Supplementary information

Language is primarily a tool for communication rather than thought

In the format provided by the authors and unedited

Supplementary Information for "Language is primarily a tool for communication rather than thought"

Evelina Fedorenko (MIT), Steven T. Piantadosi (UC Berkeley), Edward A.F. Gibson (MIT)

Corresponding author: Ev Fedorenko (evelina9@mit.edu)

SI-1. Six reasons why many find the hypothesis that language mediates thought—the hypothesis that we argue against in our piece—appealing.

1) The desire for biological discontinuity between humans and other animals.
The idea that language mediates thought (i.e., that language endowed humans with a new representational format) more straightforwardly allows for a potential discontinuity between non-human animals and humans.

2) The desire for a one-factor explanation of differences between humans and non-human animals.
The hypothesis that language gave rise to our ability to think is more parsimonious than the hypothesis whereby language and thinking independently increased in their sophistication over the course of human evolution.

3) The inner speech phenomenon.
Many have a sensation of a voice in their heads when they think—what is often referred to as *inner speech* (Fernyhough & Borghi, 2023). Attempts to systematically assess the presence of inner speech have revealed that the phenomenon is not universal: a substantial fraction of the population does not experience this phenomenon (Roebuck & Lupyan, 2020) and even those who do, only report engaging in inner speech some of the time (Heavey & Hurlburt, 2008; Alderson-Day & Fernyhough, 2015). This evidence already rules out a strong version of the idea that inner speech is critically necessary for thinking. However, given that for many, engaging in inner speech appears to be a common experience, this phenomenon deserves further investigation.

4) The intuition that linguistic explanations crystallize thinking.
Explaining (teaching) an idea to someone, including casting it in a linguistic format, can help crystallize the idea, sharpen the argumentation, and lead to an increased level of understanding (Lombrozo, 2006). Whether linguistic formulation specifically is critical to these effects remains unclear, but even if it is, such effects are unlikely to be mediated by the core language-processing mechanisms that support lexical access and syntactic structure building—the mechanisms discussed in '*The language network in the human brain*' of this piece. In particular, this intuition applies to argument chains, which connect information across a relatively long, multi-sentence timescale. However, the language network is relatively insensitive to information at this timescale. In particular, the language areas are characterized by a relatively short 'temporal receptive window' (Hasson et al., 2008), only integrating information across the span of up to ~10 words (Pallier et al., 2011; Shain, Kean et al., 2024), and do not show sensitivity to discourse-level structure (Lerner et al., 2011; Blank & Fedorenko, 2020; see Fedorenko et al., 2024 for a review). Nevertheless, these effects deserve more scientific attention. The mechanisms discussed in Cantlon & Piantadosi (2024) could play a role here.

5) The intuition that speaking a particular language (e.g., German or Mandarin) is associated with different ways of thinking.
The idea that the language you speak affects how you think, often referred to as the Sapir-Whorf hypothesis or the Whorfian hypothesis (Casasanto, 2012; Whorf, 1956/2012), has received a lot of attention in the last few decades (e.g., Majid et al., 2004; Casasanto, 2008; Lupyan, 2012; Dolscheid et al., 2013). Empirically, the approach typically involves comparing speakers of

different languages, where the languages differ in whether they have words for a particular concept or whether they mark a particular grammatical distinction, and performance of these different groups is examined on some cognitive task where the distinction is relevant, to test whether access to a certain linguistic representation is associated with better performance. Unfortunately, many experimental findings from this line of work do not appear to be robust to replication (e.g., i) Boroditsky, 2001 – failures to replicate: Chen, 2007; January & Kako, 2007; Tse & Altarriba, 2008; ii) Boroditsky et al., 2003 – failure to replicate: Mickan et al., 2014; iii) Winawer et al., 2007 – failures to replicate: Martinovic et al., 2020; Chen et al., 2023).

Furthermore, some conceptual confusion exists in the literature about the relationship between language and culture, and about the directionality of the putative cross-linguistic differences. In particular, *culture* unquestionably shapes cognition, making certain aspects of the world more salient (e.g., Pitt et al., 2022). These cross-cultural differences are, of course, *reflected* in language given that we use language to convey the contents of our minds. For example, a culture that has no utility for numbers/counting is not going to invent words for exact numbers (e.g., Frank et al., 2008) and a culture that does not have artifacts that only differ in color may not have many color terms (e.g., Gibson et al., 2017). However, cultural demands shaping cognition (and, correspondingly, language) is quite different from the hypothesis that speakers of different languages will think fundamentally differently *because* of some feature of their language, even if cultural considerations are identical.

