Linguistic Relativity: A Creature of Habit

In this essay, I will argue that linguistic determinism is thus far an untenable position. After careful consideration of the empirical data that is frequently cited as evidence for a strong linguistic determinist view, I will illustrate that these findings merely comport to two moderate formulations of the linguistic relativity hypothesis. Namely, I will discuss language's capacity to draw attention to features of the world which would otherwise go unnoticed. This is a formulation of the linguistic relativity hypothesis which has been called Trivial Whorfianism (Reines and Prinz, 2009, p.1028) and is akin to the language as spotlight hypothesis (Wolff and Holmes, 2011, p.259). Similarly, I will offer the case for what has been called Habitual Whorfianism (Reines and Prinz, 2009, p.1028), or the language as meddler hypothesis (Wolff and Holmes, 2011, p.256), which is the view that language use instils habits of thought in speakers which meddle or alter the default cognitive or perceptual functioning that would have taken effect without that language's idiosyncratic categorizations and referential systems. For the remainder of this essay I will use the term trivial to refer to the spotlight and trivial Whorfian approaches collectively. Similarly, I will use the term habitual to refer to the meddler and habitual Whorfian approaches described above, as I see these theoretical demarcations to be near identical. I will argue that there is, at best, evidence for only the trivial or habitual perspectives, both being moderate relativist positions, and that any radical linguistic deterministic view has yet to be proved. It is however difficult to separate these two positions as both emphasize language's ability to emphasize or constrain one aspect over another which may not be explicit in the cognitive processes involved in the task. furthermore, both the trivial and habitual hypotheses would allow for effects to be reversed via the formation of new habits or new languages being learned (cf. Kay and Kempton, 1984, p.77).

Linguistic determinism is the idea that languages alter the cognitive and perceptual processes of their speakers. Since languages vary greatly in how they categorize and qualify objects, the speakers of a language supposedly carve the world at its joints in conformity with the rules and tendencies of their native tongues, according to determinists. (cf. Wolff and Holmes, 2011, p.254-

255). Moreover, the way the language is structured is thought to constrain the way in which the speaker can experience the world. The implications of this view are many if determinism holds true, as every language then creates a unique way of encountering and seeing the environment and every language user is destined to approach their world from a unique footing compared to other speakers of dissimilar languages. For example, if my language uses left/right spatial metaphors for talking about time, such that the past is imagined to be to the left of me and the future to the right, I would allegedly be unable to imagine or think about the past as being behind me or the future ahead of me as is the case in English. The extreme definition provided above, has its roots in the Whorfian hypothesis (Halpern, Whorf, Carroll, & Chase, 1956), but such a radical formulation of this hypothesis has yet to find substantial empirical backing. In fact, the "weaker" hypothesis that language merely directs one's attention to specific attributes of reality is argued to reflect the original Whorfian arguments (Athanasopoulos 2009, p.83). There are several challenges to investigating linguistic determinism; the provision of adequate and clear translations to participants who speak a foreign language; methodology and design issues, finding sufficient cognitive and perceptual metrics and the confounds of language-on-language effects, to name but a few. I will illustrate that these challenges have not yet been overcome, and that the confirmatory results ascertained thus far provide only shaky evidence for the trivial and habitual relativist approaches. In the same line of argument, I will defend these moderate linguistic relativist positions against the other counter-position, that language does not affect thought and perception.

Linguistic relativity has been studied in a plethora of domains including objects and substance (Imai and Mazukai, 2003), grammatical gender (Boroditsky, Schmidt, and Phillips, 2003) mathematical reasoning, scene memory and mental calculation to name a few (Bloom and Keil, 2001, p.364), but for the purposes of this essay it will be prudent to begin our investigation with empirical findings that focus on the intrinsic differences between how languages are constructed and their idiomatic devices for abstract thinking (cf. Au, 1983, p.161). For instance, Mandarin Chinese lacks a subjunctive tense, Bloom (1981) investigated whether this difference between Mandarin and English speakers would affect their ability to reason counterfactually (e.g., posing if-then hypotheticals such

as: *if it weren't raining, I would go to the shop*). The study consisted of three versions of a fictional

story about a philosopher written In English or translated to Mandarin. the versions became

increasingly dependent upon a counterfactual interpretation. The second version of the story could be

correctly interpreted counterfactually or factually, yet the third version of the story necessitated a

counterfactual explanation. Bloom concluded that Mandarin speakers, lacking the subjunctive tense as

a resource for their understanding, were almost entirely unable to reason counterfactually, and that

bilinguals fluent in both languages, associated counter factual reasoning with English over their native

