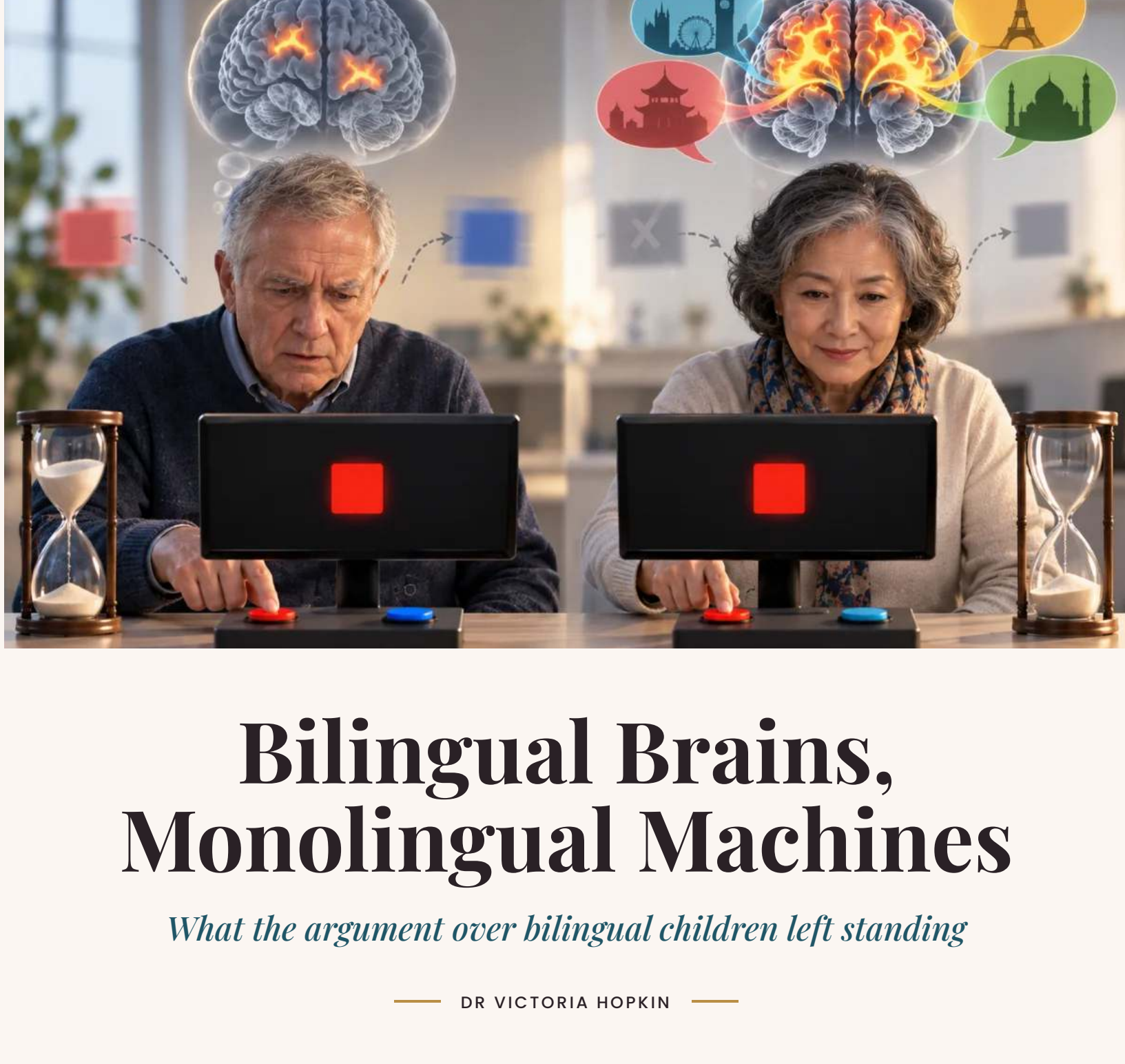




Bilingual Brains, Monolingual Machines

What the argument over bilingual children left standing

— DR VICTORIA HOPKIN —

Toronto, 2004

In 2004 a psychologist at York University in Toronto named Ellen Bialystok published a result that, for a while, changed what parents thought a second language was for.

She had been studying bilingual children for twenty years and had a suspicion that the constant work of keeping two languages apart, of choosing one and suppressing the other with every sentence, might be a kind of exercise. With Fergus Craik, a memory researcher at the Rotman Institute, she tested older adults on a task called the Simon task, in which a coloured square appears on the left or right of a screen and you press a key for the colour while ignoring the side. The side is a distraction. Suppressing it is a small act of control. The bilingual adults were faster and, more strikingly, their reaction times had slowed less with age than the monolinguals' had.

