Carving up the rainbow: how to model linguistic categorization of color

MSc Thesis *(Afstudeerscriptie)*

written by

Radek Ocelák
(born February 3rd, 1988 in Bruntál, Czech Republic)

under the supervision of **Prof. Dr. Martin Stokhof**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:	**Members of the Thesis Committee:**
July 3rd, 2013	Prof. Dr. Martin Stokhof
	Dr. Maria Aloni
	Dr. Lieven Decock
	Dr. Michael Franke

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Abstract

The thesis deals with categorization of color in language, specifically with the question how to account for its observed cross-linguistic patterns. To this end, I consider color categorization models of the most recent type, which go beyond the dated dichotomy of universalism and relativism, integrating constraints on color perception by a human individual as well as constraints on language interaction between individual agents.

After setting the color categorization problem in its historical, disciplinary, and ideological context, I proceed from the question of evaluation. I argue, contra the relativist critics, that the World Color Survey (the most extensive color naming research up to date) is suited to provide reliable data on color categorization in languages of the world, against which color categorization models should be evaluated. However, a considered reduction of the actual WCS data is desirable in order to exclude the impact of several distorting factors.

The major part of the thesis, then, is focused on the question of an appropriate perceptual basis for a color categorization model; that is, on the assumptions about individual color perception that such a model should embody. First, I examine the relevance of various color topologies for the color categorization problem. I suggest a basic modeling strategy, which involves the CIELAB color space with an updated color difference formula. The proportion of color-deficient agents in the population is likely to play a significant role. After that, I consider the possible explanatory role of two phenomena that are rather central and, as I argue, seriously misconceived in the contemporary color science: what is known as “unique hues”, and what is commonly referred to as “categorical perception of color”. Against the mainstream opinion, I argue that there is no reason to claim that some hues (in particular, red, yellow, green, and blue) are privileged or constitutive in human color perception. The widespread notion of four perceptually unique hues organized in a double-opponent fashion is flawed and has no explanatory relevance with respect to linguistic categorization of color. The phenomenon of categorical perception, as far as prelinguistic children are concerned, can be employed in explaining linguistic categorization of color, but only upon a substantial clarification. Categorical perception effects must not be explicated, as is common, in terms of warping of the perceptual color space. Moreover, the existing evidence on “infant categorical perception” does not license the usual conclusion that infants perceptually categorize color. For our explanatory purpose, the phenomenon should be conceived as differentiation of the discrimination performance over the perceptual color space.

I conclude by a brief outline of other desirable components for a color categorization model, ones that go beyond the level of individual color perception.

Contents

Acknowledgments

I am deeply grateful to my supervisor Martin Stokhof. It is not only from his supervision, but also from a couple of courses I previously took with Martin, that I have recognized in him the ideal of a teacher who, after many years in the academic traffic, still takes his work with all pedagogical seriousness. During the last two years have I, one among his many students, handed to him a heap of written pages. He has read every single of them thoroughly and offered a number of insightful comments in a short time. As for the thesis itself, his guidance was enthusiastic and invaluable for defining the subject and for improving my first elaborations of its particular aspects into the present shape. Martin took me on with a topic that would, ideally, call for a field linguist, a neurophysiologist, a psychologist and a technologist of color, and a philosopher. A thesis later, I still think the last one is not the least indispensable.

I would like to thank my thesis committee members for their willingness to read and assess my work in the beginning summer.

I am obliged to Pedro Correia, who brought me to the field of color categorization research and provoked my desire for a critical interdisciplinary treatment.

My thanks also go to Tanja Kassenaar and to Ulle Endriss, who work hard to make the MSc program run smoothly for everyone else, and solve problems even before they arise. Often, it is such a difference from elsewhere! I am also grateful to my mentor Maria Aloni for her kind advice in pursuit of my studies.

I owe thanks to the folks of ILLC in general, who on many occasions astonished me both with the formal rigor of their work and with their disinterest in the phenomena they hope to account for with their formal tools.

