1-1-2021

From Goethe to Godel: Against the Language of Thought Hypothesis

Berit Turnquist
University of Mississippi

FROM GOETHE TO GODEL: AGAINST THE LANGUAGE OF THOUGHT HYPOTHESIS

Thesis Presented for the Designation of the Master of Arts Degree

Department of Philosophy and Religion

Berit Turnquist

May 2021

Abstract:

This paper re-examines the language of thought hypothesis by considering objections raised by Johann Wolfgang von Goethe against influential views about the relation of language and thought in the late 18th and early 19th centuries, such as those posited by Herder, Schleiermacher, Schlegel, and von Humboldt. Goethe's *Theory of Colors* contains an instructive critique of the idea held by many of his contemporaries: that the bounds and limits of thought are linguistic in character. I argue that Goethe's remarks anticipate later 20th-century challenges to the language of thought hypothesis regarding implicit cognition, such as Dennett's "chess playing" example, as well as Gödel's discussions of the issue of formal incompleteness.

Table of Contents

From Goethe to Gödel: Against the Language of Thought Hypothesis

Berit Turnquist

The language of thought hypothesis is one of the strongest theoretical contenders at present for explaining the cognition-language relationship, but it does not deliver the precision or completeness that it promises in certain important cases. The language of thought hypothesis works well for mapping the propositional parts of cognition and for tracking certain kinds of implicit cognitive functions or transitions, such as deductive inferences and the cognitive sequencing that precedes the formation of propositional representations. However, the language of thought hypothesis falls short in its ability to account for the flickers of liminal cognition that are as subtle and abstract as they are impactful for the thinker's disposition and overall mental state. The language of thought hypothesis can account linguistically for the cognitive events involved in stating a logical proposition, but it cannot account for the cognitive elements involved in emotions, the subtle and instantaneous unfurling of memories, or the perceptual response to the colors in a painting or the early morning sky. If the language of thought hypothesis cannot account for these cognitive elements, then it cannot constitute a complete account of human cognition.

The view that total precision and completeness are not possible when it comes to mapping cognition linguistically is shared by Johann Wolfgang von Goethe, as evidenced by his discussion of the cognition-language relationship in his 1810 *Theory of Colors.* I use Goethe's theory as a launching pad for my argument because he offers in his color theory a penetrating critique of one of the core theses of the language of thought hypothesis, a hypothesis that was shared by influential

thinkers of his own time: that the scope and bounds of thought are linguistic in character.[1] I intend to show that the critique present in Goethe's theory is one that anticipates the ontological and conceptual objections leveled against the language of thought hypothesis by 20th-century thinkers Kurt Gödel and Daniel Dennett: Gödel by proving that the components of a linguistic or arithmetic system cannot be used to prove the consistency of the same system, and Dennett by mounting a direct conceptual criticism against the language of thought hypothesis.

My argument will begin with an exposition of the language of thought hypothesis that delineates its characteristics as well as its strengths. Next, I will set the stage of the thought-language theory debate as it arose in the late 18th and early 19th centuries by outlining Goethe's critiques of early thought-language theories, as well as the positions of notable contemporaries of Goethe who did not share his concerns. I will then explain the two types of objections I will level against the language of thought hypothesis - conceptual and ontological - before arguing that both of these critical angles are present in the objections levelled by Gödel and Dennett against the language of thought hypothesis and are foreshadowed by Goethe in his theory. I will thus trace the objections surrounding the language of thought hypothesis from the early 19th century to the 20th century, utilizing arguments from contemporary neuroscience, philosophy, and mathematics.

[1] Such hypotheses about the linguistic limitations on thought have a historical connection in virtue of influencing Noam Chomsky, and thereby the language of thought hypothesis. Chomsky references Wilhelm von Humboldt (who features in section 2 of this paper) in his book *Cartesian Linguistics* (Chomsky, Noam. *Cartesian Linguistics: A Chapter in the History of Rationalist Thought.* New York: Harper & Row: (2009).)

