The decline effect: How initially strong results tend to decrease over time

Angela de Bruin & Sergio Della Sala

1) Human Cognitive Neuroscience, Psychology, University of Edinburgh, Edinburgh, UK
2) Centre for Cognitive Ageing and Cognitive Epidemiology, Psychology, University of Edinburgh, Edinburgh, UK

Address correspondence to:
Angela de Bruin
Human Cognitive Neuroscience, Psychology
University of Edinburgh
7 George Square
Edinburgh EH8 9JZ - United Kingdom
Phone: +44 (0) 131 650 3426
Fax: +44 (0) 131 651 3230
Email: angela.debruin@ed.ac.uk

Word count: 998
Paap, Johnson, and Sawi (2015) argue that a bilingual advantage in executive functioning (EF) either does not exist or only occurs infrequently and in undetermined circumstances. They furthermore argue that biases may have inflated the evidence for a bilingual advantage. Their overview of published results shows that more than 80% of the tests conducted after 2011 showed no effect of bilingualism, especially in studies with larger sample sizes. In this commentary, we will describe how evidence for a bilingual advantage has changed over time and which reasons may underlie a decline of positive evidence.

The relatively short but extensive history of research on bilingualism and EF has seen a shift from initial studies presenting strong evidence for a bilingual advantage to more recent studies criticising the reliability or even the existence of such an advantage. One of the first and most influential studies on this topic (Bialystok, Craik, Klein, & Viswanathan, 2004) reported the highest effect sizes observed in this literature (ranging up to $d = 3$). In the following years, further studies reported positive effects of bilingualism but with smaller effect sizes. The average effect size in published studies now seems to be substantially smaller (around $d = .30$; de Bruin, Treccani, & Della Sala, 2015).

In 2011, Hilchey and Klein concluded that their review showed limited evidence for an inhibitory effect of bilingualism, but ‘robust’ evidence for a global bilingual advantage. In an updated review (Hilchey, Saint-Aubin, & Klein, in press), however, they conclude that the evidence for a global advantage has evaporated since their initial review. This change in conclusions is compatible with Figure 1 presented by Klein (2015), who compared reaction times on Simon and flanker tasks between bilinguals and monolinguals. Large bilingual-monolingual differences were predominantly found in earlier studies, but less so in recent publications. To examine this apparent decline in support for a bilingual advantage, we created an overview of studies on bilingualism and EF published between 2004 and 2014 (Figure 1). Based on the overall conclusions presented in these papers, we classified them as ‘supporting’ or ‘challenging’ a bilingual advantage, or as ‘mixed’ if no conclusion was drawn. The pattern of supporting versus challenging studies has indeed changed over time.

---

1 Publications were found through a PubMed and Scopus search using the keywords ‘bilingual executive’, ‘bilingual cognitive’, and ‘bilingual advantage’. A systematic search through the references yielded an additional five relevant papers.

2 Contrary to Paap et al.’s overview based on individual tasks and comparisons, we based our classifications on overall conclusions. Some papers were classified as ‘supporting’ even if the results included null effects. Figure 1 is therefore likely to give an exaggerated impression of the actual effects of bilingualism.
Whereas earlier studies largely supported a bilingual advantage, recent years (especially 2014) have shown an upsurge in studies challenging this view.

Thus, the support for a bilingual advantage appears to have diminished in recent years. A decrease in positive evidence after a strong initial finding is not uncommon in science and is dubbed the ‘decline effect’. In many research fields, initial studies have shown large effects whereas later studies struggle to replicate these findings or only find effects in restricted circumstances. For example, the decline effect accounts for the lack of confirmation of a widely used treatment for autism (Carter et al., 2011, linked to the decline effect by Ozonoff, 2011), of the link between depression and left-hemisphere lesion in stroke (Carson et al., 2000), of the link between type D personality and mortality rates (Coyne & Voogd, 2012), as well as in several clinical (Ioannidis, 2006) and experimental psychology studies (Francis, 2012).

Several explanations for this decline effect have been suggested (cf., Schooler, 2010, & Lehrer, 2011). One common statistical explanation is regression to the mean: If initial measurements show an inflated effect size due to errors, statistical self-correction will lead to results closer to the average in subsequent measurements. A publication bias (cf., de Bruin et al., 2015) may furthermore help to interpret the decline effect. It is difficult to publish null effects or small effect sizes when a new hypothesis is tested for the first time. Initial effects are therefore usually positive and large. However, when a theory becomes more established, challenging studies may become more interesting and easier to publish. Changes in methodology and research practices could also contribute to a decline effect. Whereas initial findings are often reported with smaller sample sizes, follow-up studies tend to include more participants. Considering that larger studies have been associated with smaller effect sizes (McMahon, Holly, Harrington, Roberts, & Green, 2008), this could explain part of the decline in effect sizes. Furthermore, due to publication pressure, studies are often reported with only one experiment without self-replication. Especially in combination with selectively reporting only those experiments that work, replication in follow-up studies may prove to be difficult. Pre-registration may be a valuable tool to diminish publication bias and selective reporting (cf., Registered Reports in Cortex, Chambers, 2013). In the literature of bilingualism and EF, the attempt to establish the boundaries of a bilingual advantage may also have led to an increase in null effects. Recent studies have examined the limits of a bilingual effect through the use of different types of tasks, different methodologies, and
different populations that often do not show an effect of bilingualism. This is likely to have increased the amount of null results. The decline effect does not mean that a phenomenon does not exist. Rather, it explains how evidence for a phenomenon can change over time and how initially strong findings are likely to be challenged in later studies. It should furthermore encourage researchers to critically evaluate published findings. Publication biases and selective reporting can affect the presentation of results over the years. For research to progress, it is essential to unbiasedly report all results regardless of the outcome.

**Figure 1.** Overview of 108 studies examining bilingualism and executive control published between 2004 and 2014. The result types (‘supporting’, ‘mixed’, ‘challenging’) are based on the overall conclusions of the paper rather than on results of individual tasks.
References


Coyne, J. C., & de Voogd, J. N. (2012). Are we witnessing the decline effect in the Type D personality literature? What can be learned?. *Journal of psychosomatic research, 73*(6), 401-407.


Paap, K. R., Johnson, H. A., & Sawi, O. (in press). Bilingual advantage in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. Cortex.