I gratefully acknowledge the support of my studies by Nuffic. I am one of the last to have been granted this great opportunity, for the Huygens Scholarship Programme has been closed as of 2012 due to education spending cuts. The present thesis may be considered a testimony as to whether it was a right decision or not.

Last but certainly not least, my gratitude is to my family back home and to my fine friends both Czech and international, without whom the last two years would have been a lot tougher. No lists, but you guys know. A Tobě, chlapečku.

Chapter 1

Introduction

For ages, people have been aware of the fact that different languages do not necessarily treat color in the same way. By way of illustration, the English color naming system does not fully coincide with that of Russian and modern Greek, where what we call blue in English is split between two salient color terms, one for lighter and the other for darker shades of blue. Reversely, in the Welsh color naming system, there is a single term covering both the English categories green and blue. Much stronger discrepancies appear when we observe color naming in languages that are geographically, genetically and culturally distant from the European languages.

Cross-linguistic patterns of color categorization constitute, at least since Berlin and Kay's *Basic color terms* (1969), a major topic in empirically, typologically oriented semantics. The issue is twofold. First, are there universal patterns, or at least non-randomly strong tendencies, in how languages divide the space of observable colors with their basic color vocabulary? If so, what are these patterns or tendencies like? Second, if there is any significant degree of universality in categorization of colors, how can we explain it? For about three decades, the debate was dominated by the opposition between two camps, the universalists and the relativists, who proposed different answers to both these questions. The domain of color became a major battlefield of the more general "nature vs. nurture" dispute about the origins of human cognitive categories, which seems to have contributed to the polarization of the color naming discussion.

Recently, the dichotomy of universalism and relativism with respect to color categorization somewhat dissolved. The result is an approach which assumes limited universal tendencies in color categorization, as confirmed by most cross-linguistic studies, and tries to account for these tendencies, as well as for the remaining variation, by means of concrete color categorization models. These models typically embody some of the explanatory principles previously proposed by either of the opposing camps. That in particular includes constraints given by the character of color perception by a human individual, on one hand, and the evolutionary influence of language interaction between perceiving individuals (simulated in game-theoretic terms) on the other. Several models of this type have been built up to now. However, for two basic reasons these models are questionable. One is that the authors usually offer little independent motivation for the concrete explanatory principles employed. Moreover, the models do not come with a transparent evaluation against empirical data that could be, with respect to the current state of knowledge, taken to reflect the cross-linguistic reality of color categorization.

The present thesis aims to provide a firmer ground for color categorization models of the recent type, in two particular respects. First, it attempts to specify a reasonable evaluation procedure for these models, by means of addressing the highly non-trivial question, what are the most reliable cross-linguistic data at hand as far as categorization of colors by languages of the world is concerned. Second, it defines a perceptual basis for these models; that is, the assumptions about the character of color perception by human individuals upon which the game-theoretically modeled language interaction should arguably take place. In this thesis I had to leave aside the third crucial aspect of this type of models which also calls for a substantial reflection: the character of the very agent-based language interaction that is ultimately supposed to lead to constitution of color categorical systems. Where necessary, I simply assume a solution roughly along the lines of the existing game-theoretic models.

The thesis is organized as follows. Chapter 2 sets the question of universality of color categorization in its historical, disciplinary, and ideological context. After this introductory exposition, a more detailed motivation for the subsequent parts of the thesis is provided, and some key concepts are preliminarily introduced. Chapter 3 deals with the first of the two main issues, defining a reasonable empirical basis for evaluating color categorization models. To this end, I extensively discuss the World Color Survey, the largest cross-linguistic research ever performed on color naming, and the reliability of the thereby provided data with respect to categorization of color in languages of the world. Chapters 4, 5 and 6 are focused on various aspects of the other main issue, the question of an appropriate perceptual basis for color categorization models. More specifically, chapter 4 examines the relevance of various color topologies for the color categorization problem and suggests a basic modeling strategy. In chapters 5 and 6, I investigate two particular issues of the contemporary color science which, in my opinion, are in an urgent need of conceptual clarification: the problem of what is known as "unique hues", and the question of "categorical perception" of color. The treatise in these chapters attempts to rectify some crucial misconceptions that are mainstream in the respective research subfields. That, hopefully, in turn leads to a more adequate assessment of the possible explanatory role of these phenomena with respect to linguistic categorization of color. Chapter 7 concludes and outlines further desirable components for a color categorization model, namely ones that go beyond the level of individual color perception.