1. The Language of Thought Hypothesis, Considered

While theorising about the existence of a mental language can be traced back to medieval times, Jerry Fodor's 1975 work *The Language of Thought*[2] marked a resurgence in popularity of this line of inquiry in contemporary philosophy. Fodor postulates that the human mind works by using a representational system similar in content and structure to ordinary language,[3] and that cognitive content is a product of the systematic combination of word-like representational units in a process reminiscent of computing behavior. The existence of systematic and interpretable mental "language" was suggested by several earlier thinkers, but I will focus on the most recent significant manifestation of the mental language concept, namely, the language of thought hypothesis itself. According to the language of thought hypothesis, such "mentalese" (as Fodor calls it) serves as the vehicle for all cognition and largely consists of propositional attitudes or representations that are linguistic in form and structure. Propositional attitudes and representations are various kinds of mental states or events (such as beliefs, desires, judgements, intentions etc.) whose truth-conditions, success-conditions, and/or cognitive values consist of or are expressible in terms of propositions. Folk psychology describes such mental representations using propositional attitude reports such as:

"*X* thinks that *P"*

"X is afraid that *P*"

[2] J. A. Fodor, *The Language of Thought*. Harvard University Press, 1975.

[3] For the purposes of this paper, I take "ordinary" language to mean written or spoken, external expressions of language, in contrast with mental language.

"*X* desires that *P*" and so on.[4]

Something like "That glass is filled with water" expresses a proposition, as would "Pineapple on pizza tastes bad," whereas "I want black olives on my pizza" expresses a propositional attitude. Propositional attitude reports provide a formal template for describing the mental state of the thinker *X*. For *X* to want or hope or fear *P* is for *X* to have a mental representation whose subject is *P,* and that positions *P* as something worth fearing, wanting, or hoping for.

Tokens, another critical component of the language of thought hypothesis' framework, can be conceived of as pieces of the thought process that work, roughly, as causal mental "events" or impulses leading to propositional attitude reports that describe the type of cognitive state. Natural language uses types (general templates or types of sentences) and tokens (discrete instances of a given type of sentence), and according to the language of thought hypothesis, cognition works in the same way: concrete tokens (instances of types) in the mind involve tokening the word-like units of mentalese and combining them in the right way. One set of considerations in favor of the language of thought hypothesis holds that thought representations, like sentences, are compositional. That is, just as an infinite number of sentences can be formed in natural language using the syntax and rules of language, so the language of thought hypothesis claims to explain how mentalese concepts can be combined with syntactical and logical rules to create an infinite number of thoughts.

The debate about whether, and how well, the language of thought hypothesis works as an adequate theory of cognition stretches back to the 19th century, with degrees of confidence in the hypothesis' success as a comprehensive cognitive theory varying widely between theorists. Fodor

[4] There are other kinds of attitude descriptions in folk psychology that take objects rather than propositions as their contents. E.g., "*X* fears *P"* or *"X* wants *P."* These are not propositional, but the language of thought hypothesis can still accommodate these kinds of reports by appealing to the subject having a certain psychological relation to the concepts that are constituents of the individual mental representations that make up mentalese.

himself asserts in his book *The Mind Doesn't Work That Way* that to assume that the language of thought hypothesis is a complete account of the truth, or the "whole story" about the contents of cognition, would be a mistake.[5] There are a number of theorists (particularly die-hard computational theory of mind proponents[6]) who consider the language of thought hypothesis to be a sufficient explanation for the content and bounds of cognition. However, the ranks of those skeptical about the language of thought hypothesis are numerous, and these skeptics began levelling objections to the hypothesis long before Fodor published his self-attenuating remarks in *The Mind Doesn't Work That Way.*

[5] Fodor, Jerry A. *The Mind Doesn't Work That Way*. Cambridge, MA: MIT, 2001.
[6] Ibid.

2. Goethe versus Early Thought-Language Theories

While the language of thought hypothesis (hereafter LTH) is a 20th-century theory, it has roots in earlier views. For instance, it accepts a version of the thought-language principle, according to which the structure of a language directly impacts its speaker's worldview. The thought-language principle (sometimes called the linguistic relativity hypothesis) became influential in the 19th century through the work of figures such as Herder, Schleiermacher, Schlegel, and von Humboldt.[7] In fact, Schleiermacher goes so far as to *identify* thought with inner speech.[8] Humboldt is particularly important in this regard because of his influence on LTH via Noam Chomsky, as well as being a notable proponent of the representational theory of thought. This view, that cognition is impossible without language and language components,[9] has had a long-standing influence on LTH. The representational theory of thought holds that anything that can be captured in mental representations, propositional attitudes, causal links, or tokening sequences can also be spoken about coherently, and that the components of cognition are only possible insofar as they attach to a descriptive vehicle. In other words, in Humboldt's perspective, all that is able to be thought must also be able to be said.

Humboldt's other, related, contribution to linguistics lies in his view that language and linguistic forms are the result of generative processes, rather than being discrete linguistic entities.

