

# How biases inflate scientific evidence

Angela de Bruin and Sergio Della Sala consider the example of the cognitive benefits of bilingualism

'Bilingual brains are more healthy'. 'Being bilingual really does boost brain power'. 'Bilingual adults have sharper brains'. These are just some of the media headlines claiming that people who speak a second language have a cognitive advantage compared with monolinguals. Although exaggerations, these headlines are based on scientific studies finding advantages in suppressing irrelevant information, switching between tasks, and mental flexibility.

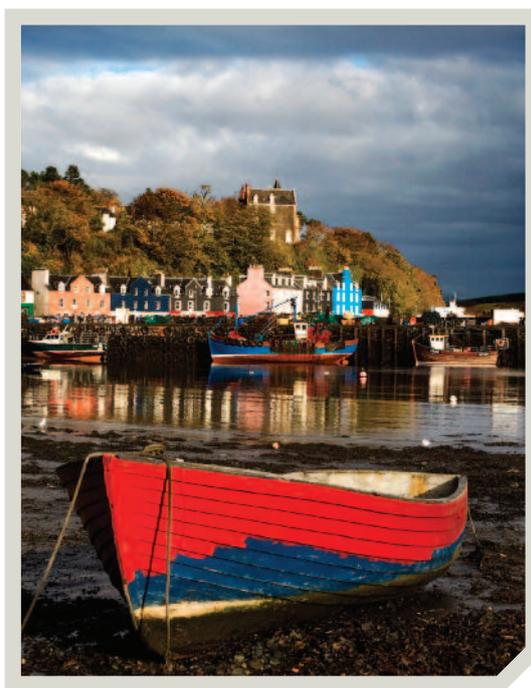
Hoping to make their child smarter, parents are increasingly asking for bilingual nannies to teach their child another language. The finding that bilingualism may delay the onset of dementia by five years has been used by companies to encourage people to buy their language-learning software, again with catchy headlines like 'Worried about Alzheimer's? Learn a second language' (Rosetta Stone) and 'Delay dementia for up to five years through language learning' (Babbel).

But do bilinguals truly have such an impressive cognitive advantage? In this article, we will discuss how a publication bias – positive results are more likely to be published than null or negative results – may have inflated the evidence.

## Bilingualism and executive control

The first studies that showed positive effects of bilingualism required participants to suppress task-irrelevant information. One relatively well-studied

task in this field is the Simon task. In this task, participants need to respond to certain shapes by pressing a button, for example the left button for a triangle and the right button for a square. Shapes appear on the left or right side of the screen, leading to incongruent (e.g. left-side screen, right button) and congruent (e.g. left-side screen, left button) trials.



The Hebrides are a valuable environment to study bilingualism

Incongruent trials commonly elicit slower reaction times (RTs) than congruent trials, which is also called the Simon effect.

Bilinguals have been found to show smaller Simon effects than monolinguals, suggesting that they find it easier to suppress irrelevant information (Bialystok et al., 2004). The explanation behind this inhibitory advantage is based on the finding that bilinguals constantly have to control their two languages. The two languages of a bilingual are apparently always active, even if only one language is needed. Thus, when a bilingual wants to speak in one language, the other language needs to be suppressed. For example, when a French speaker needs to describe a dog in English, it is not only crucial to select the English item, but also to suppress the French word *chien*.

Bilingual advantages have not just been found to be related to inhibition, but also to task switching. When bilinguals and monolinguals were asked to sort stimuli according to shape or colour, bilinguals were faster at switching between the colour and shape decisions than monolinguals (Prior & MacWhinney, 2010). This could be related to language switching: bilinguals who often switch between their two languages may also be better at switching between two non-verbal tasks.

