-based conflict, the researchers performed another experiment on a separate group of bilinguals from the same population. In this case, participants knew that some of the words might be Dutch words and responded “Yes” to each real English word regardless of whether it was also a Dutch word. In this case, interlingual homographs elicited increased activation only in left prefrontal regions, suggesting that the left prefrontal regions are sensitive to stimulus-based conflict. In contrast, the response profile of the anterior cingulate cortex (and the pre-SMA) and left caudate reveals regions that are either sensitive to, or help resolve, response-based conflict.

The precise impact of the other language might depend on how active it is. It is reasonable to expect that it will be most active when it is being used at the same time when bilinguals are in what Grosjean (1998) termed a bilingual mode and they are switching between languages. We consider the response of the control regions in the section on language switching.

**Task switching**

The second paradigm within which to examine the neural bases of cognitive control is task switching. Different types of data can help identify the structures recruited in switching between languages or between other types of tasks. Stroke damage to a specific structure can lead to difficulties in task performance and so provide evidence of its causal role in cognitive control that complements the data from neuroimaging studies. Again, we begin by establishing the neural basis of task switching in monolinguals and then compare those patterns to data from bilinguals performing task switching and language switching.

**Task switching in monolingual speakers.** The occurrence of a stroke is a tragic and dramatic event that helps to explicate the role of regions in Figure 6 for task switching. Greater damage to the left frontal cortex leads to increases in switch costs and so reflects difficulty in holding the current task in mind or in selecting the correct response, although inhibition of inappropriate tasks or associated responses might be more closely linked to the right frontal cortex (Aron, Monsell, Sahakian, & Robbins, 2004). The anterior cingulate cortex is less susceptible to stroke, but it is sometimes necessary to ablate part of it surgically. Postoperatively, such patients have difficulty responding to a cue that requires them to switch the direction in which they move a joystick (Williams, Bush, Rauch, Cosgrove, & Eskandar, 2004). Damage to the basal ganglia also severely impairs a person’s ability to switch between tasks and to overcome the interference from the prior task. We illustrate with a nonverbal task (Yehene, Meiran, & Soroker, 2008). Yehene et al. asked their patients to press one of two keys in response to the position of a target schematic face in a 2 × 2 matrix on the basis of one of two rules. In the top-down task they had to press Key 1 if the target was in the top half of the grid and Key 2 if it was in the bottom half. In the left–right task they pressed Key 1 if it was on the left side of the grid and Key 2 if it was on the right. The rule was cued on each trial. On critical trials, the correct response depended on the application of the correct rule, because Key 1 designated a target that was up or left and Key 2 designated a target on the bottom or right. Therefore, if a target was in the upper right cell of the grid, pressing Key 1 was correct for the top-down task but Key 2 was correct for the left-right task. Basal ganglia patients were severely impaired when the rule switched in this task, signaling the importance of that structure in such tasks.

In a meta-analysis of data from neuroimaging studies involving different types of task switches (e.g., rule switching, changes in target locations, and different response sets), Wager, Jonides, and Reading (2004) confirmed that the regions in Figure 6 are reliably activated on task-switch trials. The frontal cortex is sensitive to changes in demands involved in switching between tasks (Christoff & Gabrieli, 2000; MacDonald, Cohen, Stenger, & Carter, 2000) with more complex working memory demands associated with right frontal activation (Simmonds, Pekar, & Mostofsky, 2008). The anterior cingulate cortex is sensitive to changes in tasks and to errors consistent with its role in monitoring and in adaptive control in response to errors (e.g., Hyafil, Summerfield, & Koechlin, 2009). Parietal areas are also involved in remapping stimuli to response according to the new task (e.g., Corbetta & Shulman, 2002; Dosenbach et al., 2006). Finally, the basal ganglia play a role in shifting response. In a study in which participants tracked a continuous sine wave by controlling a cursor, activation increased in the left caudate when the current trial required a movement opposite to that used previously (Lungu, Binenstock, Pline, Yeaton, & Carey, 2007).

**Language switching in bilinguals.** In monolingual participants, the regions identified in Figure 6 were shown to contribute to task switching. Do they also contribute to language switching?
and interfere with linguistic behavior in bilingual patients? Patient reports indicate that damage to the prefrontal cortex, inferior parietal cortex, or basal ganglia structures affect the ability of bilingual patients to voluntarily switch from one language to another. As the anterior cingulate cortex is less susceptible to stroke, there are fewer reports for this structure, but all the other regions indicated in this control network show a clear role in language switching. Damage to either the left prefrontal lobe (Stengel & Zelmanowitz, 1933; Zatorre, 1989; Fabbro, Skrap, & Aglioni, 2000) or left inferior parietal lobe (Herschmann & Pötzl, 1920; Pötzl, 1925, 1930; Leischner, 1948/1983) can yield pathological switching, that is, unintended or inappropriate switching between languages. Lesions to the head of the caudate elicit either selective recovery of the current language, as if it is no longer possible to disengage from it (Aglioni & Fabbro, 1993; Aglioni, Beltramello, Girardi, & Fabbro, 1996), or pathological switching between languages (Abutalebi, Miozzo, & Cuppa, 2000; Mariën, Abutalebi, Engelborghs, & De Deyn, 2005). In the case reported by Abutalebi et al., A.H., a trilingual speaker of Armenian (L1), English (L2), and Italian (L3), was unable to avoid switching languages when naming simple pictures. For example, although he named the picture of a clock correctly in Armenian in an Armenian testing session, he named it in Italian in the English naming session and in English in the Italian naming session.

That the circuits underlying language switching are widespread is also indicated by data from transient cortical and subcortical electrical stimulation of the brain during surgery for treating glioma tumors or epileptic foci when the patient is awake. In the case of bilingual speakers, this stimulation can lead to involuntary switching from naming pictures in one language to naming them in another, reflecting the temporary disruption of control (Moritz-Gasser & Duffau, 2009a, b).

Neuroimaging studies of bilinguals without brain damage provide complementary data. In a study with early Spanish-English bilingual speakers, Hernandez et al. (2000) reported more activation in left dorsolateral prefrontal cortex when switching between naming pictures in English and Spanish than when naming pictures in just one language (see also Hernandez, 2009; Hernandez, Dapretto, & Bookheimer, 2001; Chee, Soon, & Ling Lee, 2003). Price, Green, and von Studnitz (1999) used single words and found that switching between languages increased activation in regions associated with phonological processing (a left inferior frontal region, Broca’s area, and parietal cortices). Taken together, these data indicate that language switching or mixing induce increased frontal and parietal activity consistent with the requirement to inhibit ongoing activity associated with one task and select a relevant response in the face of competition.

More recent research provides a fuller picture of the control regions involved in language switching. Abutalebi and colleagues (2008) studied German-French bilinguals who learned French relatively late (around 12 years of age) and were enrolled in a translation course. The task was to name pictures in their L1 under one of two conditions. In the single-language condition, a cue signaled whether they were to name the picture (e.g., “cup”) or generate an associated verb (e.g., “drink”). In the dual-language condition, the cue signaled whether they were to name the picture in their L1 or in their L2. In this dual condition, the nontarget language is very active. The key analysis is the contrast between naming a picture in L1 in the single-language condition and naming it in L1 in the dual-language condition. Abutalebi et al. found that naming pictures in the dual-language condition induced more extensive activation in the left prefrontal cortex, the anterior cingulate cortex, and the left caudate nucleus than did naming the same pictures in the single-language condition. Furthermore, the study confirmed more extensive activation in these regions when individuals were using their weaker L2. These results are strong support for the importance of these regions in selecting a language in the face of interference.