Chapters 3, 5, 6 (except for the introductory sections) and section 4.7 of chapter 4 are to be regarded as the main contribution of this thesis; my original arguments are concentrated there. Chapter 2 and most of chapter 4 have basically the character of an overview of a particular research field and are meant to provide the necessary background.

Chapter 2

Setting the problem

This chapter will proceed from a very brief historical outline of the color term debate of the last century and a half. The literature in this field has long since grown incredibly vast, virtually fractal-like for a single reader, and much of it is in one or other way relevant for what I will have to say. The following historical sketch thus can by no means provide an adequate picture of the research area in its past extent. It is only meant as a scaffolding for further introductory exposition that is necessary for the purposes of this thesis; namely, for a discussion of the major positions in the debate and of their various motivations. Neither these are dwelt on for their own sake. Rather, the discussion should provide an insight into the field that my own proposals are intended to enter, a picture of the issues considered important and of the strands of argumentation as well as rhetoric going on there.

The color term research of the last several decades spans, in the extreme, issues from fields as diverse as physics of light and philosophy of meaning. Given such interdisciplinary character of the debates, it is unusually worthwhile to pay attention not only to the "core" procedures of scientific discourse (such as reporting and evaluating evidence and drawing conclusions), but also to "sociology" and "anthropology" of the communication and its participants. By that I refer, for instance, to different backgrounds and ultimate goals of various research traditions, or to the inclination of scientists to rely on salient results from the adjacent disciplines disregarding the caveats and qualifications with which these results were originally formulated. Neglecting such aspects of the scientific traffic, I believe, one is likely to misjudge the significance of many reported findings. In the present thesis I therefore feel encouraged to track, to a necessary extent, also these "extrascientific" features of the discourse in question.

2.1 History of the color term debate

Berlin and Kay's *Basic color terms* (1969) is by far the most influential contribution in the history of the debate. For a long time it largely overshadowed the previous work, and it founded the subsequent massive research tradition in color categorization, virtually all of which bears some, however indirect, relation to the monograph's original claims. It is convenient to precede the chronological outline of the debate with a short characteristic of this milestone.

Berlin and Kay (1969), based on data gathered from speakers of 20 and literature on 78 other languages, proposed two generalizations. First, languages do not divide the space of colors in arbitrary ways, but center their color categories, or referents of their basic color terms, in at most 11 specific "focal" color points. Second, there is a universal sequence, or implicational hierarchy, according to which languages lexicalize these focal points in the course of their evolution: [white, black] $<$ [red] $<$ [green, yellow] $<$ [blue] $<$ [brown] $<$ [purple, pink, orange, gray]. The evolution of a language is found to correlate with the level of technological development of the respective culture: those with little technical ability to modify color of objects have, in general, poorer color vocabularies.

It turned out, though, that evolutionary thinking about color naming systems had emerged already a hundred years earlier, beginning with Gladstone on Homeric Greek in 1858 and proceeding with Geiger's first cross-linguistic generalizations in early 1870s. The two assumed a dependence of color naming on the evolution of color perception. Allen and Magnus, however, soon rethought the issue in terms of language evolution taking place upon invariant color perception, as biological evolution of color vision at that rate was declared unlikely by the contemporary Darwinians. Magnus' distribution of color samples to missionaries and Rivers' expedition in early 1900s pioneered modern practices of collecting typological data. (Lyons, 1995; Saunders, 2000; Dedrick, 1998.)