More recently, researchers have argued that a bilingual advantage is more global than just one specific domain of executive control such as inhibition or switching. Rather, the advantage could extend to 'conflict monitoring', 'coordination', or 'mental flexibility': in other words, bilinguals are generally better at monitoring and solving conflicts. Yet perhaps the most surprising outcome of

## references

- Alladi, S., Bak, T.H., Duggirala, V. et al. (2013). Bilingualism delays age at onset of dementia, independent of education and immigration status. *Neurology*, 81(22), 1938–1944.
- Bialystok, E., Craik, F.I., Klein, R. & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging*, 19(2), 290–303.
- Bialystok, E., Kroll, J.F., Green, D.W. et al. (2015). Publication bias and the validity of evidence: What's the connection? *Psychological Science*. doi:10.1177/0956797615573759
- Carson, A.J., MacHale, S., Allen, K. et al. (2000). Depression after stroke and lesion location: A systematic review. *The Lancet*, 356(9224), 122–126.
- Carter, A.S., Messinger, D.S., Stone, W.L. et al. (2011). A randomized controlled trial of Hanen's 'More Than Words' in toddlers with early autism symptoms. *Journal of Child Psychology and Psychiatry*, 52(7), 741–752.
- Chambers, C.D. (2013). Registered reports: A new publishing initiative at *Cortex*. *Cortex*, 49(3), 609.
- de Bruin, A., Bak, T.H. & Della Sala, S. (2015). Examining the effects of active versus inactive bilingualism on executive control in a carefully matched non-immigrant sample. *Journal of Memory and Language*, 85, 15–26.
- de Bruin, A. & Della Sala, S. (2015). The decline effect: How initially strong results tend to decrease over time [Advance online publication]. *Cortex*. doi:10.1016/j.cortex.2015.05.025
- de Bruin, A., Treccani, B. & Della Sala, S. (2015a). Cognitive advantage in bilingualism: An example of

bilingual–monolingual comparisons concerns the finding that bilingualism may delay the onset of dementia by approximately four to five years (e.g. Alladi et al., 2013).

### Bilingualism in the Hebrides

These beneficial effects of bilingualism could have important practical and societal implications and are now often presented as accrued wisdom. However, many other studies have failed to find a cognitive effect of bilingualism. We recently conducted a study in the Scottish Hebrides, comparing Gaelic–English bilinguals to English monolinguals (de Bruin et al., 2015). The Hebrides are a valuable environment to study bilingualism, as bilingual and monolingual speakers come from similar backgrounds and are living in similar environments. Many previous studies that showed an effect of bilingualism compared bilinguals with monolinguals that differed in background variables such as immigration status, education, country of origin or lifestyle. In our study, we therefore tested non-immigrant bilinguals and monolinguals and were particularly thorough in matching them on background variables. We furthermore compared active bilinguals, who still used both Gaelic and English, with inactive bilinguals, who only or mainly used English. We gave participants a series of tests, including tasks measuring their ability to suppress information (a Simon task similar to the one described above) and their ability to switch between tasks. No overall effects of active or inactive bilingualism on inhibition or switching were found.

Our study is not the only one that failed to observe an effect of bilingualism. Testing large numbers of participants, several studies did not observe a bilingual effect on tasks that had previously shown advantages compared with monolinguals (e.g. Paap & Greenberg, 2013). Similarly, the delaying effects regarding dementia have been challenged too (e.g. Lawton et

al., 2015). So despite the initial studies showing large effects of bilingualism, most recent studies have not observed effects of bilingualism at all or only in restricted circumstances. In fact, it has been estimated that 80 per cent of the studies conducted after 2011 showed no positive effect of bilingualism (Paap et al., 2015). Thus, there appears to be a shift from studies showing strong effects of bilingualism to more recent studies challenging this advantage.

### The decline effect

To examine this apparent shift in evidence, we analysed studies testing the effects of bilingualism on executive control published between 2004 and 2014 (de Bruin & Della Sala, 2015). We classified the studies as overall ‘supporting’ or ‘challenging’ a bilingual advantage, or as having ‘mixed’ results if no conclusion was drawn. Whereas the majority of initial studies supported this hypothesis, the picture has been more balanced in the past few years (when the number of overall studies has also increased). This is furthermore supported by Klein (2015), who compared results found on two specific tasks (Simon and flanker tasks) over the years. Large reaction time differences between bilinguals and monolinguals were only found in initial studies. Later studies either showed no difference at all or very small effects.