Three years later came the finding that made the newspapers. Bialystok, Craik and a neurologist named Morris Freedman went through the records of 184 patients at the Baycrest memory clinic in Toronto and sorted them by whether they had lived their lives in one language or two. The bilingual patients had begun showing symptoms of dementia, on average, about four years later than the monolingual ones. The education levels were similar. The bilinguals had, it appeared, been protected by something they had been doing all their lives without thinking of it as protection.

Parents took note. So did schools, and immersion programmes, and the makers of language apps, who now had a claim that went well beyond communication. A second language, it seemed, made a better brain. And the claim had a name, the bilingual advantage, and a laboratory, and a decade of studies pouring out of Toronto and everywhere else that found bilingual children ahead on tasks of attention and control.

Then the field did what fields are supposed to do, and the story of what happened next is worth telling in full, because the ending is not the one the parents were sold, and what survived is more useful than what was lost.

The Argument

Kenneth Paap is a psychologist at San Francisco State who had spent his career on reading, and in 2013 he did something simple. He tried to find the bilingual advantage himself. He gave 286 students the same kinds of tasks Bialystok used, and found nothing. Then he looked hard at the studies that had found something, and noticed that they were often small, that the tasks did not agree with each other about what was being measured, and that a bilingual advantage on one test rarely appeared on the next. He said so, in a paper with a title that did not hedge: there is no coherent evidence for a bilingual advantage in executive processing.

"The studies had been run. The field had simply been more interested in one answer, and the journals had followed."

Angela de Bruin, then a doctoral student in Edinburgh, went looking for the reason the literature and Paap's laboratory disagreed. She gathered the abstracts of conference presentations on bilingualism and cognition from 1999 to 2012, and then traced which of them had gone on to be published as papers. The ones reporting an advantage were published about two thirds of the time. The ones reporting no advantage were published less than a third of the time. The studies had been run. The field had simply been more interested in one answer, and the journals had followed.

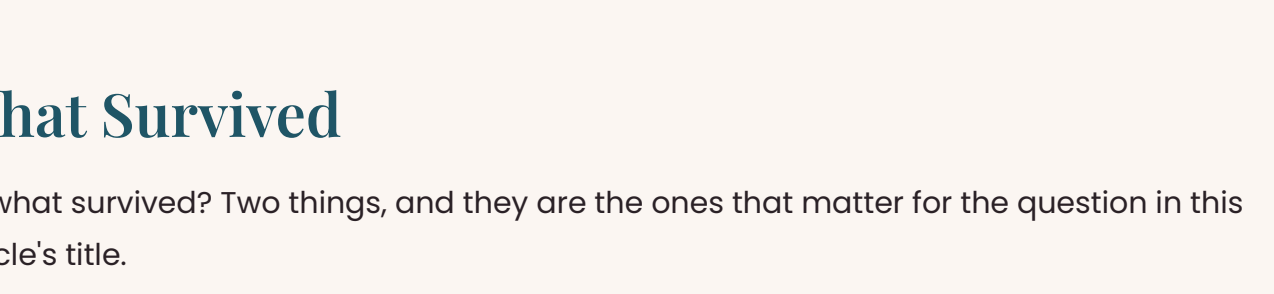


Fig. 1 — Publication bias in the bilingual-advantage literature: conference abstracts from 1999–2012 traced to publication (de Bruin, Treccani & Della Sala, 2015).

What settled the argument was scale. In 2019 a team led by Anthony Dick at Florida International University examined 4,524 children aged nine and ten from a national study of brain development, and found no bilingual advantage on executive function. In 2020 Emily Nichols and Adrian Owen in Ontario tested 11,000 adults online and found none. And in 2021 Cassandra Lowe and her colleagues at Western University gathered every study of children they could find, 1,194 separate effect sizes covering more than 23,000 children between three and seventeen, and put them together. The overall effect was 0.08, which is small, and after correcting for the fact that positive results had been more likely to be published, it was indistinguishable from zero.

Bialystok did not concede, exactly, and the argument continues in the journals. But in 2017 she published a long paper in *Psychological Bulletin* in which the word advantage had been replaced by another. She now wrote of the bilingual adaptation: minds accommodating experience, in ways that might show up in some tasks and not others, that might matter for some people at some ages. It was a more careful claim, and a much smaller one, and it was, I think, the right one.

What Survived

So what survived? Two things, and they are the ones that matter for the question in this article's title.