The evolutionary line of thinking fell into a temporary decline with the rise of the Boasian, strongly relativistic and structurally descriptive paradigm in anthropology and linguistics, with its rejection of Western ethnocentrism in cross-cultural investigations. In the post-war linguistics, formed by Boas' disciple Sapir and his disciple Whorf, it is a common notion that languages divide the color spectrum on an arbitrary basis, "except for pragmatic considerations" (Dedrick, 1998, p. 4). Berlin and Kay's immediate predecessors in color naming research were psychologists who attempted to operationalize and examine a version of the famous Sapir-Whorf hypothesis, which states an influence of language on thought (Brown and Lenneberg, 1954; Lenneberg and Roberts, 1956). To them, color appeared as an ideal area of investigation, offering a universal referential domain for arbitrary systems of naming. Their work was focused on proving a correlation between "codability" of spectral regions in particular languages on one hand, and some cognitive variables, such as memorability, concerning those particular spectral regions and the speakers in question on the other. (Lucy, 1997; Dedrick, 1998; Saunders, 2000.)

It is from Lenneberg and Roberts (1956) that Berlin and Kay (1969) adopt a good deal of methodology, despite the complete reversal of the theoretical agenda. (Saunders and van Brakel, 1997a.) This is the origin of using the array of 320 Munsell color chips in much of the subsequent research, as a representation of the assumed universal domain of color. Hereafter, the meaning of a color term is usually conceived or schematized as a region of this array; and during lab or field work, this very array as a whole or the particular color chips are presented to the informants in order to establish semantic boundaries and foci of various color terms.

Berlin and Kay's universalist turn received a strong support from the psychologist Eleanor Rosch (1972a; 1972b; 1973), who presented cross-cultural evidence (comparing American English and the Dani language of Western New Guinea) for perceptual, non-linguistic salience of certain colors, allegedly in correspondence with Berlin and Kay's (1969) findings on focal color points. Opposing Lenneberg and others, she claimed that these focal, or prototypical, colors, which are faster to learn and easier to remember than other colors, are psychological universals and a basis of color categorization.

In the contemporary, Chomsky-inspired demand for universal explanations in language, the alliance of these two theses from linguistics and from psychology founded the dominant line of subsequent research on linguistic categorization of color, which I will call universalist. This line is characterized, firstly, by defending the claim that there are universal patterns in how languages of the world categorize color, and secondly, by attempts to explain these universal patterns by appeal to universal properties of human visual perception. The rest of this historical sketch is mostly limited to the spine of the universalist tradition, that is, the works of Paul Kay and colleagues. If it neglects the immense amount of other contributions to the field, it is only because their substantial part is related to and oriented by this research line, be it negatively.

The Berlin and Kay (hereafter, B&K) research program has undergone several major revisions during the next decades, both as regards presentation of the cross-linguistic findings and as regards the exercised strategies of explanation. Originally, the color categories were identified across languages on the basis of sharing their focus, with little attention paid to categorical boundaries, which were found cross-lingustically unstable. From Kay (1975) on, it is emphasized that languages tend to fully partition the spectrum with whatever number of basic color terms they have. So, for instance, a language with three terms focused in, respectively, black, white, and red is likely to cover dark regions of the color space with the first, light regions with the second, and red together with purple, orange and yellow with the third. This is where the notion of a composite category (*e.g.*, the green/blue/black category) comes in, and the language evolution is then seen as progressive division of such composites, up to full individuation of the assumed 11 focal color points. Kay (1975) also somewhat diversifies the repertoire of attested stages that categorical systems can pass through in their development. Kay and McDaniel (1978) identified the first 6 focal colors of the universal sequence with putative 6 FNR (Fundamental Neural Response) categories,[1] thus founding a very influential nexus of anthropological linguistics and color science (Saunders and van Brakel, 1997a; Saunders, 2000).

From the later 1970s on, the World Color Survey (WCS) collected and processed data on color naming in 110 unwritten languages, in order to overcome the empirical limitations and methodological inadequacies of the original research.[2] The results of the WCS were progressively reflected in Kay et al. (1991), Kay et al. (1997) and Kay and Maffi (1999). These 1990s contributions further mitigate the strictness of the original generalizations. The cross-linguistic patterns of color categorization are presented, finally, in the form of 9 types of categorical systems spread over 5 evolutionary stages, based on how the assumed six privileged focal points are grouped by the color vocabulary of each language.[3] Kay and Maffi (1999) are also more restrained in relating focal colors directly to neurophysiology, although they maintain a universalist strategy of explanation. The WCS results appeared in their entirety as a delayed monograph Kay et al. (2009), the theoretical part of which remains on the 1990s level.