[7] Michael N. Forster. "Language." Essay in *The Cambridge History of Philosophy in the Nineteenth Century (1790-1870)*, 263–92. Eds. Allen W. Wood and Susan Hahn, Cambridge: Cambridge University Press, 2018.
[8] Forster, 269.
[9] Mueller-Vollmer, Kurt and Markus Messling, "Wilhelm von Humboldt," *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), Edward N. Zalta (ed.)

For Humboldt, the grammatical rules and dictionary of a given language are "hardly even comparable to its dead skeleton" in that they do very little when compared to the fluid process by which the mind, in simultaneously reflecting upon and synthesizing, labeling, and differentiating stimuli and sensory experiences, distinguishes itself as a separate "thing" from the person experiencing and reflecting upon the stimuli at hand. Without stimuli and sensation, there is no thought, and it is the task of the medium of language to organize, segment, and label significant parts of the endless sensory input in a way that has cognitive significance. Interestingly, although Humboldt's intuitions were shared by Herder, Schleiermacher, and Schlegel, among others, and had precedents in the work of such notable figures as Kant and Fichte, one of Humboldt's contemporaries (and indeed, a good friend of his), Goethe, took issue with the picture of thought, stimuli, and language put forward by him and other proponents of representational theories of thought.

The theory outlined by Goethe in his 1810 *Theory of Colors* is known for its objections to the empirically rigorous claims of Isaac Newton about refraction, light, and color (taken primarily from Newton's 17th century works *Opticks* and *Of Colours,)*[10] but it also points to an important philosophical question about the cognition-language relationship that cries out for explanation, and it sows the seeds of skepticism about the thought-language principle that anticipate the objections later made by 20th century thinkers. After responding to and challenging Newton, Goethe gives an account of the emotional, moral, and connotative effects that each color has on the mind; an account in which primary colors and a variety of secondary color combinations are described in

[10] Goethe openly disputes Newton's ideas in *Theory of Colors,* which was published about a century after Newton's two listed landmark works on (broadly) light and color. Most notably, Goethe claims that Newton's theory of optics, diffraction, light, and color only holds true in special cases, while his own proposed color theory is applicable in general cases and in a wider variety of circumstances. For the purposes of this paper, I have chosen to circumvent discussion of the many different objections and qualifications that Goethe levels against Newton's prismatic theory.

terms of their impressions on human cognition. Yellow is serene and lively, red stimulates feelings of gravity and nobility, green is peaceful and calming, lilac evokes shade and coldness, and so on. According to Goethe, colors leave distinct impressions on the minds of the people who see them, and they have inherent imaginative and moral connotative qualities. Goethe raises concerns about the difficulty of describing such abstract cognitions in ordinary language early in *Theory of Colors*, beginning in the preface:

> Every act of seeing leads to consideration, consideration to reflection, reflection to combination, thus it may be said that in every attentive look on nature we already theorise. But in order to guard against the possible abuse of this abstract view, in order that the practical deductions we look to should be really useful, we should theorise without forgetting that we are so doing, we should theorise with mental self-possession, and, to use a bold word, with irony.[11]

> And thus as we descend the scale of beings Nature speaks to other senses - to known, misunderstood, and unknown senses: so speaks she with herself and to us in a thousand modes. To the attentive observer she is nowhere dead nor silent (...) However manifold, complicated, and unintelligible this language may often seem to us, yet its elements remain ever the same.[12]

From the first pages of his manuscript, it is clear that Goethe does not want his color theory to be classified as an empirical schema, but rather as a pragmatic meta-investigation into the

[11] Goethe, xxi.
[12] Goethe, xviii.

impressions that colors have on an individual's mind and the way that the individual perceives colors. Goethe urges the reader to hold the investigations and concepts he describes with a light touch - with a sense of practicality and curiosity - so as not to pollute the abstractness of the most compelling parts of his theory with too much concern for rigid categorization. Later in *Theory of Colors,* Goethe laments the tendency of language to fail to capture exactly what occurs in the speaker's (or writer's) mind upon perceiving colors, and he emphasizes that avoiding unwarranted certainty is essential when it comes to discussions of abstract concepts. He also cautions against the introduction of too many specialized and technical words, lest the understanding and imagination of the speaker be stunted by excessive, overreaching specificity:

> Too many specific terms have been adopted; and in seeking to establish new definitions by combining these, the nomenclators have not reflected that they thus altogether efface the image from the imagination, and the idea from the understanding. Lastly, these individual designations of colours, employed to a certain extent as elementary definitions, are not arranged in the best manner as regards their respective derivation from each other: hence, the scholar must learn every single designation, and impress an almost lifeless but positive language on his memory.[13]

> We never sufficiently reflect that a language, strictly speaking, can only be symbolical and figurative, that it can never express things directly, but only, as it were, reflectedly. This is especially the case in speaking of qualities which are only imperfectly presented to observation, which might rather be called powers than objects, and which are ever in

[13] Goethe, 246.