Other studies have used neuroimaging to examine the neural basis of the asymmetric cost in switching between a language in which one is more proficient and a language in which one is less proficient (Meuter & Alport, 1999). We illustrate this with a functional imaging study, but there is other work using evoked reaction potentials that is consistent with the idea that switching between languages involves a process of actively inhibiting the other language (Jackson, Swainson, Cunnington, Jackson, 2001) even if that does not invariably lead to an asymmetry in switching cost (e.g., Christoffels, Firk, & Schiller, 2007; Verhoef, Roelofs, & Chwilla, 2009, 2010). Wang, Xue, Chen, Xue, and Dong (2007) examined the cost of switching into L1 (Chinese) versus a newly acquired L2 (English). In line with the view presented here that the same regions are used for cognitive control and language control, Wang et al. reported increased activation in the regions associated with control when subjects switched into L2. The pattern again is consistent with the idea that bilinguals must inhibit their L1 to speak in their L2 when they are switching between the two languages. The persisting suppression delays naming time when individuals switch back into L1.

The effects of language switching have also been examined in comprehension, and, surprisingly perhaps, there is also good evidence for the involvement of control processes. Language switching elicits a left caudate response in late bilinguals (German-English/Japanese-English) when they make semantic decisions about the meanings of words (Crinion et al., 2006). The left caudate is also activated when bilinguals encounter a language switch while listening to a narrative and make no overt response at such a juncture (Abutalebi et al., 2007). The participants in this study were Italian-French bilinguals who had acquired French before the age of three and were living at the time of testing in an Italian community in Geneva, where French predominates. Switching elicited bilateral inferior frontal activity (along with activation in a language area). Most interestingly, a switch into the less-exposed language (Italian) elicited activation of the left caudate and the anterior cingulate cortex. Such a neural response indicates the need to distinguish between the processes responsible for implementing control from processes associated with overcoming the effects of such control. In the present case, switching into the less dominant
language in a comprehension task appears to demand more neural resources to overcome (suppress) the activation of the more exposed (dominant) language. In a production task, such a neural response may give rise, as we have seen, to slower naming when switching back into the more dominant language in order to overcome its earlier suppression.

Finally, a special type of language switching occurs when bilinguals translate from one language to another, and this task also involves the cortical and subcortical structures depicted in Figure 6. Price et al. (1999) reported that, in contrast to reading in different languages, translating activated mainly the anterior cingulate cortex and bilateral subcortical structures including the head of caudate. In that study, if participants did not know the translation equivalent they responded “No” or “Nein.” However, in other studies, left inferior frontal activation was found when that option was not available, both in single-word tasks (Klein, Milner, Zatorre, Meyer, & Evans, 1995) and in auditorily presented text translation by simultaneous interpreters (Rinne et al., 2000). Further, Rinne et al. (2000) reported that, since translation into the nonnative language is the more difficult task, left dorsolateral activation was more extensive when the interpreters translated into their nonnative language. The involvement of subcortical structures along with activity in the left prefrontal cortex is also reported (Lehtonen et al., 2005). Lehtonen and colleagues studied Finnish-Norwegian bilinguals who had learned Norwegian as adults (21–36 years). Participants completed a translation task and a control task. In the translation task, they silently translated visually presented Finnish sentences into Norwegian and then decided whether a presented Norwegian probe sentence was a correct translation of the Finnish sentence. In the control task, they silently read a Finnish sentence and determined whether a Finnish probe sentence was identical to it. The contrast between the translation and control task yielded substantial activation in the left (ventrolateral) prefrontal cortex and in a region of the basal ganglia (globus pallidus) that is activated in suppressing competing responses (Atallah, Frank, & O’Reilly, 2004; Ali et al., in press). Taken together, these data provide evidence for the involvement of the cortical and subcortical regions of the control network in a task special to bilinguals.

Local switching and mixing costs in bilingual and monolingual performance. The difference between local switch costs and mixing costs was discussed in Section 2, with most studies reporting smaller mixing costs for bilinguals and with more varied evidence for local switch costs. This distinction can also be examined using evidence from neuroimaging. From a control point of view, these two types of cost are interesting because local switch costs reflect transient control processes whereas mixing costs reflect the need for sustained control. In task-mixed blocks, individuals need to keep two tasks active and monitor the world for cues as to which one to perform.

Dosenbach et al. (2006) provide a detailed analysis of the regions involved in initiating a new task, sustaining it over a sequence of trials, and responding to error. They argue that the anterior cingulate cortex, together with another bilateral frontal region (the anterior insula/frontal operculum) form a core region for implementing and sustaining a new task. As yet, there are no comparable analyses for language switching in bilinguals, so we illustrate with evidence from two studies that compare local switching and mixing costs in language tasks.

Braver, Reynolds, and Donaldson (2003) asked participants to classify words according to whether they referred to objects that were natural versus created or whether the objects referred to were large versus small. Participants carried out these tasks in separate blocks of trials or mixed in the same block of trials. The anterior cingulate cortex and prefrontal regions of the right hemisphere were activated in the mixed blocks but showed no variation with local switching. In contrast, local switching was accompanied by activation in left prefrontal and parietal regions.

Wang, Kuhl, Chen, and Dong (2009) extended these ideas to language switching. Native speakers of Chinese who started learning English around 12 years and who rated themselves as being of low to moderate proficiency in English named digits silently either in single-language or mixed-language blocks. Language of response was signaled by a verbal cue presented 400 ms before the stimulus digit. Consistent with previous research (e.g., Meuter & Allport, 1999), it took longer to switch back into Chinese than to switch into English (43 ms vs. 8 ms). There was also a mixing cost that was similar for Chinese and English (but see Christoffels, Firk, & Schiller, 2007; Kroll et al., 2006, for data showing that an L1 can reveal greater mixing costs). Importantly, however, local switching and mixing costs were associated with different brain regions. For mixing costs, there was activation in bilateral prefrontal and frontal regions. Unlike other studies, Wang and colleagues reported no differential activation of the anterior cingulate cortex, a difference they attribute to the more automatic retrieval of numeral names. In contrast, and in line with the data of Braver and colleagues (2003), local switch costs activated left frontal regions (along with other cortical and subcortical regions). Based on an analysis of individual data, Wang and colleagues also proposed that a left parietal region plays a role in overcoming inhibition or in reactivating the previous language.

Bilingualism and the neural networks for control

We have summarized research showing the neural regions involved when individuals control interference in using one of their languages and the regions involved when they switch between languages. In both cases, the set of regions depicted in Figure 6 is activated. These data suggest extensive overlap with the regions mediating cognitive control when monolingual speakers resolve interference or switch between different tasks. Such a correspondence supports the proposal that the bilingual advantage in nonverbal interference tasks and in task switching arises from their use of neural regions recruited in language control.
We have relied on commonalities in the response of the control regions in bilingual and monolingual speakers faced with different tasks, but there may be subtle differences that are missed in such comparisons. It is important to have studies that directly compare bilingual and monolingual speakers (matched on confounding variables such as IQ and socioeconomic class) performing the same nonverbal conflict or switching task. One such study has identified differences between bilinguals and monolinguals (Bialystok, Craik, et al., 2005). The researchers contrasted two groups of early bilinguals (French-English and Cantonese-English) with a monolingual English group performing a Simon task. Bialystok and colleagues used magnetoencephalography (MEG) to identify the neural basis of processing differences between the language groups and analyzed two bands of signals: one associated with attentional control (theta band) and the other associated with signal processing (alpha band). The data indicated that there is a common network used by all participants but with with subtle differences in how interference is controlled. Faster responding in the bilingual groups was associated with more activation in the signal-processing band in two left frontal regions and the left anterior cingulate cortex, as distinct from the left middle frontal region associated with faster responding in monolingual speakers. It will be important to extend such research to other tasks.