This phenomenon of a decrease in positive evidence after a strong initial finding is not unique. Dubbed the ‘decline effect’, it has been observed in various research fields. For example, Ozonoff (2011) describes how evidence for a widely used autism treatment is now diminishing (Carter et al., 2011). Similarly, initial research showed that depression was more often associated with left-hemisphere strokes than right-hemisphere strokes, whereas later systematic reviews showed no effect of stroke location on depression (Carson et

## Meet the authors

‘A few years ago, our group became interested in the claim that bilingualism may boost cognition. We investigated this question and obtained positive, negative, and null results. However, we only managed to publish the positive findings and asked ourselves whether this is a more general problem happening to other researchers. We investigated and found the issue of publication bias in the field of bilingualism and executive control, but also in other fields. This does not mean that this effect or other effects do or do not exist, but researchers and the society should be aware of these biases. We can only increase our knowledge if we share all sound data regardless of the outcome.’



**Angela de Bruin**

is a PhD student at Human Cognitive Neuroscience, Psychology, University of Edinburgh  
angela.debruin@ed.ac.uk



**Sergio Della Sala**

is Professor at Human Cognitive Neuroscience, Psychology, University of Edinburgh  
sergio@ed.ac.uk

al., 2000). More generally, decline of evidence has been discussed for clinical (Ioannidis, 2006) and experimental psychology studies (Francis, 2012).

Several reasons have been suggested to explain this decline effect (for discussion see: Lehrer, 2010; Schooler, 2011). Regression to the mean is the most common statistical explanation. A first finding may be excessively large due to errors. In subsequent studies, statistical self-correction should lead to values closer to the mean. In the field of bilingualism, conceptual rather than

publication bias? *Psychological Science*, 26(1), 99–107.  
de Bruin, A., Treccani, B. & Della Sala, S. (2015b). The connection is in the data: We should consider them all. *Psychological Science*, 26(6), 947–949.  
Easterbrook, P.J., Gopalan, R., Berlin, J.A. & Matthews, D.R. (1991). Publication bias in clinical research. *The Lancet*, 337(8746), 867–872.  
Francis, G. (2012). Too good to be true:

Publication bias in two prominent studies from experimental psychology. *Psychonomic Bulletin & Review*, 19(2), 151–156.  
Franco A., Malhotra, N. & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345, 1502–1505.  
Ioannidis, J.P. (2006). Evolution and translation of research findings: From bench to where? *PLoS Clinical*

*Trials*, 1(7), e36.  
doi:10.1371/journal.pctr.0010036  
Ioannidis, J.P., Munafò, M.R., Fusar-Poli, P. et al. (2014). Publication and other reporting biases in cognitive sciences: Detection, prevalence, and prevention. *Trends in Cognitive Sciences*, 18(5), 235–241.  
Jarrett, C. (2013). Revolutionary or stifling? *The Psychologist*, 26, 626–629. Available at

tinyurl.com/oofvcr4  
Klein, R.M. (2015). Is there a benefit of bilingualism for executive functioning? *Bilingualism: Language and Cognition*, 18(1), 29–31.  
Lawton, D.M., Gasquoine, P.G. & Weimer, A.A. (2015). Age of dementia diagnosis in community dwelling bilingual and monolingual Hispanic Americans. *Cortex*, 66, 141–145.  
Lehrer, J. (2010, 13 December). The truth

direct replications may be another explanation for the increase in null findings. Studies have aimed to replicate the bilingual advantage in different types of executive control tasks and using different types of bilingual populations. Some of these may not elicit an effect of bilingualism, for example when speakers with a low proficiency in their second language are tested instead of high proficiency speakers. Trying to find the boundaries of an effect is likely to yield more null effects.

Research practices and biases have also been suggested as the underlying mechanism. Initial studies are typically reported with smaller sample sizes, whereas replications tend to include more participants. Larger studies have been linked to smaller effect sizes (McMahon et al., 2008), which could explain why later studies with more participants also report smaller effects. The lack of self-replication is important in this respect too. Publications often only report one experiment without any replications. This single experiment with a small number of participants may yield a large effect size. However, if this positive result is due to errors, replications will struggle to obtain similarly large effect sizes, or any effect at all.

Another well-discussed bias is the publication bias. When a hypothesis is tested for the first time, it may be easier to publish positive results with large effect sizes. Null results are not meaningful if there is not yet any evidence for the existence of a certain phenomenon. However, null results may become more interesting and easier to publish once there is a more established theory. A publication bias has been described for many different research fields, including psychology (e.g. Francis, 2012), social sciences (e.g. Franco et al., 2014), and clinical research (e.g. Easterbrook et al., 1991). The existence of a publication bias could explain why positive findings

are overrepresented in the literature (cf. Ioannidis et al., 2014).