2.2 Criticism of the universalist research line

In the previous section, I wilfully avoided any deeper investigation and critique of the mentioned workings and claims and their conceptual background, which could easily grow to the extent of a book. I wish to treat the problem of color categorization from a present-day perspective, and the chronological line with its idiosyncrasies would be a rather inconvenient guide. As opposed to more recent works with a universalist flavor, including works by Kay and colleagues, I consider the so far mentioned contributions historical. They are historical in the sense that they have been subject to, in my opinion, conclusive critiques, which there is no need to repeat, unless it proves necessary for a contemporary treatment of the issues in question. After decades of both empirical and conceptual critical fire,

[1]See section 5.1.

[2]See sections 2.2 and 3.1.

[3]For instance, in Stage III (languages where the 6 focal points are split among 4 color terms) there are 3 attested types: [white; red; yellow; black+green+blue], [W, R+Y, G+Bu, Bk] and [W, R, Y+G+Bu; Bk].

original statements become obscure and hardly comprehensible, just as the simple question whether the Sapir-Whorf hypothesis is correct would seem nowadays.

The devastating critique concerns, first, the methodology of Berlin and Kay (1969), whose conclusions gained popularity, and in some related disciplines even reached a canonical status, quite disproportionally to their soundness. (See Saunders and van Brakel, 1997a; Dedrick, 1998, and their further references.) To mention only the most stunning fact, for 19 of the 20 examined languages the authors consulted only one informant, in each case a bilingual person socialized in the West. The status this work acquired would be rather inexplicable if we ignored the factors that I referred to above as "sociology" and "anthropology" of science.

Second, harsh criticism has been aimed at the explanatory principles that the B&K tradition uses. The attempt to reduce patterns of categorization in the world's languages exclusively to universal properties of human visual perception has been challenged, and so has been the very reality of the psycho- and neurophysiological entities and processes assumed in the proposed explanations. (Considerations of this kind, without a specific historical focus, are elaborated later in the thesis.)

Third, even the late 1990s contributions of this research line can be regarded historical in the sense that the "models" proposed there arguably fail to meet the desirable standards of scientific modeling. The standard I wish to relate to involves a formal model embodying a set of independently justified principles, and an evaluation against a reliable basis of empirical data. In contrast, in Kay and Maffi (1999), the last theoretical revision in this line, the boundary between the model and the data is blurred. The justification provided for the partitioning principles ("distinguish black and white", "distinguish the warm primaries and the cool primaries", "distinguish red") and for the order of their application is rather poor and not clearly independent of patterns in the data.[4] At the same time, the model is not evaluated against anything close to the raw data of the WCS, but only shown to fit an empirical pattern that is gained from the data by way of massive interpretation; namely, to the pattern resulting from classification of the WCS languages into the 9 types in 5 evolutionary stages of color categorization systems. This particular manner of classification, however, besides its lack of transparency, depends to a certain extent on the explanatory ideas of the B&K research line. In a nuthsell, the model arguably fits "the data" only through mutual adjustment of both.

Thus, the fragment of the field's history, which was introduced above, is only intended to provide a background for the following depiction of the two major positions that are represented in the broad debate on color categorization. Awareness

[4]Furthermore, my chapter 5 challenges Kay and Maffi's central assumption of primacy of the six "Hering colors" in perception. But the explicit principles of their model and their ordering remain questionable even when the assumption is granted.

of these will be important in treating the actual topics of this thesis.

2.3 Positions in the color term debate. Universalism and relativism

The B&K tradition in color naming research has been labeled universalist, and the main opposition in the debate may indeed be seen as one of universalism vs. relativism. But such a statement deserves some elaboration, since both these terms can refer to various theoretical positions.

