The Baltic International Yearbook of
Cognition, Logic and Communication

December 2014 Volume 9: *Perception and Concepts*
pages 1-13 DOI: 10.4148/biyclc.v9i0.1086

ELISABETTA LALUMERA
University of Milano–Bicocca

WHORFIAN EFFECTS IN COLOR PERCEPTION: DEEP OR SHALLOW?

ABSTRACT: This paper discusses, from the point of view of the philosophy of psychology, recent behavioral and brain studies showing effects of the diversity of language vocabulary on color perception. I examine the alternative between two different interpretations of these so-called whorfian effects, namely habitual or deep whorfianism, and shallow whorfianism. I argue that at the moment the evidence underdetermines both interpretations and the question is open. I also clarify that shallow whorfianism is not a synonym for 'trivial whorfianism', as some authors have suggested, but rather makes a case for the online and situated nature of human cognition.

1. INTRODUCTION

According to the hypothesis of linguistic relativity, or Sapir-Whorf hypothesis, linguistic differences determine cognitive differences. As the linguist B.L. Whorf wrote, "we dissect nature along lines laid down by our native language" (Whorf 1956/1970, p. 212). This thesis was neglected by the scientific community in the early decades of cognitive science, partly because of its methodologically flawed experimental evidence, and partly because of the success of universalist and nativist research programs that established some amount of prelinguistic cognitive basis for most key human cognitive faculties.

Recently, however, the pendulum seems to be swinging back, as whorfian effects, or effects of language differences on cognition, are being found in many domains (including colors, objects, emotions, time, space, and numbers). But whereas there is large consensus on whorfian effects, there is little on how to interpret them. A few researchers invoke a change of paradigm from Universalism to Whorfianism, while others defend Universalism by pointing out methodological flaws in new studies that seem to challenge it, but the main tendency seems to be to propose taxonomies of more fine-grained hypotheses about the possible roles of languages in cognition as a replacement for the traditional Universalism vs Linguistic relativity dichotomy (Wolff & Holmes 2011; Reines & Prinz 2009; Slobin 1996).

In line with this latter tendency, this paper focuses on color cognition from the point of view of the philosophy of psychology. In the domain of colors, Universalist versus Whorfian opposition now tends to be recognized in the literature as explanatorily inadequate and the alternative on the table is rather whether language affects color perception and cognition by establishing long-term, stable habits of seeing the world (habitual, or deep, whorfianism), or by providing short-term online cues during the perceptual process (Language-as-a-Meddler effect, or shallow whorfianism) (Wolff & Holmes 2011; Winawer et al. 2007). Typically — though not necessarily — if effects of the diversity of languages on thought are classified as shallow, their explanatory relevance for our broad understanding of cognition is somehow presented as minor (Pinker 1994). On the other hand, if they are described as deep, their importance is promoted — color language affects color cognition because it affects it permanently.

In this paper I illustrate the negative conclusion that even the alternative between these two neowhorfian hypotheses is underdetermined by the data. In other words, from a methodological point of view, cross-linguistic studies, behavioral studies, and FMRI studies on categorical perception at the moment fail to adjudicate between deep and shallow whorfianism in the domain of colors. This is not to say that whorfianism *per se* is untestable, as some authors have claimed (Cruz 2009), but rather that more work has yet to be done.

Moreover, I suggest that so-called shallow whorfianism in the domain of colors is not just a weak form of influence of the diversity of

languages on color cognition, but rather evidence for the view that at least in some cases color cognition is contextual, situated, and employs on-line strategies. Shallow whorfianism would then be evidence for a broad picture of embodied and situated human cognition (Barsalou 2012; Casasanto & Lupyan 2011).

This is how the paper is organized. Section 1 contains a survey of cross-linguistic studies on color representations, and explains why they are not decisive. Section 2 is about behavioral and brain studies, especially about Key and Regier's proposal that color perception is permeable to linguistic influence only when stimuli appear in the right visual field. I also discuss the view that shallow whorfianism is trivial whorfianism. Conclusions are in the third section.