Why, then, might bilinguals, at least those who use both languages on a regular basis and who acquired them early in life, show an advantage in overcoming interference and in task switching? The position that we have sought to establish is that it is due to the need to control linguistic interference with the corresponding demands to monitor and adapt behavior. Such control is required when individuals speak two languages. It may also be required when individuals use two sign languages but appears not to be important when individuals speak one language and sign in another. Consistent with this view, Emmorey, Luk, Pyers, and Bialystok (2008) found that speech-sign bilinguals responded comparably to monolinguals and did not show the advantage demonstrated by a group of speech-speech bilinguals on a flanker task; and Kovelman et al. (2009) confirmed that bilinguals who spoke one language and signed another showed no increase in prefrontal activation when they switched between the two, although they did show increased activation in language regions associated with mapping meaning to form.

As noted earlier, whether the source of the bilingual advantage is the voluntary or the involuntary nature of control is an open question, though it may prove to be the former (cf. Fernandez-Duque & Knight, 2008). But given that there is such an advantage, the control network in bilinguals may be more efficient overall, or bilinguals may adopt a more effective strategy in performing nonverbal tasks. For example, in interference tasks they might be better at maintaining the task goal and so reduce the impact of conflicting information. In task switching, they may respond more efficiently to a task cue and retrieve task goals more effectively. If this is the case, switching costs and demand on transient control processes would be reduced. Longitudinal studies will be important here, because it is known that older adults shift from a control strategy that is proactive and maintains task-relevant goals to one that is reactive and retrieves relevant information only when required (Jimura & Braver, 2010; Paxton, Barch, Racine, & Braver, 2008). The bilingual advantage shown in older adults may reflect their continued use of a proactive control strategy supported perhaps by left frontal structures and the anterior cingulate cortex.

Bilingual experience may also alter the capacity of the control network by altering the density of grey matter (i.e., the nerve cell bodies together with axons and dendrites) in one or more control regions (e.g., anterior cingulate cortex; caudate). It may even affect the white matter connections (i.e., the myelinated axons that connect regions of grey matter). Prior research indicates that cognitive, linguistic, and motor abilities can correlate with differences in brain structure, (e.g., Crinion et al., 2009; Draganski & May, 2008; Gaser & Schlaug, 2003; Lee et al., 2007; Maguire et al., 2000; Mechelli et al., 2004). If one or two regions show marked differences then this would constrain accounts of the neural basis of the bilingual advantage. Longitudinal studies are important, as they can rule out preexisting individual differences rather than bilingual experience as the source of the difference. In this context, studies of the aging brain (see Section 2) may prove particularly revealing, because age-related declines can be related to changes in specific brain structures. Our supposition is that deteriorating performance found in nonverbal-conflict tasks will also be found in tasks involving language control.

4. Implications of Bilingualism for Clinical Practice

The behavioral studies reviewed in Sections 1 and 2 reveal a number of differences between bilinguals and monolinguals in a variety of cognitive domains. These differences have proven to be useful for understanding the implications of bilingualism for cognitive development and cognitive aging. Moreover, the recent work in neuroimaging and related fields described in Section 3 is beginning to elucidate the neural correlates that underlie proficient language use. The question posed in the present section is whether these findings can be applied to help practitioners in the areas of neuropsychology, educational psychology, and speech/language pathology deal with the problems of bilingual clients and patients.

The challenge to professionals in these applied fields is that bilingual individuals vary enormously in their language skills. A few of the many factors that affect the degree of language proficiency in bilinguals are age and manner of acquisition of each language, degree of use of each language over a lifetime, and literacy and level of formal education in each language. It seems likely that these same factors will also affect the extent to which bilingualism modifies cognitive processing mechanisms. It is difficult to obtain a comprehensive assessment of all relevant factors in each individual case—yet such assessment is necessary to interpret test performance accurately. This uncertainty about the details of individual bilingualism combined
with the lack of tests developed specifically for use with bilinguals, the lack of knowledge about how bilingualism affects performance on standardized tests that were developed for monolinguals, and the strong emphasis on language-based assessment in clinical settings makes it difficult to answer some of the most common referral questions about bilinguals.

To simplify the following discussion, we assume that the bilingual individuals had early exposure to two languages and that English is the dominant language spoken by the majority of people in the environment. However, much of the discussion would apply equally well to proficient bilinguals who acquired one of their languages late in life, to bilinguals who live in bilingual communities in which one language is not clearly in the majority, and certainly to situations in which English is not the majority language.

Three general themes are common when bilingual individuals are referred to a clinician for intervention or therapy. Although the specific questions differ, these same themes are evident for children, adults, and aging bilinguals. The first theme is to establish whether there is a cognitive impairment or language impairment. In children, this question often takes the form of asking whether the child is learning English (the second language) as quickly as she or he should be, and if not, if there is a language impairment or more general developmental delay. For adults the concern is often linked to test results. As we saw in Section 1, tests of verbal fluency and naming generally reveal lower scores for bilinguals than for monolinguals, and these verbal scores are frequently lower than indicators of verbal memory or nonverbal functioning for bilinguals. In a clinical setting, this pattern raises the concern about the possibility of brain injury or developmental impairment—precisely what those tests were designed to diagnose—rather than the history of bilingual language use. For both children and adults, if language impairment is identified, there are inevitably questions about the best strategy for accommodating the impairment and for facilitating communication and recovery. For example, should treatment be provided in just one or in both languages? Would it be best to try to use primarily one language to ease the load on the compromised cognitive system by avoiding bilingualism (e.g., by switching to using only the majority language at home)?

A second theme is the need for advice on the best way to promote rapid acquisition of English as the person’s second language. For children the question is frequently framed in terms of educational options: Is total immersion in English best, or is it better to encourage parallel development of both languages by including both as part of the academic curriculum? In young adults, the concern is centered more on academic achievement, and questions attempt to determine the role of bilingual language use in academic outcomes. In middle-aged and older adults, the focus again shifts to learning the language. Some individuals are concerned about the length of time it is reasonable to live in a country without learning the environmental language.

The third theme is more specialized. Clinical intervention is sometimes sought to assess the adequacy of English proficiency for a specific purpose, such as functioning in school or in a professional setting. Adequate proficiency is also essential for safety and security, as in understanding the conversation in a medical interaction or discussing the risks of a medical procedure. Linguistic levels that may be perfectly adequate for some purposes may fail to support the ability to understand complex information for which careful thought and cautious decision making are required. These situations may also require the services of a clinician.

There are a number of reviews on cognitive and language assessment of bilinguals that provide useful information on the challenges that arise, on the kinds of questions to ask in clinical settings to obtain the necessary information to interpret bilingual performance on language-based tests, and on how bilingualism can affect performance on specific tests (e.g., Altrirri, Baker, 2000; Cummins, 2000; Kohnert, 2007; Paradis, 2008; Paradis & Libben, 1987; Peña & Bedore, 2009; Ponton & León-Carrion, 2001; Rivera-Mindt et al., 2008; Valdés & Figueroa, 1994). Here we attempt to connect questions about assessment of bilinguals more specifically with the experimental literature reviewed above.