The first is a study that Katherine Kinzler and Boaz Keysar ran in Chicago in 2015, with Samantha Fan and Zoe Liberman. They took children aged four to six and divided them three ways: those who spoke only English, those who were bilingual, and a third group who spoke only English but lived in homes where they regularly heard another language, from a grandparent or a nanny or a parent on the phone. The children were merely exposed. They could not speak the second language.

Then they played a game. A grid of objects sat between the child and an adult, and the adult asked the child to move "the small car". There were three cars, but the smallest of them was hidden from the adult's side by a screen. The child could see it. The adult could not. A child who understood what the adult could see would move the car the adult meant, the smallest one visible to her. A child who did not would reach for the smallest car on the grid.

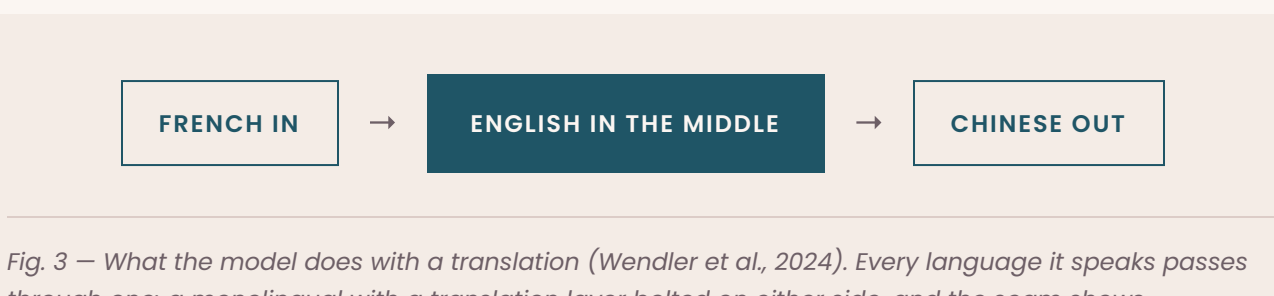


Fig. 2 — The small-car task (Fan, Liberman, Keysar & Kinzler, 2015). Whatever this was, it was not the product of managing two grammars. It was the product of growing up in a house where you had to work out, constantly, who could understand what.

The monolingual children got it right about half the time. The bilingual children got it right about three quarters of the time. And so did the exposed children, the ones who could not speak a second language at all. Whatever this was, it was not the product of managing two grammars. It was the product of growing up in a house where you had to work out, constantly, who could understand what. The paper called it the exposure advantage, and it is, as far as I can tell, the most robust thing in the whole literature.

The second survivor is older and quieter. In 2001 Bialystok herself, in a book that predates the advantage claim, documented what linguists call metalinguistic awareness: the bilingual child's early understanding that a word is a label rather than a property of the thing, that the same object has two names, that any sentence is one way of saying something and not the only way. A monolingual five-year-old, asked whether you could call the sun "the moon" if everyone agreed, tends to say no. A bilingual five-year-old tends to say yes, and can tell you why. That awareness is not an advantage in the Simon task. It is an awareness that language is a choice, and it is exactly the awareness that a person needs in order to work with a machine that has made its choices already.

Two Monolinguals

In 1989 François Grosjean, a Swiss psycholinguist, wrote a short paper with an exasperated title: neurolinguists, beware, the bilingual is not two monolinguals in one person. His point was that researchers kept testing bilinguals as if each language were a separate, slightly worse version of a native speaker's, when in fact a bilingual has one repertoire, used whole, and moves through it according to who is in the room. The paper became a founding text of the field. It is also, it turns out, the most precise description available of what a language model is not.

Chris Wendler and Robert West, at the Swiss Federal Institute of Technology in Lausanne, published a study in 2024 whose title asks the question directly: do llamas work in English? Llama is Meta's open language model, trained on text in dozens of languages, and it will happily translate a French sentence into Chinese. Wendler and his colleagues opened the model up and watched what happened in the layers between the French going in and the Chinese coming out. In the middle, the representations were English. The model was not translating French into Chinese. It was translating French into something very like English, and then that into Chinese. Every language it spoke passed through one.