2. CROSS-LINGUISTIC STUDIES ON COLOR

Color cognition has always been a battlefield for Relativism versus Universalism. The spectrum of visible wavelength could be potentially carved up in indefinitely many ways, and in fact the number of color terms available varies from language to language, from a minimum of 2 to a maximum of eleven. Notice that here, 'color term' is intended in the restrictive sense of a monolexeme referring to a specific color, so that "green" is a color term, but "dark green" and "olive" are not. In the 70s and 80s Berlin and Kay (1969) and Heider (1972) found support for universal constraints on color perception, and proposed the view that there is a class of so called 'focal colors' that are perceptually more salient, easy to re-identify and discriminate even in the absence of corresponding color terms in one's language, and easy to remember: black, white, red, green, yellow and blue. According to their hypothesis, languages evolve from a two-focal colors lexicon to an eleven focal color lexicon. More precisely, in a language with only two color names, the two terms stand for 'dark' and 'bright', and additional color terms are added in a fixed order as the lexicon evolves: first green or yellow, then blue, etc.

This is a universalist hypothesis on color cognition, because the characteristics of languages depend on focal colors, and not the other way round. Thus evidence in favor of focal colors gathered by means of the World Color Survey has reinforced the view that there are universal

constraints on color cognition for many decades.

Recent cross-linguistic studies, however, failed to replicate some of Heider's earlier empirical results, casting doubt on the hypothesis that focal color terms correspond to perceptual universals, and on the impenetrability of color perception to language (Davidoff et al. 1999; Roberson et al. 2000, 2005). In one study, Berimno people of New Guinea, whose language has five color terms, appeared to make more mistakes than the English-speaking control group in reidentification tasks, where the stimuli were focal colors not lexicalized in their language — like yellow. Their performance was as good as the control group where the color stimuli were lexicalized in their own language. Another study of the same group of researchers was designed to tap categorical perception, namely, perception at the boundaries between categories. Berimno speakers judged within-category stimuli to be more similar to each other than cross-category stimuli where their own linguistic boundary coincided with the boundary of the set, but not otherwise. For example, they judged a shade of yellow as more similar to white than to another shade of yellow. This squares with the whorfian prediction, because it implies that categorical perception is language-dependent, and whorfian. The same research team replicated the experiments with another population living in a different natural environment, the Himba of Southern Africa, so as to eliminate the possibility that the environment, and not language, was the key influence on color perception (Roberson et al. 2005). Their conclusion was, again, that color perceptual categories are language-dependent, rather than universal.

However, it is important to notice that to establish that there are language-dependent differences in color categorization is not yet to exclude the possibility that there are universal tendencies, too. Universalism about constraints on color cognition can be compatible with recognition of whorfian effects. This is the key idea of Lindsey & Brown (2006), who analyzed the data contained in the World Color Survey, and by performing cluster analyses on the color naming systems of individuals, rather than populations, obtained cluster categories that closely resemble the traditional focal colors of Heider's seminal findings. In an similar vein, Kay & Regier (2007) replied to the Berimno data by pointing out that they still do match the categorization tenden-

cies that can be extracted from the World Color Survey — specifically, Berimno categorize colors just like the other five-color-terms languages of the Survey do, and thereby they exemplify a linguistic universal.

Acknowledging these results, three points are worth noting about cross-linguistic studies on color perception. First, there is a very general point — that the ecological quality of cross-linguistic studies makes them extremely difficult to replicate and control in all the validity parameters. Second, the more conservative interpretation of the results, taken collectively, is orthogonal to the traditional Universalist versus Whorfian divide: cross-linguistic studies show that there are both universal tendencies *and* language-dependent variation in color categorization. In fact, this has been readily acknowledged by some of the very participants in the experimental controversy (Kay & Regier 2007).

Third, generally (with some exceptions) cross-linguistic studies address the question whether language diversity influences color cognition or not, but fail to dig deeper into how, if yes, such an influence works. In other words, even when the traditional Universalist versus Whorfian opposition is left behind, the question of the interpretation of whorfian effects is still open. As noted above, commentators have already distinguished between two views, habitual or deep whorfianism on the one hand, and online or shallow linguistic influence on the other (Wolff & Holmes 2011). I think a proper understanding of the distinction is crucial for understanding the case of colors, so let me elaborate a bit on it.