Mandarin (Ibid., pp.31-32). As such bloom adopted a strong deterministic interpretation, however,

this conclusion is not without contention. Most notably, Au (1983) observed that the fictional stories

utilized in the experiment were translated inadequately into Mandarin. When the study was replicated

using idiomatic Mandarin, she found no significant difference in reasoning, except that Chinese

speakers may construct counterfactuals through a more complex process (ibid. pp.181-182). This

study highlights the challenge of successful translation in investigating the Whorf hypothesis, and the

need for caution when interpreting results from bilingual participants. This area of research was

revisited more recently by Yeh and Gentner (2005) who found support for a more moderate version of

Bloom's initial claim. As English speakers have syntactic markers for counterfactuals in the form of

the subjunctive, they are more able to readily identify counterfactuals when lacking contextual cues in

non-transparent stories. Whereas: "Chinese speakers may be disadvantaged when counterfactuals

must be detected with respect to specific current context." (P.2415), as they rely on world-knowledge

and context to ground their counterfactual assumptions, there is then the risk of greater inefficiency

and inaccuracy in the detection of counterfactuals in everyday processing as more cognitive resources

are required for Chinese speakers. This is in line with both the trivial and habitual formulations of the

Whorf hypothesis, yet Chinese speakers can still form counterfactuals and their reasoning of them is

only slightly disadvantaged in certain instances, thus a stronger deterministic position, such as

Bloom's (1981), is unfounded.

Language's affect on color perception has also undergone scrupulous investigation, as hue is

dependent on the wavelength of light and wavelength is a continuous discrete dimension ('The

Sorites Paradox', see Davidoff, 2001, p.383), a language's colour categorizations are the ideal starting point for investigating linguistic relativity, as they occupy a fuzzy set where: "there are not obvious discontinuities in nature" (Roberson et al. 2005, p.406). However, many findings have been contradictory. Berlin and Kay (1969) set the stage for a universalist account of colour perception, they gathered colour area samples and their respective category foci (i.e., the best exemplar of that colour area) from 20 different languages and concluded that colour categorisation is not language-specific and that all languages focused their colour boundaries around focal colours which correspond to black, white, red, yellow, green and blue. They further claimed that these focal colours followed an evolutionary course, such that a language with only two colour terms would utilize white and black, as these represent the most basic contrast to allow for light and dark distinctions to be drawn. Beginning with these most rudimentary colours, a language can grow to include a total of 11 foci according to their research (p.2). There were many criticisms of this research, namely, the very small sample size of only 20 participants, all of whom were bilingual English speakers and mostly hailing from industrialized societies. However, the universalist stance found further empirical grounding soon after with Heider's (1972) cross-linguistic study of English and Dani speakers.

The Dani language possesses only two colour terms which reflect the light/dark basic distinction identified by Berlin and Kay (1969; cf. Heider & Olivier, 1972, p.340). Despite this, Heider found no significant differences in memory of colour chips between English speakers and the Dani, moreover, the Dani found it easiest to learn and recognize an additional eight foci colours from English, providing further evidence for Berlin and Kay's evolutionary path (Heider, 1972, p. 17). This thus led Heider and Olivier to suspect that:

> "Not proven, but certainly suggested by the visual shape of the scaling configurations, was the further finding that Dani and American color memory structures were quite similar to each other, although the naming structures were not." (1972, p.350)

4

Such a result is obviously at odds with both a trivial and habitual stance on linguistic relativity (cf. Regier & Kay, 2009, p.439), but this universalist position on colour perception has been challenged more recently with cross-linguistic studies using the Berinmo of Papa Guinea (Roberson, Davies & Davidoff, 2000) and the Himba of South Africa (Roberson et al., 2005). Roberson, Davies and Davidoff (2000) argued for the counter-position based on their cross-linguistic studies with the Berinmo of Papa Guinea, they concluded that language is pivotal in allocating specific regions of colour space with their associated colour terms. They failed to replicate the findings of Heider (1972) which espoused the view that focal colours were more readily remembered in short term memory by both English speaking participants and Dani participants, this is not surprising given that there were many confounds to the original experimental design of Heider's, such as making the test array biased in favour of focal chip selection, moreover, the Dani may have exhibited better recall for focal chips, yet their discriminability was very poor in comparison with the American subjects (Roberson, Davies and Davidoff, 2000, pp.4-5). Not only did they fail to replicate the results for a privileged memory for focal colours, they also found that similarity judgements were not at the focal colours, rather categorical perception straddled the language-defined boundaries: "Color categories arise only when two items become qualitatively dissimilar; these boundaries arise from the language of experience…" (ibid. p.110). This finding is furthered with evidence from another culture which also only possess five basic colour terms, the South African nomadic Himba. Himba participants failed to exhibit categorical perception for the eleven ('universal') colours outlined in English, and also failed to demonstrate categorical perception of the five basic colour exponents of the Berinmo language (Roberson et al., 2005). These studies compound a view more in line with the trivial and habitual hypotheses, and converge to form a more moderate view; that there are universal, physiological constraints on colour perception, but that the category boundaries are largely demarcated by one's language (cf. Athanasopoulos, 2009, p.84; Regier & Kay, 2009, p.442). Developmentally, it has been noted that Himba and English speaking children initially make similarity judgements from universal perceptual grounds before they acquire the ability to name colours, they are thus constrained solely by their neurophysiology. There is then a divergence in their conception of colour categories as they become language bound and culturally determined, the linguistic codes learned from their parents