> movement throughout nature. They are not to be arrested, and yet we find it necessary to describe them; hence we look for all kinds of formulae in order, figuratively at least, to define them.[14]

Here, Goethe states that there are certain qualities in reality that cannot be "arrested" via descriptive language and therefore can only be imperfectly or figuratively defined. In subsequent sections, Goethe asserts that metaphysical explanations of the linguistic-cognitive relation tend to be vague, whereas mathematical accounts run the risk of being limited and inflexible. He sums up the problem with cognition and language like this: "[H]ow difficult to keep the essential quality still living before us, and not to kill it with the word."[15] Goethe recognizes how difficult it is to make statements about the vivid perceptions and judgements in one's mind without rendering these thoughts bland and truncated when expressed in ordinary language.

> From henceforth everything is gradually arranged under higher rules and laws, which, however, are not to be made intelligible by words and hypotheses to the understanding merely, but, at the same time, by real phenomena to the senses.[16]

Goethe pauses in the middle of a conversation about the arrangement and categorization of sensory qualities of nature to point out that phenomenal perception and cognition in response to natural stimuli often outpace language's capacity to describe. While Goethe's remarks on language only implicitly criticize the key ideas of the language of thought hypothesis, they are nonetheless

[14] Goethe, 301.
[15] Goethe, 302.
[16] Goethe, 72.

incisive.[17] As we will see in the following sections, Goethe's intuitions about language and cognition are strengthened and justified by 20th-century iterations of similar objections that address conceptual and ontological shortcomings of the language of thought hypotheses.

[17] Late in *Theory of Colors,* there is a shift in tone from the "ironic," free curiosity that Goethe espouses in earlier passages to an implicit demand that the reader treat Goethe's account of each color's effect on the mind as fact, and that each color-to-mental-effect relationship be treated as a strict, categorical relationship. Why, after 300 pages of criticizing Newton's highly empirical methodology, does Goethe embrace a similar rigidity for talking about his catalogue of color perceptions? This puzzling inconsistency may be the result of Goethe's lack of access to a comprehensive framework for talking about, and thus consistently rejecting, the language of thought hypothesis. However, this question is outside the scope of this paper, and will need to be considered at more length in a different project.

3. Gödel versus Completeness

Before I transition into a discussion of more recent criticisms of the language of thought hypothesis, I will broadly outline the methodology by which I will object to LTH. My primary objection to LTH is twofold: I think that LTH fails both ontologically and conceptually in important cases. The conceptual objection involves arguing that an important claim that LTH holds to be true is actually false, thus showing that any discourse about cognition and language that relies on this claim is incoherent. Goethe foreshadows the conceptual objection to LTH in his claims that certain important cognitions, especially those that are perceptual in nature, escape language's power to describe them, thus calling a central tenet of LTH into question. The ontological angle will demonstrate that there exists no sufficiently exhaustive lexicon of linguistic signs and components that can hope to capture both the full complexity and the transience of human cognition. The ontological critique provides justification for the claim that some cognitions are simply not description-compatible. The conceptual objection is found in both Dennett's and Gödel's critiques, as well as in Goethe's. The ontological criticism (and the thought experiment that I will propose) follow closely from Goethe's remarks that there will always be a margin of descriptive error when translating cognition into language, and that, in fact, there is no language with enough words to avoid the necessity for figurative, indirect description in some important cases.

The first iteration of my conceptual objection stems from the basic tenets of Kurt Gödel's second incompleteness theorem, and it challenges the compositionality requirement central to the