The source of the distinction between shallow and deep whorfianism is Dan Slobin's concept of *thinking for speaking* (Slobin 1996). Simply put, the idea is that when speakers of different languages speak, they are likely to employ different concepts, and not just different words for them, and the concepts employed depend on what one's language forces or facilitates one to express. For example, in the English sentence "The bird flew down from out of the hole in the tree" the concepts of manner and direction are employed, according to Slobin, because English motion verbs generally encode manner and direction (whereas, for example, Spanish or Italian verbs are different). Slobin's idea was that thinking for speaking may be whorfian, whereas some areas of thought that involve sensorimotor and perceptual processes may be not. In contemporary terms, thinking for speaking is usually called 'shallow whorfianism', and 'deep whorfianism' would be the view that languages modifies concepts even when they are not recruited in a linguistic task. A version of deep whorfianism is Francisca Reines and Jesse Prinz (2009), that is, the idea that language differences may induce different thinking habits, or modifications of the facility and frequency with which concepts are employed, even in non-linguistic tasks.

Going back to the case of colors, the opposition is between deep whorfianism, the view that language diversity modifies or favors certain perceptual color representations and/or concepts even when they are not employed in a color recognition task — i.e., they are offline – and shallow whorfianism, according to which linguistic representations are active when recruited for color perception tasks, and they exert their influence on categorization judgments only then. In other words, shallow effects are task-dependent and temporary, while deep effects are not. Deep effects suggest that experience with language use sets up default categorization procedures. As Lupyan (2012) explains, linguistic labeling gradually makes certain representations more similar, and results in a separation of the parts of the color spectrum that color words denote. Shallow effects, on the other hand, are compatible with two radically alternative views of human cognition. On the one hand, they can be accommodated within a classical cognitivist view such as the traditional Chomskyan view (Pinker 1994), where language comes after the outputs of perception — thus favoring the traditional Universalists' stance. On the other hand, shallow whorfianism can also be part of a more revisionary view, according to which representations are shaped up or formed online for the needs of a specific task.

Though cross-linguistic studies are crucial in providing data in favor of whorfian effects on color perception, they are not sufficient to adjudicate between deep and shallow whorfianism. This is due to the very nature of the experimental paradigms that can be employed on the field. Typically, they involve memory or recognition tasks of color chips. Even when no verbal report is recorded, it is generally left open whether the linguistic influence comes before or after perception and categorization. So how exactly does language variability affect color perception, even in the attested Berimno cases? How deep are whorfian effects?

Notice here that "deep" and "shallow" are not meant as synonyms

for "interesting" and "uninteresting" effects of language on cognition, though strong Universalists such as Steven Pinker tend to endorse such synonymy (Pinker 1994, p. 64). According to Pinker, the influence of language variety tested in most studies is uninteresting, because trivial, because the linguistic strategy is consciously employed by subjects in order to accomplish the task. For example, he claims that it is obvious that when subjects have no hint of which color chips go together, they think to themselves that they just might group together those chips that have the same name in their own language. In fact, experimental paradigms involving subjective similarity tasks are subject to this kind of objection. The very nature of the task enables subjects to search consciously for the more rational strategy available. If linguistic labeling is a conscious strategy recruited online in order to perform a specific task, then language comes on top of perception, and does not affect it deeply (Winawer et al. 2007).

In the next section we shall consider experimental paradigms where conscious linguistic strategies are appropriately screened off, and still some whorfian effects show themselves to be task dependent and temporary. A question on this point is worth raising here. Is that enough to deem such effects as uninteresting, *just because* they are task dependent and temporary? The answer is that it would be enough, but at the price of committing to the view that *only* stable and context-free representations are employed in perception and cognition. However, there is growing agreement to the contrary, as different lines of research converge on the view that conceptual and perceptual representations are intrinsically task-dependent and temporary, and that they are built ad hoc from stored material as a result of the way retrieval cues interact with the physical and linguistic context. This kind of flexibility is a characteristic of human cognition, which enhances our performance in reasoning, understanding and categorization tasks (Barsalou 2012; Casasanto & Lupyan 2011; Mazzone & Lalumera 2010). Evidence for such a broad view, however, cannot be given within the space of this paper. The aim here is just to show that experimental studies so far underdetermine both deep whorfianism and shallow whorfianism, and the above specification is meant simply to clarify that shallow whorfianism, if proved right in some cases, would not be a trivial result.