begin to prime and direct them toward the colours specified by their culture (Roberson et al. 2005, p.406).

While the above studies represent substantive evidence for a relativist view, there are nonetheless issues with their design. Fault can arguably be drawn with their over-reliance on memory and subjective judgement and there remains the possible confound that subjects rely on visual memory as well as verbal labels (cf. Pinker, 1994, p.65). Winawer et al. (2007) sought to bypass these methodological issues in their investigation of linguistic codes as primes. The study hinged on the fact that Russian makes an obligatory distinction between light blue "goluboy" and dark blue "siniy", which is absent in English. The researchers posed the hypothesis that the pre-existing demarcation in Russian would act as a prime for the Russian speaking participants, thus making their selections more accurate and quicker than their English speaking counterparts. Subjects were shown a triad of three coloured squares (one on top, two on bottom) and asked to identify which of the bottom two squares was perceptually identical to the square on top. This objective design boasted the advantage of not relying on participants' memory, using explicit cues that would be hard to misinterpret and the triad allowed for a cross-category and same-category condition as the bottom tiles could be from the light blue or dark blue dimensions or both be from the same shade (ibid. p.7780). Furthermore, there were additional interference tasks which tested the effect of a verbal and spatial distractor, the verbal distractor acted as a test to see if language's effect on perceptual processing is online, this also acted as a control parameter to ensure that language's affect was in fact the key variable the design was tapping into.

 "We found that Russian speakers were faster to discriminate two colors if they fell into different linguistic categories in Russian (one siniy and the other goluboy) than if the two colors were from the same category (both siniy or both goluboy). This category advantage was eliminated by a verbal, but not a spatial, dual task." (ibid. p.7783)

6

These results afford further grounds for the trivial and habitual hypotheses as it clearly shows via the verbal interference task that linguistic codes were being utilized as primes for the Russian participants, i.e., the knowledge or habitual use of the Russian language made these participants attend to this colour boundary more readily. While this is compelling evidence, it must be noted that the difference in discrimination speed was only 124 milliseconds. McWhoter (2016) criticizes this finding, agreeing that it is indeed a confirmatory result for an effect of language on thought, but for it to be only one tenth of a second of a difference makes it very hard to conclude any qualitative difference about how Russians have a different experience of the world (ibid. pp.9-10). I am sympathetic this position, but when dealing with primes and reaction times the margin often seems negligible when scaled back up to the real time of experience.

I will draw on but one more interesting support from the domain of categorical perception, which is related to the Russian blues study. Athanasopoulos (2009) aimed to see whether learning a second language leads to an adoption of its categorical divisions and so causing a cognitive restructuring of the bilingual's mind. Greek distinguishes between light blue "ghalazio" and dark blue "ble", similarly to Russian. Using Greek bilinguals who have learned English, which does not subdivide the shades of light and dark blue, Athanasopoulos sought to test the hypothesis that acquisition of English would diminish the salience of the light/dark blue distinction, thus English proficiency and light/dark discriminability were posited to negatively correlated. This was found to hold true and the author interprets the results as further indication that concepts are not fixed nor stable and are "susceptible to both linguistic and cultural influence" (Ibid. P.93). For a trivial or habitual position to be shown to be tenable, the ability to acquire new categories and perceptual demarcations by learning a new language is crucial. The degree to which this becomes habituated is also of interest, Athansopoulos found that the Greek natives' length of stay in England was a pivotal variable in shifting the foci from the primary to secondary language's distinctions. I do not believe, however, that this lends support to a more radical determinist stance. A determinist interpreting these findings would be forced to conclude that the Greek speaker is fundamentally altering the way in which they experience the world. I would hesitate to jump to such a conclusion, for one it is highly

unlikely that the Greek speaker would note any qualitative difference in how they see (or how it feels for them to see) light blue or dark blue. They are not gradually beginning to see things as an English person would, rather they are becoming subtly aware that there are different ways in which to see things. Their attention has been implicitly drawn to other features of the environment, as the language acts as a spotlight priming the speaker to be predisposed to perceiving a certain way.