We conducted a meta-analysis of published studies to get an impression of the effect sizes found in research on bilingualism and executive control. This analysis showed an average effect size of  $d = .30$ , which can be interpreted as a small, but positive effect of bilingualism on executive control. However, the meta-analysis was based on published results only and also showed evidence for the existence of a publication bias.

We wanted to examine whether a publication bias could have inflated this apparent positive effect (de Bruin et al., 2015a). We collected conference abstracts on the topic of bilingualism and executive control presented at 169 conferences between 1999 and 2012. We then classified these abstracts in four groups: studies fully supporting the bilingual advantage; studies with mixed data that, on the whole, supported the bilingual

advantage; studies with mixed data that mainly challenged the bilingual advantage; studies that fully challenged the bilingual advantage.

Next, we checked which results presented in the conference abstracts were

eventually published in a scientific journal. In total, half of the results were published in a journal. Whereas 68 per cent of the studies fully supporting a bilingual advantage were published, only 29 per cent of the fully challenging studies were published. The two types with mixed results scored in-between: Studies with supporting mixed results had a publication rate of 50 per cent, whereas studies with challenging mixed results were published 39 per cent of the time. There was thus a clear difference in publication outcomes – studies fully or mainly challenging the bilingual advantage were less likely to be published than studies fully or mainly supporting this idea.

Why are some studies not published? Sometimes studies presented at conferences are not published for a good reason. Researchers might only conduct an experiment to pilot a new idea or task, or the design may be flawed. They might not test enough participants to draw reliable conclusions and might therefore decide not to publish their results. In our analysis, we therefore also tested for potential background differences. Challenging and supporting studies used very similar tasks, similar numbers of participants and did not differ in the year of the conference abstract or the likelihood of detecting an effect. Thus, the differences in publication outcomes were not due to quality differences.

A publication bias could have occurred at different stages of the writing and publishing process. Researchers might decide not to publish certain data, because they deem them uninteresting or the results do not fit their theories. Reviewers and editors might furthermore be more likely to reject null and negative data compared with positive data. These rejections are often based on the idea that null results are not interesting enough, the result of flawed experiments, or the result of a small participant sample. The general aversion to null results has been nicely demonstrated: Mahoney (1977) asked journal reviewers to referee manuscripts with positive, negative or mixed results. Although the papers differed in their result types, the methodological procedures were identical. Yet, reviewers scored papers with positive results as methodologically better than manuscripts with negative or mixed results. Manuscripts with negative, null or mixed results were furthermore mostly rejected, whereas positive results were accepted with moderate revisions.

### A distorted representation

Deciding to publish some, but not all, studies depending on the type of results is very problematic for data interpretation in any field. A publication bias does not

“a publication bias does lead to a distorted representation of the actual effects”

wears off. *The New Yorker*. Retrieved from [www.newyorker.com/reporting/2010/12/13/101213fa\\_fact\\_lehrer?currentPage=all](http://www.newyorker.com/reporting/2010/12/13/101213fa_fact_lehrer?currentPage=all)

Mahoney, M.J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research*, 1(2), 161–175.

McMahon, B., Holly, L., Harrington, R. et al. (2008). Do larger studies find

smaller effects? The example of studies for the prevention of conduct disorder. *European Child & Adolescent Psychiatry*, 17(7), 432–437.

Ozonoff, S. (2011). The first cut is the deepest: Why do the reported effects of treatments decline over trials? [Editorial]. *Journal of Child Psychology and Psychiatry*, 52(7), 729–730.

Paap, K.R. & Greenberg, Z.I. (2013). There is no coherent evidence for a

bilingual advantage in executive processing. *Cognitive Psychology*, 66(2), 232–258.

Paap, K.R., Johnson, H.A. & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*. doi:10.1016/j.cortex.2015.04.014

Prior, A. & MacWhinney, B. (2010). A

bilingual advantage in task switching. *Bilingualism: Language and Cognition*, 13(2), 253–262.