LTH, a requirement that was shared by Humboldt and other 19th-century proponents of representational theories of thought. In the simplest terms, Gödel's second incompleteness theorem holds that the consistency of arithmetic systems cannot be proven by arithmetic itself. Given that arithmetic is a language, why assume that ordinary human language is any different in regard to the standards that it would be held to in attempting a reflexive proof? The second incompleteness theorem (and hence, the cognitive-linguistic application of the theorem) is built on the following basic idea: an arithmetically true statement, ρ, can be proven using a deductive sequence that follows from one or more basically true axioms. However, outside of these axioms and deductive sequence components, there remains another deductively true arithmetic statement, σ, that isn't included in the deductive sequence or set of axioms used to prove ρ. One could place σ inside of the set of axioms and repeat the deductive sequence in an attempt to achieve a more complete end result sentence, but the problem from the initial instance remains. No matter how many true statements are added to the body of axioms used in the deductive sequence, there will always remain a further true statement that cannot be deduced by the extant collection of axioms and true statements. According to the second incompleteness theorem, once you codify something, you make it so that it can never be complete; in other words, it will always be an open system. A linguistic version of the incompleteness theorem would assert that the completeness and consistency of a linguistic system cannot be fully proven using only itself (language), thus supporting my conceptual objection to the idea that it is possible to exhaustively catalogue cognition using the same purported language components.

An objector to my first conceptual objection a la Gödel's second incompleteness theorem might remind me that linguistic incompleteness is only a problem as long as there is only one primary language at play. It is certainly possible to conceive of a higher-order language that could

completely catalogue and prove the consistency of ordinary language. However, my worry with this suggestion is one of vicious regress: the use of a higher-order language may be able to externally prove the consistency and completeness of an ordinary, primary language, but the completeness of the higher-order language itself would remain unproveable until yet another higher-order language is employed to prove the consistency of the "first" higher order language, and so on.[18]

My second criticism of the language of thought hypothesis follows closely from the issue of formal completeness, in that it asserts that translating abstract processes and entities into language will result in something important being left out. This ontological objection against LTH's claimed ability to account for all of the important parts of cognition follows from Goethe's worries about how to speak of what is thought "without killing it with the word," as well as from Gödel's notion of incompleteness. The ontological claim that I intend to make is that there simply are not enough words in existence to represent all of the important details of cognition. While it is true that a given language may have more words for a particular thing than another, this often pairs with generalizations in other areas of the language.[19] It may be true that the quantity of words in a given language is only a correlative indicator for its ability to capture complex cognitive states, but the sheer number of things that could possibly be described and the fact that every language reflects the specific limitations and needs of its parent person-group implies that limited-ness is endemic to all languages.

[18] Further, as far as the language of thought hypothesis goes, turning to a higher-order language to make mentalese complete is incoherent precisely because LTH, by hypothesis, constitutes the bounds and limits of human thought.
[19] For a more thorough discussion of the way that language and thought interplay in different linguistic contexts, Steven Pinker's book *The Stuff of Thought* (Pinker, Steven. *The Stuff of Thought: Language as a Window into Human Nature*. London: Penguin Books, 2010) includes a fascinating discussion of the Sapir-Whorf hypothesis (linguistic determinism, linguistic relativity) and its problems, contrasted with the conceptual semantics picture of cognition and language.

It is theoretically possible to imagine outsourcing the recall and contextualization of words to a powerful, ever-expanding database designed to collect and disseminate information about words and their definitions as quickly as they are invented, but this thought experiment doesn't assuage my concerns. For one, at this point in human history, there exists no such technology, so at least for now the ontological claim holds, even if it might fail in practice depending on what happens in the future. Second, I fear that even the existence of an unimaginably powerful word and meaning database wouldn't do away with an infinite regress worry: no matter how many words are invented and recorded, despite this ever-growing multiplicity of terms, there will always, always remain something that could yet be felt or cognized for which a word needs to be invented.

Information theory, which involves the storage, content, quantification, and communication of information, holds that the more possible outcomes and/or variables there are in a given system, the higher the degree of entropy, or, the higher the uncertainty of the value of a random variable component.[20] Goethe references a similar notion in *Theory of Colors.* He remarks that the presence of excessive categorizing terminology stunts the imaginative pulse that keeps language alive and amenable to being understood, rather than being memorized by rote. In other words, according to Goethe, the endless invention and assignment of highly specific words for every single possible perception will increase confusion and reduce genuine understanding, rather than lend precision and robustness. My ontological objection to LTH follows a similar vein, and it should weaken LTH's central ontological claim, which is that there *are* enough words and symbols in existence to account accurately and with certainty for all of the important details of cognition. In order to show, *a posteriori,* that my ontological objection has legs, I propose a short thought experiment.