Before considering how bilinguals differ from monolinguals in their performance on neuropsychological tests, it is helpful to review what typically happens during a cognitive assessment. Neuropsychologists receive referrals from parents, schools, and physicians, usually with a very specific question attached (e.g., Is there a language disability? Is the person beginning to show signs of early Alzheimer’s disease?). The neuropsychologist will subsequently review the patient’s academic record or medical chart and schedule an appointment to obtain a case history and administer cognitive tests. The general questions related to case history are the same for bilinguals and monolinguals: Were there any complications at birth? Was a learning disability ever suspected? What was academic performance like through school? What was the highest level of education attained? What is the education history? Were there any losses of consciousness? Is there any history of substance abuse or other psychiatric conditions? In some cases, there will also be a detailed language history for bilinguals, to determine which language is dominant, when and how both languages were learned, the extent to which both languages are currently being used, and other factors (e.g., Marian, Blumenfeld, & Krashansky, 2007).

Subsequently, the neuropsychologist will administer a series of tests to assess a variety of cognitive domains (e.g., mental status, IQ, language, memory, executive functions, and visuospatial skills), usually with heavier emphasis on tests that will be useful in answering the specific referral question. Often vocabulary tests are used to estimate verbal IQ, picture naming tests are used to identify the presence of cognitive impairment, and timed verbal fluency tests are given to look for frontal lobe pathology (Lezak, 1995). Verbal fluency performance is sometimes also used to look for patterns of performance that are associated with certain types of disease (e.g., deficits in semantic fluency are associated with Alzheimer’s disease whereas deficits in letter fluency are associated
with Huntington’s disease; Rohrer, Salmon, Wixted, & Paulsen, 1999). Assessment of bilinguals is complicated by the problem that bilingualism itself influences performance on these measures, and it is often not clear what adjustments should be made to interpret performance relative to that of monolinguals on the same tests.

**Assessing vocabulary knowledge in bilinguals**

A staple of neuropsychological testing is the assessment of vocabulary, but as we have seen in Section 1, bilinguals, especially bilingual children, often control a smaller vocabulary in each language than comparable monolinguals do, even in the absence of other compromising factors. How can clinical assessment make reliable judgments about the potential for a disability or disease in contrast to a normal outcome in the context of bilingual language use?

The approach taken to testing and interpretation often depends on the nature of the referral question. In some cases, relatively simple referral questions that can be successfully addressed without much knowledge about bilingualism arise. For example, parents may wonder how their child’s English vocabulary knowledge compares to that of his or her monolingual peers (note that in bilingual societies this question may be less relevant, particularly if monolinguals are few in number). In such cases, it is obviously appropriate to administer a test that was developed for use with monolingual English-speaking children, and the score obtained will provide a valid answer to the question being asked. However, the possibility of interpreting that same test score will not extend beyond the answer to this one simple question. As a group, bilingual children who speak a minority language at home (e.g., a non-English language in an English-speaking environment) will obtain lower receptive English vocabulary scores than will monolinguals, even if their parents report that they are “proficient speakers of English” (Bialystok, Luk, et al., 2010). These lower English vocabulary scores may be found even in children without much proficiency in the minority home language if the parents are not native speakers of English, because such children have reduced exposure to English vocabulary at home, at least compared to children whose parents are native English speakers and use English exclusively.

The difference in vocabulary size in bilinguals is probably a better reflection of experience than of ability to learn. In 6-year-olds, the vocabulary deficit associated with bilingualism seemed to be restricted to test items classified as “unlikely to occur in a classroom context” (Bialystok, Luk, et al., 2010). Similar results may be obtained in older bilingual children and in bilingual adults and, if so, such information could ultimately be useful for developing vocabulary tests that cater to specific profiles of bilingual language exposure. In addition, item analyses may be useful for interpreting individual test scores. For example, if a bilingual child misses a home-context item (e.g., “toaster”) it may simply mean that there have been no opportunities to learn this word in English because it is unlikely to come up in a school context.

Although a group of bilinguals will, on average, score lower than a group of monolinguals, individual scores will not necessarily be lower. In the large-scale study of “fluent English-speaking” bilingual children between the ages of 3 and 10 years (Bialystok, Luk, et al., 2010), the distributions of bilingual and monolingual scores overlapped much more than not. This means that although the average bilingual score was about 10 standard score points (2/3 of a standard deviation) lower than the average monolingual score, only a small number of bilinguals scored completely outside the range of performance for monolinguals. Thus, the majority of bilingual children described as “fluent in English” will obtain “normal” scores on tests developed for monolinguals. However, it is also probable that these same normal scores will fail to provide an accurate representation of learning potential.

Vocabulary scores reflect the combined forces of the ability to learn new vocabulary and the opportunities to learn new vocabulary. Bilinguals who score within the average range for monolinguals may have better-than-average ability to learn, which has allowed them to achieve an average monolingual score despite having fewer learning opportunities. An important consideration in such cases is that comparisons between monolingual and bilingual children with matched vocabulary scores may be invalid because bilingual children with monolingual-like vocabulary scores may be precocious learners. Conversely, bilinguals whose vocabulary scores fall 2 standard deviations below the monolingual average could be learning disabled, or they may simply have had less opportunity to learn English than their case histories suggest—two conclusions with very different implications but with equally serious consequences. Bilinguals who score below average may be inaccurately diagnosed with impairment when none is present, or could be diagnosed as “normal for a bilingual” even though impairment is in fact present and treatment is needed. The less-frequent cases in which bilinguals obtain scores that are higher than the case histories suggest—again, two conclusions with very different implications. Much of this discussion likely applies as well to bilingual adults, who also typically obtain lower vocabulary scores than do monolingual adults (e.g., Bialystok et al., 2008a; Portocarrero et al., 2007).

This discussion demonstrates the tremendous challenge in interpreting individual test scores in bilinguals. Even with the availability of normative data about bilingual performance on a given test, several factors continue to complicate interpretation. Further difficulty arises if one considers a broader range of bilinguals at different proficiency levels. The previous discussion applies only to children who are judged by their parents to be “fluent in English.” Such children can reasonably be tested in English (and specifically should be tested in English if English is their dominant language). However, even in such cases, a more accurate estimation of language skills will emerge if both languages are tested. Parents may sometimes overestimate the degree of majority-language fluency that their children have achieved. Bilinguals who are not dominant in
English must be tested in their dominant language, but often tests for those languages have not been developed, and there are virtually no tests for different combinations of bilingual types. One exception that is available in many different language combinations is the Bilingual Aphasia Test (the BAT; Paradis & Libben, 1987). However, the BAT was designed to assess fluent adult bilinguals for possible language impairment (i.e., aphasia), and it is not known how bilingual children should perform on this test or even if the test is useful in assessing bilingual adults who don’t have a high degrees of fluency in their two languages.

Finally, these simpler cases of “relatively fluent-in-English bilinguals” are perhaps least likely to present for referral in a clinic because they have already been successful in attaining second-language fluency. A more typical presentation will be someone who seems to be having trouble acquiring second-language fluency. Parents of young preschool children may suspect a problem if their child seems to be avoiding English speakers in the classroom, preferring instead to socialize only with the small number of other children who happen to speak the same minority language at home. Parents of older school-aged children may become concerned about low academic test scores or large discrepancies between verbal (e.g., reading/writing) and less-verbal (e.g., math) academic domains. (Here, “less verbal” is meant to emphasize that all academic subjects require at least some verbal skills; for example, math problems sometimes come in paragraph format or require ability to read instructions.) In such referral cases, it is necessary to assess what the opportunities to learn English have actually been—sometimes children have actually had less exposure to English than is assumed—and whether or not normal amounts of learning have taken place given those opportunities. Even with adequate assessments of opportunities to learn, test interpretation is difficult because little to no information about exactly how much exposure is needed to perform within a particular range on any given test is available to clinicians.