Rhodes, E. (2015). Is science broken? *The Psychologist*, 28, 348–349. Available at [tinyurl.com/ncbvpg8](http://tinyurl.com/ncbvpg8)

Schooler, J. (2011). Unpublished results hide the decline effect. *Nature*, 470, 437.

Sumner, P., Vivian-Griffiths, S., Boivin, J. et al. (2014). The association between



Although initial studies showed large effects of bilingualism, these have been challenged in more recent studies

mean that an effect does not exist, because it does not directly validate the quality of published evidence. However, a publication bias does lead to a distorted representation of the actual effects.

Although we have taken the literature on bilingualism and executive control as an example, publication bias is a common phenomenon that affects many research areas. Effects of publication bias may be most damaging in medicine studies. For example, the drug Tamiflu, used as a treatment against influenza, was approved after several clinical trials that showed that the drug worked. Five years later, when the drug maker disclosed the full findings, 70 additional and unpublished trials were discovered, many with negative or inconclusive results. Including these hidden trials led to a more complete interpretation, and many of the assumed effects of Tamiflu could no longer be proved (see Tavel, 2015).

Only having access to 'successful' studies and only reading about studies that have found an effect will let us believe that this effect is strong and unchallenged. What we do not know, however, is how many 'failed' studies with no effects were hidden in a drawer. We should be aware that there are many well-conducted studies without positive effects that remain unpublished – in the field of bilingualism, but possibly also in all other research fields. This is firstly problematic for the scientific interpretation of a phenomenon. The results of a meta-analysis are not reliable when based on published studies only. We can only discuss the effects of bilingualism when we have access to all study results rather than just the positive effects. Similarly, even if an effect does exist, null or negative results can help researchers to establish the boundaries of this effect. In the case of the potential bilingual advantage, we now know that this effect is small at best. It has been found in many studies, but also challenged in others. It is often absent in studies with younger adults or certain language groups, or is only found in certain (parts of) tasks. Positive results can inform us about the tasks and participant groups that do elicit an effect. Null or negative effects, however, can be equally informative and can tell us about the circumstances that do not show an effect.

Publication biases and selective reporting could be diminished through

pre-registration and more transparency (cf. Registered Reports in *Cortex*: Chambers, 2013; and the TOP guidelines on Transparency and Openness Promotion in Science, 2015; see discussion in *The Psychologist*: Jarrett, 2013, and Rhodes, 2015). Registering and reviewing ideas and methods before the study is conducted will encourage researchers to specify their hypothesis before data collection. Moreover, it would enforce reporting all pre-registered analyses and results.

The interpretation of null effects can be improved through the use of Bayesian analyses that provide quantifiable evidence for a null result. Whereas the traditional  $p$ -values used in null hypothesis statistical testing can only say that there is lack of evidence for an effect, Bayes factors allow one to directly compare evidence favouring the null ('no effect of bilingualism') with evidence for the alternative hypothesis ('an effect of bilingualism'). In this way, data that show no difference between bilinguals and monolinguals can be supported by statistical evidence, thus strengthening the interpretation of null effects.

Finally, researchers themselves should be careful not to exaggerate their findings in press releases (Sumner et al., 2014). Taking all these steps could help to ensure that the field of bilingual advantage receives headlines that are scientifically justified.

### Expanding minds

The evidence for a bilingual advantage is not as pervasive as commonly assumed. Although initial studies showed large effects of bilingualism, these have been challenged in more recent studies. Of course, there are still scientists who believe in a strong and unchallenged bilingual advantage. In a commentary on our paper discussing publication bias, leading researchers in this field doubted the importance of null effects (Bialystok et al., 2015; cf. de Bruin et al., 2015b, for a response). Yet for research to progress, we must surely have the full story: all data should be shared, regardless of the outcome.

We should emphasise that bilingualism and second language learning is an advantage by definition. It allows you to travel to other countries, learn about other cultures, meet new people, broaden horizons and expand minds. When will psychology truly broaden its own horizons, by ensuring that research – especially on issues of societal relevance and popularity – is free from publication bias?

exaggeration in health related science news and academic press releases: Retrospective observational study. *British Medical Journal*, 349, g7015.  
Tavel, M.E. (2015). Bias in reporting of medical research: How dangerous is it? *Skeptical Inquirer*, 39(3), 34–38.