[20] Burnham, K. P. and Anderson D. R. (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, Second Edition* (Springer Science, New York)

Imagine that you and a friend visit a museum. Upon entering a gallery of Impressionist works, you encounter Van Gogh's "Starry Night." Your companion asks, "What do you think of this one?" In response, maybe you say something as simple as, "All that blue makes me feel a little sad," or something as complex and precise as, "The blue in this particular area of the painting makes me feel happy and sad at the same time, melancholic, bittersweet, if you will. This particular shade of blue reminds me of the walls of my maternal grandmother's kitchen in the evening between the hours between 4pm an 6pm in the late Autumn, just as the sun was starting to sink below the horizon. I am thinking about Sunday afternoons when she was preparing my grandfather's lunch for work the next day, and I knew I'd have to go home in an hour or two so I could get to school early in the morning. She wouldn't turn the lights on in the house until my grandfather would come in and find her hunched over the stove or the cutting board, squinting in the dying light, and insist on turning on a lamp for her."

Even if this detailed description could theoretically be explained in terms of explicit propositional attitudes, mental representations, and tokening sequences, something is still missing, something that LTH fails to account for. As detailed as the above narrative description is, one can easily imagine it being many, many more times as descriptive. You might mention other colors, smells, and specific memories within the same blue-wallpapered kitchen, ages of the characters in the memory, and any other factors that inform the sweet but melancholy feeling elicited by that particular Van Gogh blue, and so on, perhaps infinitely. If you and your friend sat in front of "Starry Night" from the minute the museum opened until the minute it closed, with you explaining all the small and subtle details of your cognitive experience of blue in detail and your friend asking clarifying questions throughout your explanation, you and your friend would still have two utterly distinct mental pictures of the scene. In the course of telling this long story,

other flashes of memories or impulses or emotions might whisk across your mind, perhaps when you say the word "grandfather" or "school the next morning." These flashes might go unnoticed, or you might simply forget to mention them after explaining all the details that came before them in your cognitive chain, but they nonetheless impact your disposition and spark further cognitive impulses. Besides the problem of detail, there is also the problem that speaking about cognitions requires that we make them linear in order to make them comprehensible, even though cognitions themselves are not linear, but rather web- or spoke-like in the way that parent thoughts spark offshoot thoughts and impulses. As such, cognitions are not always amenable to precise linear description. In short, LTH's claim that "everything that can be thought must be able to be said" fails when taking into account the infinitely rich and complex world of salient qualities that *could* be said, but are subconscious, or difficult to put into words that would be understandable to a conversation partner, or that are forgotten quickly but still leave a trace on the cognitive experience.

The defender of LTH might respond to my ontological objection by saying something like, "There actually *are* enough words in language to account for every single minute detail of cognition in a way that can be categorized under one of the acceptably explicit categories recognized by LTH and representational theory of thought theorists, but it would be impractical to do so. It wouldn't be impossible by any means; it would just take a long time and more detailed explanation than anyone has the patience for."

This objection highlights the important distinction between impracticality and metaphysical impossibility. However, my position is not that LTH's application for these kinds of questions is impractical; it seeks to recognize with humility the fact that human cognition is neither stable nor fully understood. The scientific community recognizes that cognition as we

understand it is transient, varies depending on situational factors, and involves processes that are partially or completely opaque.[21, 22, 23, 24, 25] LTH claims, without justification, the existence of a precise mapping account of cognition and that this cognitive account is exhaustive. In other words, LTH is an empirical claim without a test.

[21] Keppler, Joachim, and Itay Shani. "Cosmopsychism and Consciousness Research: A Fresh View on the Causal Mechanisms Underlying Phenomenal States" in *Frontiers in Psychology* 11 (2020).
[22] Prettyman, Adrienne, "The persistent problem of targetless thought" in *Consciousness and Cognition* 82, (2020).
[23] Feltz, Bernard, Marcus Missal, and Andrew Sims, eds. *Free Will, Causality, and Neuroscience*. LEIDEN; BOSTON: Brill, 2020.
[24] Murphy, Elliot. *The Oscillatory Nature of Language*, University of Cambridge ESOL Examinations, 2020.
[25] This holds especially true given the tenets of the second incompleteness theorem - completeness is impossible under a codified system.

4. Dennett versus The Language of Thought Hypothesis

Fifty years after Gödel's second incompleteness theorem shook the mathematics world, cognitive scientist and philosopher Daniel Dennett returned to the lingering problem of formal completeness and rule-based generative processes in his 1981 review of the language of thought hypothesis.[26] The second iteration of my conceptual objection to LTH is in keeping with Dennett's in that it suggests that there is good reason for *not* believing that all mental content must correspond directly to a linguistic-representational vehicle in order to be considered real. What I hope to show is that the (central to LTH) assertion that thoughts cannot occur without explicitly being linked with words and signs is incoherent. If the representational theory of thought (RTT) component of LTH is incoherent, then it can be argued that LTH doesn't succeed in doing what it purports to do, which is to provide an account of cognition that links cognition necessarily to speech.