6) The term "language of thought (LOT)".
One popular hypothesis about the representational format of thoughts is called a "language of thought" (Fodor, 1975). The core idea is that thoughts are built out of smaller atomic pieces, just like programs are built out of a small collection of operations, or sentences are built out of words. Fodor in particular argued that such a compositional, rule-based system would be required to explain the compositionality of our thinking, as well as the systematic patterns of reasoning that we engage in. Modern versions of this theory have found empirical support across many domains (Goodman et al., 2008; Piantadosi et al., 2012; Piantadosi et al., 2016; Amalric et al., 2017; Rule et al., 2020; Dehaene et al., 2022; Sablé-Meyer et al., 2022; Yang & Piantadosi, 2022; Ellis et al., 2023). However, although some have argued that natural language *is* the language of thought (e.g., Chomsky, 1965), the evidence reviewed in the main text does not support this possibility. Instead, it appears that we possess a language of thought—or, more likely, multiple languages of thought (see Mandelbaum et al., 2022 for a recent discussion)—that are linked to, but separable from, our linguistic abilities. In fact, this is how most current computational implementations are structured, where the representational systems hypothesized for the language of thought are distinct from the types of systems that are thought to underlie natural language.

SI-2. Select quotes about the role of language in thinking and cognition (relevant to Box 1).

Chomsky (in Piatelli-Palmorini, 1980):
- *" ... **language** is so **deeply involved in many aspects of thinking**... a substantial part of what we call thinking is simply linguistic manipulation, so if there is a severe deficit of language there will be severe deficit of thought"*
- *" ... thinking is a domain that is quite different from language, even though language is used for the expression of thought, and **for a good deal of thinking we really need the mediation of language**"*

Carruthers (2002):
- *" ... natural language syntax ... is crucially necessary for inter-modular integration... **non-domain specific thinking operates by accessing and manipulating the representations of the language faculty**"*

Gleitman & Papafragou (2012):
- *" ... several credible lines of experimental and developmental evidence suggest **significant influence of linguistic representation during on-line processing in many cognitive and perceptual domains**: Insofar as languages differ in the short-term processing demands that they pose to listeners, interpretational outcomes and styles, including characteristic ambiguity resolution, may look quite different cross-linguistically..."*

Gentner (2016):
- *" ... **language** serves as a cognitive tool kit that **allows us to represent and reason in ways that would be impossible without such a symbol system**"*

Lupyan (2016):
- *" ...The unique **design features of language** such as its categorical and discrete format **allow it to augment basic cognitive and perceptual processes**. ... Normal human cognition is language-augmented cognition."*

References for SI:

Alderson-Day, B., & Fernyhough, C. (2015). Relations among questionnaire and experience sampling measures of inner speech: a smartphone app study. *Frontiers in Psychology*, *6*, 517.

Amalric, M., Wang, L., Pica, P., Figueira, S., Sigman, M., & Dehaene, S. (2017). The language of geometry: Fast comprehension of geometrical primitives and rules in human adults and preschoolers. *PLoS Computational Biology*, *13(1*), e1005273.

Blank, I. A., & Fedorenko, E. (2020). No evidence for differences among language regions in their temporal receptive windows. *NeuroImage*, *219*, 116925.

Boroditsky, L. (2001). Does language shape thought?: Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, *43*(1), 1-22.

Cantlon, J. & Piantadosi, S. (2024). Uniquely human intelligence arose from expanded information capacity. *Nature Reviews Psychology*.

Carruthers, P. (2002). The cognitive functions of language. *Behavioral and brain sciences*, *25*(6), 657-674.

Casasanto, D. (2008). Who's afraid of the big bad Whorf? Crosslinguistic differences in temporal language and thought. *Language learning*, *58*, 63-79.

Casasanto, D. (2012). Whorfian hypothesis. In J. L. Jackson, Jr. (Ed.), *Oxford Bibliographies Online: Anthropology*. Oxford: Oxford University Press. doi:10.1093/OBO/9780199766567-0058.

Chen, J. Y. (2007). Do Chinese and English speakers think about time differently? Failure of replicating Boroditsky (2001). *Cognition*, 104(2), 427-436.

Chen, S., Conway, B. and Gibson, E. (2023). Russian blues do not reveal an effect of language on perception. Poster presented at the Human Sentence Processing conference, University of Pittsburgh, March 2023.

Chomsky, N. (1965). Aspects of the theory of syntax. *Special technical report no. 11* (No. NASA-CR-79495).

Dehaene, S., Al Roumi, F., Lakretz, Y., Planton, S., & Sablé-Meyer, M. (2022). Symbols and mental programs: a hypothesis about human singularity. *Trends in Cognitive Sciences, 26*(9), 751-766.

Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The thickness of musical pitch: Psychophysical evidence for linguistic relativity. *Psychological Science*, *24*(5), 613-621.

Ellis, K., Wong, L., Nye, M., Sable-Meyer, M., Cary, L., Anaya Pozo, L., ... & Tenenbaum, J. B. (2023). DreamCoder: growing generalizable, interpretable knowledge with wake–sleep Bayesian program learning. *Philosophical Transactions of the Royal Society A, 381*(2251), 20220050.

Fedorenko, E., Ivanova, A.I. & Regev, T.I. (2024). The language network as a natural kind in the human brain. *Nature Reviews Neuroscience.*

Fernyhough, C., & Borghi, A. M. (2023). Inner speech as language process and cognitive tool. *Trends in Cognitive Sciences*, *27*(12), 1180-1193.

Fodor, J. A. (1975). The Language of Thought. Crowell, New York.

Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition*, *108*(3), 819-824.

Gentner, D. (2016). Language as cognitive tool kit: How language supports relational thought. *American psychologist*, *71*(8), 650.

Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., ... & Conway, B. R. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences*, *114*(40), 10785-10790.

Gleitman, L. R., & Papafragou, A. (2012). New perspectives on language and thought. In K. Holyoak and R. Morrison, eds. (2012), *Cambridge Handbook of Thinking and Reasoning.* 2nd ed. New York: Oxford University Press.

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science*, *32*(1), 108-154.

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *Journal of Neuroscience, 28*(10), 2539-2550.

Heavey, C. L., & Hurlburt, R. T. (2008). The phenomena of inner experience. *Consciousness and Cognition*, *17*(3), 798-810.

January, D., & Kako, E. (2007). Re-evaluating evidence for linguistic relativity: Reply to Boroditsky (2001). *Cognition*, *104*(2), 417-426.

Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *Journal of Neuroscience*, *31*(8), 2906-2915.

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences*, *10*(10), 464-470.

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology*, *3*, 54.

Lupyan, G. (2016). The centrality of language in human cognition. *Language Learning*, *66*(3), 516-553.

Majid, A., Bowerman, M., Kita, S., Haun, D. B., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences, 8(*3), 108-114.

Mandelbaum, E., Dunham, Y., Feiman, R., Firestone, C., Green, E. J., Harris, D., ... & Quilty-Dunn, J. (2022). Problems and mysteries of the many languages of thought. *Cognitive Science, 46*(12), e13225.

Martinovic, J., Paramei, G. V., and MacInnes, W. J. (2020). Russian blues reveal the limits of language influencing colour discrimination. *Cognition*, *201*, 104281.

Pallier, C., Devauchelle, A. D., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences*, *108*(6), 2522-2527.

Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2012). Bootstrapping in a language of thought: A formal model of numerical concept learning. *Cognition*, *123*(2), 199-217.

Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2016). The logical primitives of thought: Empirical foundations for compositional cognitive models. *Psychological Review*, *123*(4), 392.

Piattelli-Palmarini, M. (1980). Language and learning: the debate between Jean Piaget and Noam Chomsky.

Pitt, B., Carstensen, A., Boni, I., Piantadosi, S. T., & Gibson, E. (2022). Different reference frames on different axes: Space and language in indigenous Amazonians. *Science Advances*, *8*(47), eabp9814.

Roebuck, H., & Lupyan, G. (2020). The internal representations questionnaire: Measuring modes of thinking. *Behavior Research Methods*, *52*, 2053-2070.

Rule, J. S., Tenenbaum, J. B., & Piantadosi, S. T. (2020). The child as hacker. Trends in cognitive sciences, 24(11), 900-915.

Sablé-Meyer, M., Ellis, K., Tenenbaum, J., & Dehaene, S. (2022). A language of thought for the mental representation of geometric shapes. *Cognitive Psychology*, *139*, 101527.

Shain, C.*, Kean, H.*, Lipkin, B., Affourtit, J., Siegelman, M., Mollica, F., & Fedorenko, E. (2023). Graded sensitivity to structure and meaning throughout the human language network. *bioRxiv*, 2023-06.

Tse, C. S., & Altarriba, J. (2008). Evidence against linguistic relativity in Chinese and English: A case study of spatial and temporal metaphors. *Journal of Cognition and Culture*, *8*(3-4), 335-357.

Whorf, B. L. (1956/2012). Language, thought, and reality: Selected writings of Benjamin Lee Whorf. MIT press.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., and Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*(19):7780–7785.

Yang, Y., & Piantadosi, S. T. (2022). One model for the learning of language. *Proceedings of the National Academy of Sciences*, 119(5), e2021865119.