3. BEHAVIORAL AND BRAIN STUDIES ON COLOR

This section contains a discussion of some recent studies on the shallow versus deep whorfianism opposition, which appropriately replaces the traditional relativist versus universalist divide. From a methodological point of view, in order to understand the nature of whorfian effects — once it is established that there are some — we need to see what happens before linguistic thought is consciously activated, that is, we need to avoid subjective similarity tasks, and tasks involving memory, which are more vulnerable to conscious recruitment of linguistic strategies. Simply put, what is at stake is what subjects see as similar, not what they tell us. Following these desiderata, Winawer and collaborators tested Russian and English speakers on a color discrimination task. Russian marks the distinction between lighter blues ("*goluboy*") and darker blues ("*siniy*"). Each subject was presented with three colour squares, one on the top and two at the bottom, and was instructed to pick up which one of the two bottom squares matched the colour of the top one. Results show that Russian subjects are faster in their response when the two squares at the bottom belong to different linguistic categories in Russian, i.e. one is goluboy and the other one is siniy, while English speakers do not show any difference in performance with respect to this variation. This confirms a categorical perception effect (better discrimination of cross-category stimuli), and a whorfian effect (categories coincide with linguistic labels). Moreover, the study provides evidence for Shallow Whorfianism, as it also includes tests with a verbal task interference. If discrimination performance is disturbed by a verbal task, then it is likely that language is involved online in discrimination. This was confirmed by data on Russian speakers, whose reaction time increased when they were simultaneously asked to rehearse silently a string of digits. As the authors remark, "if the language-specific distortions in perceptual discrimination had been a product of a permanent change in perceptual processors, temporarily disabling access to linguistic representations with verbal interference should not have changed the pattern in perceptual performance" (2007, p. 3774).

An uncertain verdict about the shallow versus deep alternative, however, comes from brain studies, which prima facie provide the opportunity of obtaining objective measures of similarity and discrimination tasks, but as we shall see, their interpretation remains open. Athana-

sopoulos et al. (2009) tested Greek and English speakers in a color oddball detection task, and measured their brain wave patterns. Subjects were asked to press a button when an odd shape appeared on the screen, for example a blue square after a series of circles of different shades of blue, or a pink circle after a series of pink squares of different luminescence. The results confirmed the relativist prediction that Greek speakers would perceive the difference between light blue and dark blue as more salient than English speakers, as the Greek language marks the two shades with distinct color terms. The interesting point of this study is that color recognition was not part of the core task, so color perception was below the level of conscious attention, yet was nevertheless showed to be permeable to language variability. Athanasoupoulous and colleagues concluded that Whorfian effects appear at a very early stage of color perceptual processing, suggesting deep habitual whorfianism.

However, there is also a different interpretation of the same results, as the researchers themselves admit. It may be that language is accessed online very early in the process of color perception, but anyhow affects an initial stage where color vision is common to speakers of different languages. Rather than a deep shaping effect on color perception, language on this interpretation would have a facilitating effect — it provides a useful but dispensable tool (shallow, Language-as-a-Meddler effect).

This latter kind of shallow interpretation of whorfian effects on color perception is again *prima facie* supported by behavioral studies by Gilbert et al., again on categorical perception (2006; 2008), and Roberson et al. (2008). These studies give provide evidence for the hypothesis that the influence of language on color perception is remarkable only when stimuli are presented in the right visual field, due to projection of the right-visual field representations to the left hemisphere, where most lexical information is stored. Gilbert et al.'s experiments are similar to the ones by Winawer described above, in that they tested categorical perception effects. Results show that cross-category color targets are identified more quickly than same-category targets only when they are presented in the right visual field. When subjects have to perform a simultaneous task requiring verbal resources (remembering an eight-digit number), the categorical perception effect is no longer present.