Dennett levels a widely cited objection to the RTT as it relates to LTH, based on a thought experiment involving a chess-playing computer program. Dennett postulates a situation in which an observer of a computer chess-playing program notes that the computer typically puts its queen into play very early, thus deploying a strategy in which the agile queen is chased around the board by the other clumsier or less-free-moving pieces. However, upon investigating the specific coding that determines the computer program's behavior, nothing is found that explicitly orders that the

[26] Dennett, Daniel C., 1977 [1981], "Critical Notice: Review of The Language of Thought by Jerry Fodor", *Mind*, 86 (342): 265–280. Reprinted as "A Cure for the Common Code," in *Brainstorms: Philosophical Essays on Mind and Psychology*, Cambridge, MA: MIT Press, 1981.

queen be played early; it just happens as the natural result of other strategic programming processes and rules. According to Dennett, the link between cognition and speech works in a similar fashion.

> For all the many levels of explicit representation to be found in that program, nowhere is anything roughly synonymous with 'I should get my queen out early' explicitly tokened. (...) I see no reason to believe that the relation between belief-talk and psychological talk will be any more direct.[27]

If something in mentalese can't be represented by either a propositional attitude, mental representation, or tokening process leading up to it, then it seems clear that there must be parts of cognition that are non-explicit, and thus not amenable to being represented mentally in terms of ordinary logical, propositional thought-language. If what LTH says is true, then everything that a human thinks, believes, and does can be linked to explicit explanatory vehicles via ordinary language, much like a computer program's behavior can be traced to an explicit set of codes and rules. As Dennett's thought experiment shows, if even the highly mechanical artificial intelligence of computer programming can end up exhibiting behavior that cannot be traced back to an explicitly coded command, then what reason do we have for believing that everything that the human mind "program" thinks, believes, or does can be traced to an explicit linguistic "code" component?

Garry Kasparov, chess grandmaster and opponent of the Deep Thought chess computer, famously notes that human minds transcend and confound the abilities of even the most sophisticated chess-playing computer program in an unusual way.

[27] Ibid.

> The human mind isn't a computer; it cannot progress in an orderly fashion down a list of candidate moves and rank them by a score down to the hundredth of a pawn the way a chess machine does. Even the most disciplined human mind wanders in the heat of competition. This is both a weakness and a strength of human cognition. Sometimes these undisciplined wanderings only weaken your analysis. Other times they lead to inspiration, to beautiful or paradoxical moves that were not on your initial list of candidates.[28]

In Kasparov's view, human chess players at the highest level work as a counterexample to LTH: yes, they have rules in mind, but the human player's intuitions about the harmony of the pieces, perhaps unspeakable even to the player him or herself, cannot be captured linguistically or codified. Dennett's chess-playing computer case illustrates my conceptual objection against LTH by pointing out that the connection between cognition and language is not always explicit. The unconscious mental use of deductive inferences that aren't explicitly represented in propositional attitudes would be an example of one such non-explicit but nonetheless important part of cognition.

The ideological tension between theorists who hold that all of cognition can be encapsulated by a combination of explicit representations and implicit generative rules and processes and those who are suspicious of attributing all thought to linguistic anchors dates back to the 19th century, with the former view touting Humboldt as its champion. In one of his sixteen theses in *Über Denken und Sprechen,* Humboldt says:

[28] Kasparov, Garry. *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins*. Perseus Books. LLC, 2017.

> The sensory designations of those units, into which certain portions of our thinking are united, in order to be opposed as parts to other parts of a greater whole as objects to the subjects, is called in the broadest sense of the word: language (*Sprache*).[29]

What Humboldt is suggesting is that the continuous flow of cognitive impulses and sensory stimuli can only be transformed into discrete, intelligible parts through language. Without language's mediating, organizing, and prioritizing power, no thought is able to escape and be understood apart from the impenetrable mental stream. Goethe's response to this way of thinking appears to be similar to Dennett's: while Humboldt holds that a linguistic "code" that assigns order to raw cognition is the source of all intelligible thought, Goethe is not convinced of this, as evidenced by his remarks in *Theory of Colors*[30], suggesting that there are intelligible, active parts of thought that cannot be traced back to linguistic coding. In asserting that language sometimes fails at capturing cognitive processes that are subtle and yet intelligible (intuitions, for example, or sensory and perceptual responses to certain stimuli), Goethe indirectly takes aim at Humboldt's claims, as well as subsequent theories (like the language of thought hypothesis) that rest on the assertion that cognition is restricted by language.[31]

---

[29] Mueller-Vollmer, Kurt, "Thinking and Speaking: Herder, Humboldt and Saussurean Semiotics. A Translation and Commentary on Wilhelm von Humboldt's 'On Thinking and Speaking: Sixteen Theses on Language,'" *Comparative Criticism*, 11: 159–214 (1989).