A creative approach around these problems has been to provide a learning opportunity during the assessment session itself and then to determine how much learning takes place, an approach sometimes called Dynamic Assessment (Gutiérrez-Clellen & Peña, 2001; Peña, Iglesias, & Lidz, 2001). This approach is based on interaction between the clinician and the child. Three types of dynamic assessment are (a) “testing the limits,” in which feedback is provided and errors pursued through further questioning; (b) “graduated prompting,” in which the level of contextual support is manipulated; and (c) “test-teach-retest,” in which alternative versions of tests of the same material are repeated after teaching to areas of weakness, in order to assess learning (Gutiérrez-Clellen & Peña, 2001). With these methods the amount of exposure is controlled—it is provided during the testing session itself. Children who fail to learn (i.e., do not show significant improvement on “measures of modifiability”; Peña, Resendiz, & Gillam, 2007) are flagged, with a high rate of accuracy, as probable cases of developmental delay. Such techniques are extremely useful for bilinguals and monolinguals alike, and they provide a means for obtaining accurate assessments with less concern about how to interpret past opportunities to learn.

In theory, bilingual disadvantages in vocabulary knowledge should decrease with age as their time to learn words in both languages increases. Although vocabulary knowledge continues to increase well into older age (Verhaeghen, 2003), new words may be learned at a faster rate before knowledge reaches a particular point (perhaps a typical adult-vocabulary repertoire). In other words, bilinguals should “catch up” to monolinguals as years of immersion in English accumulate. One way to test whether this is indeed the case is to ask whether the vocabulary deficit associated with bilingualism decreases in children as they progress through school and beyond that across the life-span. Indeed there has been some suggestion that bilingual children achieve monolingual-like vocabulary scores with increased time in school (Hamers & Blanc, 2000). However, the “catching up” notion is best tested with a longitudinal design, and to our knowledge such studies have not been reported. Moreover, bilinguals may appear to be catching up only because the test materials are not difficult enough to reveal persistent differences between bilinguals and monolinguals. When tested exclusively for their knowledge of very-low-frequency words in the relatively dominant language, for example in studies of the tip-of-the-tongue phenomenon, adult bilinguals consistently report recognizing fewer of the targeted vocabulary words than monolinguals do (e.g., Gollan & Silverberg, 2001; Gollan & Brown, 2006). Tip-of-the-tongue experiences are retrieval failures in which partial phonological information is available; they generally occur for low-frequency words but appear to be more broadly based for bilinguals. Thus, differences between bilinguals and monolinguals in opportunities to learn vocabulary will be less apparent in settings that only require knowledge of relatively easy, frequently occurring words than they will be in settings that require knowledge of difficult, low-frequency words (Gollan et al., 2008). This may be because, by virtue of using each language only part of the time, bilinguals will have had relatively less exposure to words in each language than will monolinguals (the weaker-links hypothesis described in Section 1), although they will have had sufficient exposure to learn frequently encountered words.

Confrontation naming

Confrontation naming is a testing method in which pictures are presented to participants, who are asked to name them as rapidly as possible. One of the most commonly used such neuropsychological tests is the Boston Naming Test (BNT; Kaplan et al., 1983). This test contains 60 black-and-white line drawings that show a single object that speakers try to name. The pictures are easy at the beginning of the test (e.g., a bed) but become progressively more difficult, ending with uncommon objects encountered in limited contexts. The ability to name pictures is sensitive to changes in cognitive functioning and is therefore useful for detecting subtle brain injuries.
language, then older bilinguals should have the dominant language during retrieval of the nondominant cognitive control. If the frontal lobes (Raz, 2000; West, 1996) and unexpected from the perspective of bilingualism as an exercise in some ways “better bilinguals” than younger bilinguals may seem low-frequency words. The finding that older bilinguals are in allows for the accumulation of experience to deal with (for review see Gollan & Brown, 2006). It may be that aging deficits in bilinguals could arise for the same reasons as receptive vocabulary deficits—namely, less frequency of use of specific words than for monolinguals. Alternatively, it may be because of dual-language activation—that is, the need to select one language in the face of competition from the other one. It is also possible that both factors may be operating. Some of the burden associated with bilingualism seems to be better managed with increased age—a result that is consistent with the notion of a frequency lag for bilinguals. In one picture-naming study, older bilinguals were relatively faster to produce low-frequency picture names in a nondominant language than would be expected based on their otherwise relatively slowly naming times relative to proficiency-matched young bilinguals (Gollan et al., 2008). Because low-frequency words in the nondominant language will be most vulnerable to the frequency-of-use lag associated with bilingualism, these words are also most likely to benefit from the increased exposure to language associated with age.

The age-related advantage for producing low-frequency words is also evident in studies comparing older to younger monolingual speakers: Like older bilinguals, older monolinguals consistently produce names for pictures with very low-frequency words with greater success than matched young monolinguals (for review see Gollan & Brown, 2006). It may be that aging allows for the accumulation of experience to deal with low-frequency words. The finding that older bilinguals are in some ways “better bilinguals” than younger bilinguals may seem unexpected from the perspective of bilingualism as an exercise in cognitive control. If the frontal lobes (Raz, 2000; West, 1996) and executive control decline in older age and are needed to suppress the dominant language during retrieval of the nondominant language, then older bilinguals should have more difficulty than young bilinguals in producing low-frequency words in the nondominant language. It might be asked whether older bilinguals perform better because the low-frequency words are archaic words more familiar to older than to younger participants. However, controlled studies select materials that are highly familiar to both young and old adults, and in the timed picture-naming study with bilinguals, the low-frequency targets were all highly familiar and current (e.g., crutches, a whistle, a scarf, a dustpan; see appendix in Gollan et al., 2008). Most importantly, the relative age-related advantage appeared only in the nondominant language, whereas the same concepts and words did not demonstrate any age-related advantage in the dominant language (or in monolinguals). Thus, it seems that accumulated use over a lifetime has its greatest influence on the very lowest-frequency words, thereby offsetting some aging-related deficits in retrieval.

A number of factors have been shown to reduce or even eliminate the bilingual disadvantage in picture naming, and this raises the question of what would be the best way to adjust tests of picture naming to accommodate bilingual ability and enable clinicians to perform reliable assessments. The answer to this question may vary with the referral question, and the implications of these findings for diagnosis and treatment of bilinguals are not yet established. For example, bilinguals name pictures more quickly (Costa, Caramazza, & Sebastián-Gallés, 2000; Hoshino & Kroll, 2008) and, in some cases, with no disadvantage relative to monolinguals (Gollan & Acenas, 2004) if the test consists of pictures with cognate names. Cognates reduce bilingual disadvantages via joint activation of target phonemes (sounds) through separate lexical representations in each language (for a review, see Costa, Santesteban, & Caño, 2005; for research showing increased activation for cognates, see Broersma & de Bot, 2006). To illustrate, the lexical repre- sentations of lemon and its Spanish translation limón activate many shared sounds, but grape and its translation uva activate no shared sounds. A similar reduction in bilingual disadvantage may be obtained by asking participants to retrieve names of people (Gollan, Bonanni, & Montoya, 2005). Bilinguals’ relative ease at producing proper names may have a different mechanism from cognate effects; bilinguals may effectively be monolingual for proper-name production because proper names are generally shared between languages (e.g., Golda Meir is basically the same in Hebrew, English, Spanish, etc).