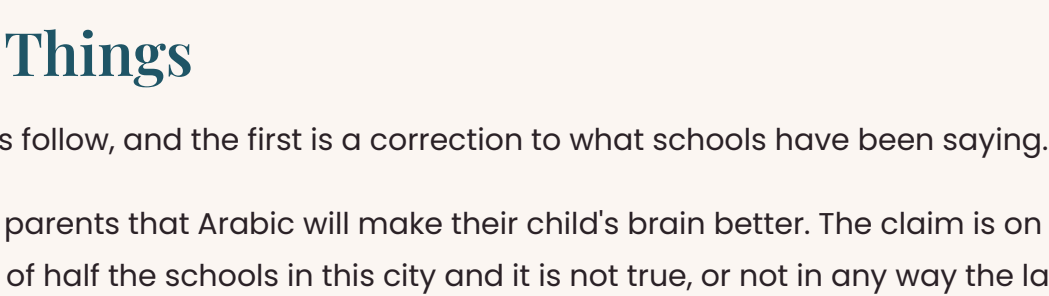


Fig. 3 — What the model does with a translation (Wendler et al., 2024). Every language it speaks passes through one: a monolingual with a translation layer bolted on either side, and the seam shows.

A team at Kyoto, led by Sadao Kurohashi, then asked what a model built in Japan would do, and reported in 2025 that a Japanese-centric model thinks in Japanese. The finding generalises: every model has a dominant language, the one it saw most in English, and it routes everything else through that. Which is Grosjean's warning realised in silicon. The machine is precisely two monolinguals in one person, or rather one monolingual with a translation layer bolted on either side, and the seam shows.

Where it shows most clearly is code-switching, the thing every bilingual does without effort and most do several times a sentence. Sarah Hopkyns, whose interviews with Emirati students were described in the first piece in this series, records young people moving between Arabic and English mid-sentence for emphasis, for irony, for intimacy, for the pleasure of the word that only exists in one of the languages. In 2023 Ruochen Zhang and Alham Fikri Aji at MBZUAI in Abu Dhabi tested whether the language models could do this, and gave their paper a title that is also the finding: multilingual large language models are not (yet) code-switchers. The models could understand a mixed sentence, roughly. They could not produce one that a bilingual would recognise, and when a switched sentence was used to ask them something, their performance fell. Hyunwoo Yoo at Seoul found in 2025 that the same weakness runs through the models' safety training: switch languages mid-prompt and the guardrails loosen, because they were built in one language at a time.

The Exposed Child

Now put the human findings beside the machine findings, because the pairing is the argument.

The bilingual advantage that parents were sold, the faster Simon task, the better brain, has not survived. But the child in Kinzler's Chicago experiment, who could see that the adult could not see the small car, has. That child is not better at control. She is better at the question every conversation turns on: what does this person mean, given what they can see? And the machine is exactly the interlocutor for whom that question matters most, because it cannot see anything. It has no room, no grandmother on the phone, no idea which of its own outputs is a feature of its dominant language and which is a fact about the world.

Consider what the exposed child does. When the writing assistant corrects her sentence, as the first piece in this series described. A monolingual child sees a red line and a better sentence, and accepts. The exposed child has grown up knowing that the same thing has more than one name, and that the name depends on who is asking. She is in a position to ask whether the correction is a preference or a preference. Nothing in the executive-function literature would have predicted that. Everything in the exposure–advantage and metalinguistic–awareness literature does.

And now the Gulf, where the policy described in the second piece in this series is, whether or not its authors thought of it this way, a policy for manufacturing exposure. Two hundred minutes of Arabic at week from the age of four do not make an English-speaking child bilingual. Kinzler's finding says it does not need to. It makes her an exposed child, in a city where the exposure continues at the school gate and the bakery and the neighbour's house, and the exposed children in Chicago took the adult's perspective as well as the bilingual ones did. The KHDA policy is buying the thing the literature says is real, at the age the literature says it forms, in a place where it is reinforced every day, and nobody has yet measured whether it is working.

Three Things

Three things follow, and the first is a correction to what schools have been saying.

Stop telling parents that Arabic will make their child's brain better. The claim is on the prospectus of half the schools in this city and it is not true, or not in any way the largest studies can find, and a parent who reads the Lowe meta-analysis will know it. The claim that survives is smaller and more useful: a child who grows up hearing another language becomes better at working out what other people mean, at an age when that skill is forming, and the effect does not require fluency. That is a claim a school can make honestly, and a claim that, unlike the other one, describes something a parent can see at the dinner table.

Teach the switch. The thing the machine cannot do, moving between languages for effect, is the thing Emirati teenagers do for pleasure and Arabic B pupils will do badly at first and then well. Most schools treat code-switching as a lapse. The literature since Grosjean treats it as the signature of a single repertoire being used whole, and the machine studies treat it as the clearest boundary between what a bilingual is and what a model is. A lesson in which pupils are asked to say the same thing three ways, in Arabic, in English, and in the mixture they would actually use with a friend, and then to ask a model to do the same, teaches more about both languages and about the machine than any grammar drill.