[30] Refer to section two of this paper, "Goethe versus Early Thought-Language Theories."

[31] Goethe's remarks on generative principles are mentioned in Note W, par. 608, at the end of *Theory of Colors.* According to Goethe, the process by which a thing in reality or in the mind can seemingly cause the reaction, repulsion, or inception of a new, separate thing or several things (whether it be a chain reaction of thoughts, decisions, physical or chemical processes, etc.) remains mysterious. Goethe suggests that his contemporaries have not given sufficient attention to understanding the implications of this sort of generative phenomenon.

5. Coda

Goethe is already concerned about the consequences of adopting too rigid a view of the relationship between cognition and language in his neglected *Theory of Colors,* a work that indirectly (though not necessarily unintentionally) responds to the representational theories of thought held by his early 19th-century contemporaries. Gödel presents a mathematical proof against the idea of systemic completeness, and is followed by Dennett who, walking in Goethe's ghostly footsteps, challenges Fodor's claims by denying that a clear and explicit relationship between cognition and linguistic components exists. Gödel's and Dennett's contributions suggest that Goethe's remarks on the language of thought hypothesis have cast a long shadow on the debate regarding the nature of the language-cognition relationship. I have added my voice to this debate by showing that challenges against the language of thought hypothesis, particularly ones that uphold the view that certain cognitions are not description-compatible, are historically significant and ought to remain significant in the interest of preserving openness and curiosity about the enduring mysteries of cognition.

BIBLIOGRAPHY

BIBLIOGRAPHY

Burnham, K. P. and Anderson D. R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, Second Edition* (Springer Science, New York): 2002.

Chomsky, Noam. *Cartesian Linguistics: A Chapter in the History of Rationalist Thought.* New York: Harper & Row: (2009).)

Dennett, Daniel C., 1977 [1981], "Critical Notice: Review of The Language of Thought by Jerry Fodor", *Mind*, 86 (342): 265–280. Reprinted as "A Cure for the Common Code," in *Brainstorms: Philosophical Essays on Mind and Psychology*, Cambridge, MA: MIT Press, 1981.

Feltz, Bernard, Marcus Missal, and Andrew Sims, eds. *Free Will, Causality, and Neuroscience*. LEIDEN; BOSTON: Brill, 2020.

Fodor, Jerry A. *The Language of Thought*. Harvard University Press, 1975.

Fodor, Jerry A. *The Mind Doesn't Work That Way*. Cambridge, MA: MIT, 2001.

Forster, Michael N. "Language." Essay in *The Cambridge History of Philosophy in the Nineteenth Century (1790-1870)*, 263–92. Eds. Allen W. Wood and Susan Hahn, Cambridge: Cambridge University Press, 2018.

Goethe, Johann Wolfgang von, *Theory of Colors*. Translated by Charles Locke Eastlake, London: John Murray, Albemarle Street, 1840.

Kasparov, Garry. *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins*. Perseus Books. LLC, 2017.

Keppler, Joachim, and Itay Shani. "Cosmopsychism and Consciousness Research: A Fresh View on the Causal Mechanisms Underlying Phenomenal States" in *Frontiers in Psychology* 11 (2020).

Mueller-Vollmer, Kurt and Markus Messling, "Wilhelm von Humboldt," *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), Edward N. Zalta (ed.)

Mueller-Vollmer, Kurt, "Thinking and Speaking: Herder, Humboldt and Saussurean Semiotics. A Translation and Commentary on Wilhelm von Humboldt's 'On Thinking and Speaking: Sixteen Theses on Language,'" *Comparative Criticism,* 11: 159–214 (1989).

Murphy, Elliot. *The Oscillatory Nature of Language*, University of Cambridge ESOL Examinations, 2020.

Prettyman, Adrienne, “The persistent problem of targetless thought” in *Consciousness and Cognition* 82, (2020).

VITA

Bethel University, Bachelor of Arts

University of Mississippi, Master of Arts