The finding that bilinguals are better able to name pictures with cognate names could be useful clinically. One possibility is that bilingual picture-naming tests should focus on cognates (or proper names) for which bilinguals perform much like monolinguals. However, removing the disadvantage may compromise a test as an assessment instrument. For example, the presence of cognate effects on dominant-language production implies the presence of dual-language activation even when bilinguals are tested exclusively in their relatively more dominant language. Thus, a possible problem with using cognates is that cognates may increase the extent to which both languages are active, and this may have other undesired effects on test performance (note that cognate-facilitation effects have also been found in bilingual children, but this literature has focused primarily on receptive vocabulary rather than on picture naming; August, Carlo, Dressler, & Snow, 2005; Mendez Perez, Peña, & Bedore, in press).

Similar considerations apply to another way to reduce bilingual disadvantages in a testing or assessment situation: to allow bilinguals to use either language to name pictures (Kohnert, Hernandez, & Bates, 1998; Gollan & Silverberg, 2001). This approach is sometimes called “composite” or “conceptual” scoring. The scoring method improves
Additional studies are needed to determine the relations between language might be magnified in cognitively impaired bilinguals. Alternatively, cognate effects in the nondominant possible that cognate effects in the dominant language occur degree of proficiency in the L3 is high enough. Therefore, it is knowledge because even an L3 can have an effect on L1 if the speed in the dominant language. The clinical significance is level of proficiency even in an L3 can influence processing disadvantages in both the dominant and the nondominant language switching, whereas they have considerable difficulty with cued language switching (Hernandez & Kohnert, 1999). Thus, although language mixing might allow bilinguals to communicate better in natural settings, it is not necessarily the case that allowing language mixing and switching in a clinical setting will lead to more effective diagnosis and treatment, because the either-language scoring method may actually obscure differences between patients and controls (Gollan et al., 2010), which is counterproductive if the goal is to identify impairments in bilinguals. In bilingual language assessment, the costs associated with language switching and mixing can be avoided by testing each language in a separate testing block.

The opposite outcome may be found for cognates. It may be, for example, that language-impaired bilingual children are less able to benefit from cognate manipulations than typically developing bilingual children are. If this is so, then the ability to benefit from cognate status itself could function as a kind of bilingual-specific litmus test for cognitive impairment. In other words, failure to demonstrate improved lexical access for cognate words relative to typically developing bilingual children would signal some type of language impairment. Importantly, however, it is necessary to consider the relative dominance of the two languages for the bilingual child and the relation between that dominance and the language of assessment. In relatively balanced bilinguals, cognates can reduce bilingual disadvantages in both the dominant and the nondonominant languages (Gollan & Acenas, 2004; Gollan et al., 2007), but such reductions are most robust when bilinguals are tested in their nondonominant language (e.g., Costa et al., 2000; Gollan et al., 2007). A study by van Hell and Dijkstra (2002) using lexical-decision and word-association tasks showed that high level of proficiency even in an L3 can influence processing speed in the dominant language. The clinical significance is that it is not possible to discount nondonominant language knowledge because even an L3 can have an effect on L1 if the degree of proficiency in the L3 is high enough. Therefore, it is possible that cognate effects in the dominant language occur only in relatively balanced bilinguals who are also cognitively intact. Alternatively, cognate effects in the nondonominant language might be magnified in cognitively impaired bilinguals. Additional studies are needed to determine the relations between cognate effects on the one hand, and language and cognitive assessment of bilingual children on the other.

**Verbal fluency in clinical practice**

Research using the verbal fluency test as an experimental tool was described in Section 1. The results showed consistent bilingual disadvantages on semantic fluency (except when receptive vocabulary knowledge is matched), with somewhat less severe or less certain disadvantages on letter fluency. Clinically, the greater bilingual disadvantage in semantic fluency than in letter fluency can be misleading, because this is the same pattern of fluency performance that is found in monolinguals with early Alzheimer’s disease as compared with normals (Butters, Granholm, Salmon, Grant, & Wolfe, 1987). This creates a dilemma for neuropsychologists: Is an individual showing signs of early Alzheimer’s disease or is she simply showing the effects of bilingualism on fluency? The verbal fluency test is an important instrument in the battery to assess patients for cognitive decline, so the ambiguity of the results obtained from bilinguals presents a clinical problem. To develop fluency tests for bilingual speakers, it is necessary to understand why semantic fluency is more affected by bilingualism than letter fluency is. As we explained earlier, letter fluency requires greater recruitment of executive control, perhaps offsetting bilinguals’ disadvantages in lexical retrieval.

A different approach to assessing older bilinguals is to use a task related to verbal fluency, one that reflects semantic processing yet distinguishes the cognitive mechanisms that underlie the effects of bilingualism from those that are involved in Alzheimer’s disease. In the semantic-association task (de Groot, 1989), speakers are given a cue (e.g., “bride”) and are asked to produce the first response that comes to mind in relation to the cue. The overwhelming majority of responses in this task are semantically related to the cues, and this is true for all speakers, whether they are monolingual or bilingual and whether or not they are cognitively impaired. However, bilinguals produce slightly but significantly different (or “less typical”) responses than are normally found in monolinguals. For example, given the cue “bride,” they might say “pretty” instead of the more typical “groom” (Antón-Méndez & Gollan, in press). A similar effect was reported in monolinguals with Alzheimer’s disease as compared to cognitively healthy controls (Gollan, Salmon, & Paxton, 2006). To this point, therefore, there is the same interpretation problem as there is for verbal fluency, because both bilingualism and Alzheimer’s disease produce the same outcome. However, further experiments with the semantic-association task demonstrated that only the bilingual effect is modulated by lexical frequency. Bilinguals produced the same associations as monolinguals do when the cues were strongly associated to high-frequency words. In contrast, speakers with Alzheimer’s disease produced atypical responses regardless of associate frequency (Antón-Méndez & Gollan, in press). This evidence is consistent with the notion that Alzheimer’s disease impairs semantic representations themselves (Butters, Salmon, & Heindel,
1990), whereas in bilinguals, difficulty with lexical access can sometimes leads them to perform in ways that imply semantic deficits when none are present.

As with confrontation naming, there is an important role for cognate status in the performance of verbal-fluency tests, so the interpretation of results, especially for clinical assessment, needs to account for this factor. Specifically, in both semantic and letter fluency, bilinguals who speak languages with many cognates spontaneously produce as many cognate responses (e.g., “lemon”) as monolinguals do but fewer responses for words that are not cognates (Sandoval et al., 2010). Put another way, words that are cognates across the two languages are generated as often by bilinguals as they are by monolinguals who only know them in one language, but unique words are produced less often by bilinguals. In this sense, the greatest difference in performance is in the lower production of noncognate words by bilinguals, who appear to have easier access to words that occur in both their languages. These findings suggest that bilinguals who speak languages with an extremely high proportion of cognates (e.g., Catalan-Spanish bilinguals) may exhibit no fluency disadvantage, even for semantic fluency.

Another similarity between verbal fluency and confrontation picture naming is that bilinguals retrieve a greater number of concept names if they are tested in both languages (Bedore, Peña, García, & Cortez, 2005). However, unlike picture naming in the BNT, fluency scores do not increase if bilinguals are allowed to use whichever language comes to mind during a single trial and so to switch between languages (Gollan et al., 2002; De Picciotto & Friedland, 2001). The lack of an improvement in fluency scores when both languages are used may reflect the costs of language switching. The timing allowed to name each picture in the BNT, about 6 seconds, is too long to detect the millisecond cost of language switching, so on this task no switching costs are reported. Presumably, on a more tightly timed picture-naming task allowing responses in either language, bilinguals would name fewer pictures than would monolinguals in a fixed amount of time (e.g., 60 seconds), because of the additional time needed to carry out the language switch (Gollan & Ferreira, 2009).