As an empirical claim, universalism involves that there are certain patterns of color categorization (naming) that are universal, or shared across languages. The strength of such claims has varied, from the firm implicational hierarchy of color terms stated in Berlin and Kay (1969), to statements of cross-linguistic *tendencies* or non-randomly strong patterns in the data, as in MacLaury (1997b); Kay and Regier (2003); Regier et al. (2005). Universalism of this sort usually aligns with appeal to perceptuo-biological explanation, that is, explanation in terms of neurophysiology and psychology of color perception in human individuals. This is the case with the dominant, B&K research tradition, but not necessarily so. Wierzbicka (1996; 2006) instantiates another possible position in that she defends semantic universals (not in the domain of color though; her views on color will be discussed later) while relating them to universal human experiences based on shared features of the environment.

To the relativist side of the dispute we can count, in principle, all the voices that have cast some sort of doubt on the univesalists' empirical claims of whatever strength. This is often done by pointing out categorization systems of particular languages that do not easily fit the proposed generalizations. (Cf. Lyons, 1995; Lucy, 1997; Roberson et al., 2000, 2005) As the mainstream universalism has progressively mitigated its empirical claims, the relation of the two camps has become less of a direct conflict and more of emphasizing two sides of the same facts. (Cf. Dedrick, 2006.) A more substantial relativist critique concerns the very methodology behind the universalist findings, and in this sense there has been little rapprochement so far. In chapter 3 I extensively discuss the views of John Lucy and Barbara Saunders, prominent representatives of the relativist opposition.

As regards explanation (of whatever categorization patterns they recognize in languages), the relativists emphasize particular cultural, socio-historical processes, over which it is very hard to generalize cross-culturally. In reaction to the mainstream universalism, they tend to downplay the role of perceptuo-biological constraints. This position seems to adopt as much of the pre-B&K notion of arbitrary categorization systems as can be defended in the light of the whole subsequent research. At the same time, a relativist of this kind is not bound to subscribe to the relativism of the Sapir-Whorf thesis in an interpretation that claims an influence of language on perception or other "low" levels of cognition: it is possible to defend culturally induced variation in color categorization even upon acceptance of perceptuo-biological universality.

The universalist program is especially aligned with cognitive sciences having to do with color perception. On the contrary, the relativist positions are mostly grounded in cultural anthropology and typological linguistics, with ultimately Boasian roots, where relativism is a sort of default methodological stance, at times turning to conviction, though, or even value. (Cf. Dedrick, 1998, 2006; McManus, 1997; Saunders, 2006.) It is useful to keep in mind that, to some extent, the universalist vs. relativist dispute is not only a controversy over factual matters, but also an expression of divergence in fundamental attitudes. As Dedrick (1998, p. 146) has it: "The particularity of a culture's colour language—or of the culture itself—is as uninteresting to the cognitive psychologist as coincidence with the universalist scheme [...] is to the Boasian anthropologist." The latter's methods are description, interpretation and understanding (p. 135), rather than "objective" experimental procedures and modeling. The relativist side of the debate is, on principle, not interested in unified modeling – or in modeling at all, given that it is hard to think of a manageable model without substantial universalist assumptions. In Lucy (1997) or Saunders (2007) this is quite patent. It seems to be based on a deep conviction that the abstraction from particularity that any unified treatment necessarily brings about is unacceptable, thus ultimately on the value for understanding which one assigns to particularity.

In this sense, the present thesis unavoidably sides with the univeralist line of research, as it assumes the perspective of modeling. Yet the relativist critiques provide a lot of material for careful consideration, once we decide that, after all, we *want* to push unified treatment of the issue as far as it can reasonably go, and not to give up in advance. In general, there has so far been little mutual appreciation to that effect; with the relativists usually discarding the universalist program as a whole, and the universalists ignoring serious critical points from the other camp.