Measure the thing that is real. The LEAP Monitor, the termly study of early Arabic provision in Dubai that the Cambridge Consultancy Group begins publishing this year, will from its second issue include a version of Kinzler's small-car task with children in the two-school panel. It is a ten-minute game with a grid and a screen, it has a published protocol, and if the exposure the policy is paying for is producing the perspective-taking the literature predicts, that is where it will show. No school in the region has data on this. It is cheaper to collect than almost anything else a school measures.

The Small Car

Bialystok is still at York, and still, in her eighties, publishing. The Baycrest finding about dementia has held up better than the Simon task did, and the truth is that nobody yet knows exactly what a life in two languages does to a brain, only that it does something, and that the something is smaller and stranger than the newspapers said in 2007.

"It cannot see the grid, and it does not know its own English is a choice, because it was never given one."

What the twenty-year argument left standing is not a better brain. It is a child who knows that the adult cannot see the small car, and that child who knows that the sun could be called the moon if everyone agreed. Those are not advantages in the way the field once meant. They are habits of attention to other minds, and to the fact that every sentence is a choice. The machine has neither. It cannot see the grid, and it does not know its own English is a choice, because it was never given one. Which is why the bilingual child, the one the studies could not find an advantage in, may turn out to be unusually well placed to work with it.

Sources referred to in this piece

- Bialystok, E. (2001) *Bilingualism in development: language, literacy, and cognition*. Cambridge: Cambridge University Press.
- Bialystok, E. (2017) 'The bilingual adaptation: how minds accommodate experience', *Psychological Bulletin*, 143(3), pp. 233–262.
- Bialystok, E., Craik, F.I.M., Klein, R. and Viswanathan, M. (2004) 'Bilingualism, aging, and cognitive control: evidence from the Simon task', *Psychology and Aging*, 19(2), pp. 290–303.
- Bialystok, E., Craik, F.I.M. and Freedman, M. (2007) 'Bilingualism as a protection against the onset of symptoms of dementia', *Neuropsychologia*, 45(2), pp. 459–464.
- de Bruin, A., Treccani, B. and Della Sala, S. (2015) 'Cognitive advantage in bilingualism: an example of publication bias?', *Psychological Science*, 26(1), pp. 99–107.
- Dick, A.S. et al. (2019) 'No evidence for a bilingual executive function advantage in the nationally representative ABCD study', *Nature Human Behaviour*, 3, pp. 692–701.
- Fan, S.P., Liberman, Z., Keysar, B. and Kinzler, K.D. (2015) 'The exposure advantage: early exposure to a multilingual environment promotes effective communication', *Psychological Science*, 26(7), pp. 1090–1097.
- Grosjean, F. (1989) 'Neurolinguists, beware! The bilingual is not two monolinguals in one person', *Brain and Language*, 36(1), pp. 3–15.
- Hopkyns, S., Zoghbi, W. and Hassall, P.J. (2021) 'The use of English and linguistic hybridity among Emirati millennials', *World Englishes*, 40(2), pp. 176–190.
- Lowe, C.J., Cho, I., Goldsmith, S.F. and Coyle, J.T. (2021) 'The bilingual advantage in children's executive functioning is not related to language status: a meta-analytic review', *Psychological Science*, 32(7), pp. 1115–1146.
- Nichols, E.S., Wild, C.J., Stojanowski, B., Battista, M.E. and Owen, A.M. (2020) 'Bilingualism affords no general cognitive advantages: a population study of executive function in 11,000 people', *Psychological Science*, 31(5), pp. 548–567.
- Paap, K.R. and Greenberg, Z.I. (2013) 'There is no coherent evidence for a bilingual advantage in executive processing', *Cognitive Psychology*, 66(2), pp. 232–258.
- Wendler, C., Veselovsky, V., Monea, G. and West, R. (2024) 'Do llamas work in English? On the latent language of multilingual transformers', *Proceedings of ACL 2024*, pp. 15366–15394.
- Yoo, H., Yang, Y. and Lee, H. (2025) 'Code-switching red-teaming: LLM evaluation for safety and multilingual understanding', *Proceedings of ACL 2025*, pp. 13392–13413.
- Zhang, R., Cahyawijaya, S., Cruz, J.C.B., Winata, G.I. and Aji, A.F. (2023) 'Multilingual large language models are not (yet) code-switchers', *Proceedings of EMNLP 2023*, pp. 12567–12582.
- Zhang, C., Liu, Q., Cheng, F., Jiang, J., Wan, Z., Chu, C., Murawski, Y. and Kurohashi, S. (2025) 'What language do non-English-centric large language models think in?', *Findings of ACL 2025*, pp. 26333–26346.