Because bilingualism affects verbal fluency in a number of interesting ways, there are various possibilities for reducing the bilingual fluency disadvantage. However, reducing this disadvantage may compromise the reliability of the instrument as an assessment tool for bilinguals, so it is not clear what combination of fluency tests would be most useful for diagnosis of cognitive impairment in bilinguals. Minimally, interpretation of the test scores needs to be modified to accommodate the systematic differences that accompany bilingual performance, but ultimately it may be possible to develop fluency tests that are specifically targeted to a bilingual population.

The assessment of executive functions

Many of the linguistic skills that bilinguals generally perform more poorly than monolinguals (reviewed in Section 1) are included in typical assessment batteries, often using the same instruments as those used in research. Therefore, understanding how to interpret bilingual performance on those tests is a crucial concern for neuropsychologists. However, in Section 2 we described a variety of nonverbal cognitive tasks on which bilinguals generally perform better than monolinguals. These tasks were measures of executive control and, as we have argued, the experience of bilingual language use has the beneficial outcome of enhancing these levels. What are the clinical implications of this advantage?

The implications of this bilingual advantage for clinical assessment are more limited than the bilingual disadvantage in lexical retrieval for several reasons. Perhaps most important is the great emphasis on verbal skills in clinical assessments, with a more minor role for nonverbal cognitive performance. Therefore, the bilingual advantages found in nonlinguistic tasks will have relatively little effect on the cognitive profiles generated in clinical settings. Another important point is that many of the tasks showing bilingual advantages in experimental studies (e.g., the Simon task and the Attentional Network Task) are not used in clinical settings.

An important exception is the Stroop color-word-naming task, which is commonly used to measure attention and is diagnostic of a variety of conditions associated with cognitive impairment (e.g., Lezak, 1995). As we have seen earlier, bilinguals generally suffer less Stroop interference and greater Stroop facilitation than monolinguals do (Bialystok et al., 2008a; Hernández et al., 2010). Several considerations make it difficult to interpret these differences, however. For example, performance on the Stroop is affected by language proficiency (Tzelgov, Henik, & Leiser, 1990; Rosselli et al., 2002). Because of this, it is possible that only highly proficient bilinguals will exhibit the advantage in their dominant language and that disadvantages may be found if bilinguals are tested in a less dominant language. Equally, it may be that a smaller Stroop effect would be found for less-proficient bilinguals, since the meaning of the color word would be less automatically activated and therefore less interfering. However, Bialystok et al. (2008a) considered that possibility and divided each of the monolingual and bilingual groups into subgroups based on the speed with which they read the name of the color word when it was written in black ink. The idea was that faster reading times should lead to more interference and therefore a larger Stroop effect. Therefore, comparing the fast bilingual readers with the slow monolingual readers should reduce the size of the Stroop effect, possibly reversing the direction. Nonetheless, the analysis showed that bilinguals continued to record a smaller Stroop interference effect than did monolinguals, even when considering only the bilinguals for whom reading the English words was the most automatic.

The facilitation effects found for bilinguals in the Stroop task might be interpreted as a bilingual disadvantage. Increased facilitation effects have been found in monolinguals with Alzheimer’s disease when compared with healthy controls (Sieler, Balota, & Faust, 1996) and in children when compared with adults (Wright & Wanley, 2003). The disadvantage view of facilitation is that these effects indicate increased
inadvertent focus of attention on the word during color naming (MacLeod & MacDonald, 2000; Spieler et al., 1996). Note, however, that the version of the Stroop task used in experimental research is not exactly the same as the version used in the clinic. For example, experimental studies typically use raw interference scores whereas clinic assessment relies on a speed-adjusted interference score. Similarly, congruent trials are typically not administered in clinical settings. Therefore, more information about precisely what types of bilinguals exhibit a Stroop advantage, the origin of bilingual effects on the Stroop task, and perhaps most importantly the distribution of scores is needed.

5. Bilingualism in the World

The constant use of two languages is an experience that leaves its mark far beyond the immediate and obvious domain of communication. As we have seen in this review, it modifies the level to which some features of linguistic systems may be learned and the way in which they are used; it enhances aspects of cognitive processing, particularly those involved in the executive control system; it recruits, and most likely adapts, the neural networks involved in the control of nonverbal processes to modify their use for verbal processes; and it intervenes in clinical assessment by presenting a profile that may not be accurately captured by monolingual norms. These are significant consequences that cover individual (e.g., cognitive development and decline) and public (e.g., assessment and dementia) outcomes. Given this context, the questions posed in this final section concern the implications of bilingualism for public policy decisions, especially perhaps in the areas of education and health care. The current prevalence (and rapid growth) of bilingualism in today’s highly interconnected world make these questions relevant and urgent. In light of the dramatic numbers noted in the Introduction, we conclude by addressing specific questions about bilingualism that concern both individual and social issues.

Bilingual education

Not all parents have the opportunity to expose their children to a second language at home, yet many understand the value of being able to communicate in another language. One option in these cases is to find alternatives in formal education. A popular program in this regard is immersion education. In these programs, school instruction takes place in a language that is not the language of the home or the community (e.g., French instruction in English Canada, Spanish instruction in the United States) and children are expected to use this language in all their communication with teachers and friends while at school. Therefore, children develop fairly high competence in this language, even though they do not typically achieve the level of a native speaker (for review, see Genesee, 1985; Johnson & Swain, 1997). But does this limited school exposure make these children “bilingual” by the criteria used in this review and, therefore, affected by the cognitive and linguistic outcomes we have described?

The question can be cast more broadly as an inquiry regarding the degree of bilingualism necessary for the outcomes observed for more fully functioning bilinguals. There is little evidence on this point, but the available studies suggest that there is a correlation between the degree of bilingualism and the extent of the impact of bilingualism on cognitive and linguistic processing. Early studies with children in French immersion programs showed that both metalinguistic (Bialystok, 1988) and cognitive (Bialystok & Majumder, 1998) outcomes for these children were between those found for monolingual children and those found for bilingual children who were fully fluent in both languages. More generally, Luk (2008) compared 120 bilingual adults with varying degrees of bilingualism to a group of 40 monolinguals on linguistic and cognitive outcomes and again found larger effects to be associated with greater degree of bilingualism.

Extending this pattern to education, it is reasonable to assume that there is a cumulative effect of learning language that, at least in the intense environment of immersion programs, confers some of the cognitive advantages on children even if they do not become highly fluent speakers. Importantly, there are few if any costs of immersion education for most children, although individual cases may present special challenges that need to be considered.

More languages, more benefits?

Bilingualism, as we have explained, leads to specific benefits in cognitive processing, and even the limited bilingualism that comes from immersion education produces some minimal form of this effect. By the same logic, then, does trilingualism lead to even greater benefits than bilingualism, acting as something like super-bilingualism? The evidence on this point is scant. An interesting study by Kave et al. (2008) compared general cognitive level in a large sample of older adults living in Israel as a function of how many languages they spoke (there were no monolinguals in the group). They reported significantly higher maintenance of cognitive status in older age in trilinguals than in bilinguals, and even greater maintenance by multilinguals who spoke four or more languages than by trilinguals, although the measure of cognitive level used was not very precise. Similarly, others have reported later age of onset of Alzheimer’s disease in multilinguals as compared with bi- and trilinguals, as we will describe (Chertkow et al., 2010). However, perhaps for bilinguals but almost certainly for multilinguals, it is possible that people who are able to maintain knowledge of multiple languages may start out advantaged in certain ways. It is too early to conclude what the effect of knowing more than two languages might be on cognitive outcomes.