2.4 Further structure of the thesis

Nowadays, one thing seems clear: it has been harmful for the debate that a clearer division has not been made between the cross-linguistic data presented in favor of universalism (ultimately, the outcomes of the WCS), and the universalist explanatory strategies employed. Both have largely been put forward by the very same people, in the same publications, to some extent even in mutual shaping.[5] This

[5] See section 2.2.

rhetorical conglomerate then provided a single visible target, namely the B&K program, for both respective strands of relativist critique, one aiming at worth of the gathered data, the other at plausibility of the proposed explanations. Problems of both types have been pointed out in attacks on the B&K research line, yet the logical independence of the two issues passed rather unnoticed. The obvious space for diversified positions concerning both these issues has been concealed by the rhetorical opposition to the universalist program. Thus, the literature produced until now offers either full adoption of the WCS results as a representation of the cross-linguistic reality of color categorization, or a complete dismissal of these data, but nothing in between. Given that the relativists give serious reasons to back up the latter choice, it is clear that the WCS data (not to speak of their usual universalist interpretations) should not be taken at the face value. However, it is as implausible a conclusion that this body of work, involving a few dozens linguists, thousands of informants all around the world and long years of processing, focused on color categorization in the world's languages, fails to reveal *anything* about the phenomenon, due to known methodological problems. All those people would need to have worked really hard in order to conceal from us everything there is to be known on that matter.

If we assume a modeling stance with all it involves, we cannot avoid a question that is external to any proposed model; namely, what are the most reliable empirical data on the basis of which we could evaluate the model? As argued above, for the phenomenon of color categorization by languages of the world, this question has so far received very little unprejudiced treatment. My assumption is that such data are likely to be found in the outcomes of the WCS, the most extensive research ever done on the topic.[6] However, this "reliable core" can be localized in the superset of all WCS results only upon close consideration of the relativist criticism that has been raised against the WCS. This is the agenda of the next chapter.

As to explanatory strategies, the notion that it takes both perceptuo-biological constraints *and* cultural processes to derive the existing cross-linguistic regularities and variance in color categorization has become fairly common in the last two decades. (Cf. Dedrick, 1998, 2006; Foss, 1997; McManus, 1997; Poortinga and van de Vijver, 1997; van Kruysbergen et al., 1997; reserved formulations to the same effect are also found in the opponent camps, cf. Kay and Berlin, 1997, and Saunders and van Brakel, 1997b.) It is one thing, though, to vaguely notice the role of both types of factors, and quite another thing to capture the nature

[6] One more impressively wide cross-linguistic survey on color categorization is the Mesoamerican color survey by MacLaury (1997a). I leave this empirical work completely aside, with the idea that it is a possible further instance in verifying that a model with good predictive power with respect to the WCS outcomes presents a plausible explanation of color categorization in general.

and interaction of such factors in any concrete manner. For some time, general modeling approaches to the phenomenon backed away, in contrast to the boom of various branches of specialized research on color categorization, as witnessed by Biggam and Kay (2006a,b); MacLaury et al. (2007); Biggam et al. (2011). These flourishing branches include at least the following: reflections of recent progress in technical color appearance models and in neurophysiology of color; detailed descriptions of color naming systems in particular languages from an ethnosemantic, diachronic or sociolinguistic perspective; research of (what is called) categorical perception of color and how it is influenced by language; investigations into the role of individual differences in perception of color; research of color preferences.

Much of what happens in these research branches is potentially relevant for purposes of a general modeling approach concerning our focal phenomenon. In the last couple of years, the interest in color categorization modeling has increased again and the research took a game-theoretic drift, after the impulse of Steels and Belpaeme's (2005) case study of the evolution of shared color categories. However, most of these late models reflect little of the relevant progress in the related fields, and the authors do not occupy themselves with an appropriate justification of their explanatory principles. (Cf. Regier et al., 2007; Baronchelli et al., 2010; Loreto et al., 2012; Correia and Ocelák, 2013.)[7] A somewhat different case is that of Jameson and Komarova (2009a) and Jameson and Komarova (2009b), where certain explanatory components of the model are motivated very thoroughly (see section 4.9); nevertheless, other explanatory factors are taken for granted or abstracted from and there is no evaluation against empirical categorization data.