Converging results across several perceptual and cognitive domains express a trivial or habitual conclusion. The issue for future studies will now be to tease apart the differences between the trivial and habitual hypotheses and test their validity against one another, though this task will likely be unsatisfactory as both formulations of the moderate relativist views possess little to differentiate between them except for what it is to eventually make something a habit. As such, I agree with Reines and Prinz that "language draws our attention to features of the world, and noticing these features becomes habitual" (2009, p. 1030). I think that they may eventually be demonstrated to be identical, as you can't have a habitual effect without it first being trivial, it would be impossible to start with an instant habit. The habitual hypothesis (and this is where I see it most associated with the 'language as meddler' hypothesis) hinges on a speaker's habituation of the language; on it becoming properly engrained in their thinking, i.e., non-linguistic and linguistic codes being recruited together, they may be consistent with, or compete with, one another and that is where they express their effect (Wolff & Holmes, 2011, pp. 255-256). But for something to become a habit it would first have to be something noticed and enforced trivially, but for that trivial feature to be habituated means it is likely salient to begin with. As such these positions are somewhat recursive and circular. It will likely prove difficult to disentangle the trivial and habitual positions, because of this progressive evolution from trivial to habitual and the possibility of habitual back to trivial if both hold true. Even if the trivial hypothesis alone is correct, there is then the issue of illustrating when something becomes a habit and automatic or not. Nonetheless, the alternatives of linguistic determinism, i.e., that language constrains our thinking into such a fixed state of affairs that we cannot think outside of that mode, and its antithesis – that language does not affect thinking at all are much less tenable positions.

REFERENCES:

Athanasopoulos, P. (2009) Cognitive representation of colour in bilinguals: The case of Greek blues. *Bilingualism: Language and Cognition*, 12, 83–95.

Au, T. K. (1983). Chinese and English Counterfactuals: The Sapir-Whorf hypothesis Revisited. *Cognition, 15,* 155- 187.

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution.* Berkeley: University of California Press.

Bloom, A. H. (1981). *The linguistic shaping of thought: A study in the impact of language on thinking in China and the West.* Hillsdale, NJ: Erlbaum Associates.

Bloom, P., & Keil, F. C. (2001). Thinking through language. *Mind and Language*, 16(4), 351-367.

Boroditsky, L., Schmidt, L. A., & Phillips, W. (2003). Sex, Syntax, and Semantics in D. Gentner, & S, Goldin-Meadow (Eds.) *Language in Mind: Advances in the Study of Language and Thought,* (pp. 61-79). Cambridge, MA: MIT Press.

Davidoff, J. (2001). Language and Perceptual Categorisation. *TRENDS in Cognitive Sciences*, 5(9), 382-387.

Halpern, A., Whorf, B., Carroll, J., & Chase, S. (1956). Language, Thought, and Reality. *American Sociological Review,* 21(5), 653.

Imai M, Mazuka R. Re-evaluation of linguistic relativity: language-specific categories and the role of universal ontological knowledge in the construal of individuation. In: Gentner D, Goldin-Meadow S, eds. *Language in Mind: Advances in the Issues of Language and Thought*. Cambridge, MA: MIT Press; 2003, 430–464.

Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf Hypothesis? *American Anthropologist,* 86, 65-79.

McWhorter, J. (2016)*. The Language Hoax: Why the world looks the same in any language*. Oxford: Oxford University Press.

Pinker, S. (1994). *The Language Instinct.* New York: Morrow.

Regier, T., & Kay, P. (2009). Language, Thought, and Color: Whorf was half right. *Trends in Cognitive Sciences,* 13(10), 439-446.

Reines, M. and Prinz, J., 2009. Reviving Whorf: The Return of Linguistic Relativity. *Philosophy Compass*, 4(6), pp.1022-1032.

Roberson, D., Davidoff, J., Davies, I., & Shapiro, L. R. (2005). Color Categories: Evidence for the Cultural Relativity Hypothesis. *Cognitive Psychology,* 50, 378-411.

Roberson, D., Davies, I., & Davidoff, J. (2000). Colour Categories Are Not Universal: Replications and New Evidence from a Stone-Age Culture. *Journal of Experimental Psychology: General*, 129, 369–398.

Winawer, J. et al. (2007) Russian blues reveal effects of language on color discrimination. Proceedings of the National Academy of Sciences 104 (19), 7780–7785.

Wolff, P., & Holmes, K. J. (2011). Linguistic Relativity. *WIREs Cognitive Science,* 2, 253-265.

Yeh, D., & Gentner, D. (2005). Reasoning Counterfactually in Chinese: Picking up the Pieces. *Proceedings of the Twenty-Seventh Annual Meeting of the Cognitive Science Society,* 2410-2415.