A different kind of outcome can be found in language learning. Monolingual children learning their first language sometimes use a strategy of disambiguation to rapidly figure out the meaning of new words by assuming that each object has one unique name, as discussed in Section 1. However, Byers-Heinlein and Werker (2009) extended this idea and
compared 1½-year-old children who were being raised in monolingual, bilingual, or trilingual homes. The results showed a strong reliance on this disambiguation strategy by monolingual children, a marginal and nonsignificant use of the strategy by bilingual children, and no evidence at all for this strategy in trilingual children. Thus, the number of languages in the environment modified children’s expectations about words and their meanings, possibly setting the stage for different paths of language learning.

**Bilingual aphasia and its treatment**

Aphasia (word-finding difficulties) is the commonest outcome of stroke, and yet our understanding is largely restricted to monolingual speakers, whereas a significant portion of stroke patients are bilingual—a proportion that is set to increase. Clinical management is hampered because there is no current basis for predicting speech-production difficulties following stroke in bilingual speakers. Recovery patterns are diverse (Green, 2005; Paradis, 2004): For instance, both languages may recover to the same relative premorbid level (parallel recovery), one may recover better than another, or the progressive recovery of one language may impair the recovery of the other. Without an understanding of the causal bases of these recovery patterns, including the nature of the control processes involved, there can be no principled basis for treatment and no rational basis for identifying the resources required for treatment. For instance, if treatment in one language (e.g., the L1 or current dominant language) transfers to another, then monolingual speech therapy could help in the recovery of both languages. However evidence on this point is equivocal, largely because there are few well-controlled studies (see Kohnert, 2009, for a recent review). Even the decision to treat in one language rather than two reflects an untested assumption that may or may not be appropriate to the individual case. For instance, individuals with a parallel recovery pattern frequently self-cue and produce a correct word in the nontarget language in order to retrieve the intended word. Prescribing use of the nontreated language may not be justified (Ansaldo, Marcotte, Scherer, & Raboyeau, 2008). A case study reported by Ansaldo, Saidi, and Ruiz (2010) exemplifies the value of using the patient’s behavior in both languages and of considering the control processes involved. They treated a highly proficient Spanish-English bilingual with a subcortical lesion that included the left caudate. He had word-finding difficulties in both languages and involuntarily switched between languages within conversations with monolingual partners. On the supposition that distinct control processes mediate translation and speech in just one language (Green, 1986), Ansaldo et al. developed an elegant procedure (“switch back through translation”) that made use of these involuntary language switches and treated the patient successfully.

Our review indicates the intimate relationship between language control and the processes of cognitive control. We expect that successful language recovery will be associated with a tighter coupling between regions linked to language processing and regions (frontal and subcortical) associated with control (Green, 2008). Preliminary data using functional neuroimaging to examine changes in regional coupling during recovery support this conjecture (Abutalebi, Della Rosa, Tettamanti, Green, & Cappa, 2009). If control functions are a strength of bilingual patients, then treatment should make use of them (Penn, Frankel, Watermeyer, & Russell, 2010). More generally, treatments aimed at enhancing or making more effective use of cognitive-control processes may prove to be a useful adjunct to conventional treatment derived from research on monolingual patients with aphasia.

**Protection against dementia**

In previous sections, we reviewed the evidence showing that bilingual children and adults enjoy an advantage over their monolingual counterparts in aspects of attention and cognitive control. In some cases (e.g., Bialystok et al., 2004), this bilingual advantage actually increases in older adulthood, in the sense that performance falls off more steeply with increasing age in monolinguals than it does in bilinguals (see Fig. 3b). This result may be interpreted as showing that bilingualism serves to protect against some aspects of age-related cognitive loss, and prompts the question of whether bilingualism might offer some protection against pathological decline, specifically against the onset of dementia. Such protection might be considered one form of “cognitive reserve”—the protection of cognitive function by stimulating activities (Stern, 2002). Bialystok, Craik, and Freedman (2007) conducted a study of hospital records and found that a sample of 93 lifelong bilinguals experienced the onset of symptoms of dementia some 4 years later than a comparable sample of 91 monolingual patients. The two groups were essentially equivalent on other factors that might have influenced the result. This initial study was followed by another (Craik, Bialystok, & Freedman, 2010) in which approximately 100 bilingual and 100 monolingual patients diagnosed with probable Alzheimer’s disease were questioned about age of onset and other relevant factors. In this sample, the bilingual group had their first clinic visit more than 4 years later than did the monolinguals and had experienced symptoms of dementia more than 5 years later than their monolingual counterparts. As in the first study, the groups were equivalent in cognitive level (MMSE score) and the monolinguals had the greater advantage in terms of education and occupational status. There were no differences in these results in subgroups of immigrants and nonimmigrants. A recent study from a Montreal group (Chertkow et al., 2010) has given partial support to these first findings. In their investigation, Chertkow and colleagues found a bilingual delay in the onset of symptoms in an immigrant group, as well as in a nonimmigrant group whose first language was French, but not in a nonimmigrant group whose first language was English. For people who were multilingual (defined as speaking three or more languages), the delay of onset was again found.

Taking a different approach, Schweizer, Ware, Fischer, Craik, and Bialystok (2010) examined smaller samples of
monolingual and bilingual patients diagnosed with probable Alzheimer’s disease who had also received a CT scan. The samples were matched on cognitive level, so if bilingualism boosts cognitive reserve—maintaining cognitive functions despite accumulated brain pathology—the bilingual group should show more evidence of lesion burden. This was exactly the result: The bilingual group showed substantially more atrophy in temporal regions than did their monolingual counterparts, although the bilingual patients were still able to function at the same cognitive level. These studies support the possibility that the bilingual advantage in cognitive control extends to benefit patients suffering from Alzheimer’s disease and also possibly to other forms of dementia. If confirmed, these findings would make bilingualism one factor that contributes to cognitive reserve, with effects similar to those found for social, intellectual, and physical activity. How exactly cognitive reserve acts to provide compensation for brain pathology is an exciting question for future research.

Conclusion

As described earlier, bilingualism is already common in many parts of the world and is certain to become even more common as the 21st century unfolds. We have summarized the current state of knowledge about language development and cognitive control throughout the lifespan, associated changes in the brain, and the implications of bilingualism for clinical practice. Much remains to be learned, but it is already clear that the consequences of speaking two or more languages are profound, in some cases dramatically so. As one example, if the finding that bilingualism delays the onset of Alzheimer’s disease by 4 to 5 years is confirmed by further research, there are potentially important implications for the concept of cognitive reserve. How exactly does bilingualism change the brain, for example, and which aspects of these changes confer protection against the onset of dementia? Once this is known, findings from bilingualism research may help to focus the search for other environmental conditions with comparable effects. In the same vein, what about countries such as Belgium and the Netherlands, where substantial proportions of the population speak more than one language? Is this associated with a generally later onset of Alzheimer’s disease relative to countries that are largely monolingual? Other intriguing questions include ones concerning the length of time that a person is bilingual: Does learning a second language from infancy provide special benefits, for example, or is it sufficient to speak two languages consistently from the teenage years or even later? What about the similarities of the two languages? Is the bilingual advantage greater (or less?) following the acquisition of highly similar languages such as Spanish and Italian compared to such dissimilar languages as Chinese and English? Given the rapidly accelerating interest in bilingualism as a research topic, answers to these and many other questions should be available in the very near future.

Funding

Preparation of this manuscript was supported by Grant R01 HD052523 from the National Institutes of Health to EB, Grant MOP57842 from the Canadian Institutes of Health Research and a Grant from the Alzheimer’s Society of Canada to EB and FIMC, Grant 089220/Z/09/Z from the Wellcome Trust to DWG, and Grant R01 HD050287 from the National Institutes of Health to THG.

References