As the categories in the categorical perception effects are those labeled by the languages tested, whorfian effects are also inferred. The whorfian effects appear to be shallow for two reasons, first, because they are relative to the position of the stimulus, and therefore definitively eliminable, and second, because they can be disrupted by verbal interface. But is lateralization per se evidence for shallow whorfianism? The key assumption of the studies described above is that categorical perception in the left visual field (for stimuli presented in the right visual field) brings with it whorfian effects, and categorical perception is caused by the online influence of language. However, this is far from unquestionable.

Let us abandon colors for a while. A recent line of research aims at proving that left-lateralized categorical perception in the left visual field is not linguistic. Holmes & Wolff (2012) reasoned as follows: if categories with no linguistic labels would give rise to left-lateralized categorical perception, then language has no online role. Results confirmed their hypothesis. In the main experiment, participants were presented with a discrimination task of novel object silhouettes. They were asked to indicate whether a target object presented within a ring of identical distractors was located on the left or on the right hand side of the display. Targets placed in the right visual field were discriminated more quickly if they belonged to different categories than if they belonged to the same, both in the no-label and in the label condition.

What is at stake here? According to Holmes and Wolff, these findings "provide the first unambiguous demonstration in adults that the left hemisphere is associated with categorical processing independently of language" (2012, p. 442). This conclusion is indirectly very relevant for the issue of shallow versus deep whorfianism. If whorfian effects are prominent when the stimuli are processed in the left visual field, and if in the left visual field categorical perception is not performed by recruiting language online, then whorfian effects do not arise from the online recruitment of language. That is to say, whorfian effects are deep, rather than shallow. Taken at face value, these results therefore counterbalance those obtained by Kay and Regier, and sever the connection between lateralization and shallowness of whorfian effects.

4. CONCLUSION

Contemporary research on the relativity of color cognition has recently abandoned the traditional Universalism versus Whorfianism divide, and faces the issue of deciding between a shallow whorfian and a deep whorfian interpretation of results. It has been claimed (Cruz 2009) that traditional whorfianism is untestable, largely due to the methodological difficulties of cross-linguistic studies. Here, I partially acknowledged the point – also shallow versus deep whorfianism cannot be assessed by usual paradigms employed in cross-linguistic research on the field. However, main trends in both behavioral and brain studies are still far from accrediting one interpretation over the other. This is not to say that the two views are untestable for methodological or a priori reasons, but rather that alternative models of the language-cognition interface are still compatible with data on color perception. Available evidence suggests that shallow or contextual effects of color language on color cognition are pervasive, in line with a broad view of human cognition according to which online temporary resources are employed across the board — thus giving us shallow whorfianism as a mark of flexibility (Casasanto & Lupyan 2011). However, the hypothesis that some of these linguistic strategies become permanent is not yet to be discarded.

References

Athanasopoulos, P., Wiggett, A., Dering, B., J-R., Kuipers & Thierry, G. 2009. 'The Whorfian mind: Electrophysiological evidence that language shapes perception'. *Communicative & Integrative Biology* 2, no. 4: 332–334.

Barsalou, L. W. 2012. 'The human conceptual system'. In 'The Cambridge Handbook of Psycholinguistics', 239–258. Cambridge University Press.

Berlin, B. & Kay, P. 1969. *Basic color terms–Their universality and evolution*. University of Caifornia Press, Berkeley/Los Angeles.

Black, S. R. 2003. 'Review of semantic satiation'. *Advances in psychology research* 26: 63–74.

Casasanto, D. & Lupyan, G. 2011. 'Ad hoc cognition'. In 'Proceedings of the 33rd Annual Conference of the Cognitive Science Society: 826', .

Cook, R. S., Kay, P. & Regier, T. 2005. 'The World Color Survey database: History and use'. In Cohen H. & Lefebvre C. (eds.) 'Handbook of Categorisation in the Cognitive Sciences', 223–242. Amsterdam and London: Elsevier.

Cruz, H. De. 2009. 'Is linguistic determinism an empirically testable hypothesis?' *Logique et Analyse* 52, no. 208: 327–341.

Davidoff, J., Davies, I. & Roberson, D. 1999. 'Colour categories in a stone-age tribe'. *Nature* 398, no. 6724: 203–204.

Drivonikou, G. V., Kay, P., Regier, T., Ivry, R. B., Gilbert, A. L., Franklin, A. & Davies, I. R. L. 2007. 'Further evidence that Whorfian effects are stronger in the right visual field than the left'. *Proceedings of the National Academy of Sciences* 104, no. 3: 1097–1102.

Gendron, M., Lindquist, K. A., Barsalou, L. & Barrett, L. F. 2012. 'Emotion words shape emotion percepts'. *Emotion* 12: 314–325.

Gilbert, A. L., Regier, T., Kay, P. & Ivry, R. B. 2006. 'Whorf hypothesis is supported in the right visual field but not the left'. *Proceedings of the National Academy of Sciences of the United States of America* 103: 489–494.

——. 2008. 'Support for lateralization of the Whorf effect beyond the realm of color discrimination'. *Brain and Language* 105: 91–98.

Gilbert, C. D. & Sigman, M. 2007. 'Brain states: Top-down influences in sensory processing'. *Neuron* 54: 677–696.

Heider, E. R. 1972. 'Universals in color naming and memory'. *Journal of experimental psychology* 93: 10–20.

Heider, E. R. & Olivier, D. C. 1972. 'The structure of the color space in naming and memory for two languages'. *Cognitive Psychology* 3: 337–354.

Holmes, K. J. & Wolff, P. 2012. 'Does categorical perception in the left hemisphere depend on language?' *Journal of Experimental Psychology: General* 141: 439.

Kay, P. & Regier, T. 2007. 'Color naming universals: The case of Berinmo'. *Cognition* 102: 289–298.

Lindsey, D. T. & Brown, A. M. 2006. 'Universality of color names'. *Proceedings of the National Academy of Sciences* 103, no. 44: 16608–16613.

Lupyan, G. 2012. 'Linguistically modulated perception and cognition: The label-feedback hypothesis'. *Frontiers in psychology* 3.

Mazzone, M. & Lalumera, E. 2010. 'Concepts: Stored or created?' *Minds and machines* 20, no. 1: 47–68.

Pinker, S. 1994. *The Language Instinct*. New York: William Morrow and Company.

Regier, T. & Kay, P. 2009. 'Language, thought, and color: Whorf was half right'. *Trends in cognitive sciences* 13, no. 10: 439–446.

Reines, M. F. & Prinz, J. 2009. 'Reviving Whorf: The return of linguistic relativity'. *Philosophy Compass* 4, no. 6: 1022–1032.

Roberson, D. & Davidoff, J. 2000. 'The categorical perception of colors and facial expressions: The effect of verbal interference'. *Memory & Cognition* 28: 977–986.

Roberson, D., Davies, I. & Davidoff, J. 2000. 'Color categories are not universal: Replications and new evidence from a stone-age culture'. *Journal of Experimental Psychology: General* 129: 369–398.

Roberson, D., Davidoff, J., Davies, I. R. L. & Shapiro, L. R. 2005. 'Color categories: Evidence for the cultural relativity hypothesis'. *Cognitive Psychology* 50: 378–411.

Roberson, D., Pak, H. & Hanley, J. R. 2008. 'Categorical perception of colour in the left and right visual field is verbally mediated: Evidence from Korean'. *Cognition* 107: 752–762.

Slobin, D. L. 1996. 'From "thought and language" to "thinking for speaking"'. In JJ Gumperz & SC Levinson (eds.) 'Rethinking linguistic relativity', 70–96. Cambridge: Cambridge University Press.

Tuborg, J. 2008. 'The evolutionary role of shame'. *BioScience* 24: 715–722.

Whorf, B. L. 1956/1970. 'Language, thought, and reality: Selected writings of Benjamin Lee Whorf'.
Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R. & Boroditsky, L. 2007. 'Russian blues reveal effects of language on color discrimination'. *Proceedings of the National Academy of Sciences* 104, no. 19: 7780–7785.
Wolff, P. & Holmes, K. J. 2011. 'Linguistic relativity'. *Wiley Interdisciplinary Reviews: Cognitive Science* 2: 253–265.