After dealing with the question of the relevance of the WCS data in chapter 3, the rest of the present thesis (chapters 4 to 6) is devoted to examination of what would be plausible, independently justifiable (that is, reasonably up to the present state of the research fields in question) components for an integrative model of the phenomenon at hand. Importantly, the thesis only covers what I wish to call the "perceptual basis" of explanation, that is, the factors that have to do with the perceptual makeup of human individuals – with how individual agents perceive color. It is meant as a basis upon which agent-based evolutionary game-theoretic interaction can take place, roughly in the spirit of the recent models. Various aspects of such interaction require at least as much (and have so far received as little) attention and justification as the perceptual layer of color categorization modeling. However, investigation into these aspects of the complex

[7]Again, sociological attention to peculiarities of the contemporary scientific circulation would be advisable, if one were to explain why, *e.g.*, in Loreto et al. (2012) highly advanced statistical methods and simulation techniques can combine with quite precarious explanatory assumptions, such as using a physical color space as opposed to a perceptual one (cf. chapter 4), and with adoption of the unqualified color sequence of Berlin and Kay (1969) as the empirical base of evaluation.

color categorization problem had to be left for another occasion.

2.5 Some reservations

At this point, it seems apt to state some qualifications regarding what my project is *not*, in order to avoid certain objections that have been raised against the B&K program by the relativist side of the dispute. (Cf. Saunders and van Brakel, 1997b; Saunders, 2000, 2006.)

The project is not primarily engaged in philosophy of semantics. In particular, it does not attempt to promote the view of meaning that Saunders labels "cryptographic" and rightly derides, the picture of color terms as direct encodings for sets of color sensations, construed in the lines of old British empiricism, let alone for sets of Munsell color chips. Irrespective of whether workings of the B&K line subscribe to such a conception of meaning or not, a modeling strategy can do well without it. I fully side with Saunders' (2006) expressed view that meaning of a color term is a normative matter, constituted in social practices of the particular community. No less am I sympathetic to the Brandomian notion, referred to by her, that concepts (*a fortiori*, color concepts) are (co-)defined by their inferential roles, rather than being mere classifiers. (Brandom, 1998; cf. also Marconi, 1997, on referential and inferential semantic competence; Lyons, 1995, on color term use of the blind; Malcolm, 1999, on Wittgensteinian "color grammar".) Yet, in my opinion, there is no deep incompatibility between these semantic views on one hand, and *schematizing* meaning of a color term on the Munsell color chart, or letting informants name color chips with color terms, on the other. Admittedly, given the endorsed semantic views, these procedures should be considered shortcuts or heuristics. Still, with careful interpretation, they appear to be fairly reliable with respect to what could be revealed by observing social interaction, the "true" domain of meaning. This is not to say that outcomes of such heuristic procedures have always been interpreted with appropriate care, as will be made clear in the following chapter.

Secondly, and in connection to the previous, I do not maintain a specific epistemological position. In particular, I am not interested in defending the empiricist foundationalism which Saunders (2000; 2006) tends to see in the B&K line and criticizes with reference to Sellars' (1956) "myth of the given". I do not think that anything about my project involves a commitment to some sort of empiricist, non-inferential foundations for judgments; and for my purposes I do not consider it necessary to keep impressions, noticings, and observations reports of the informants explicitly apart, as Saunders (2000) does.

2.6 Central notions

I will now introduce a couple of notions that are quite central to the debates on color categorization. The definitions here are only preliminary, in the sense that, at some point of this work, most of these notions will be called into question. The point of proceeding from tentative, popular definitions is to provide basic understanding of what theoretical work people in the field tend to load on the concepts.

Hue, *lightness/brightness* and *saturation* are three commonly acknowledged dimensions of the phenomenal color space of normal observers. By "color", in English, people preferentially refer to differences in hue; hue is the dimension along which one proceeds, say, from blue to yellow via green. Lightness (for light-reflecting surfaces) or brightness (for light-emitting bodies) of a color is loosely correlated with intensity (luminance) of the emitted or reflected light. Saturation can be in other words defined as the phenomenal purity or pronouncedness of a chromatic color. While one can use the physical parameter of luminance in order to roughly mediate the idea of lightness as a phenomenal attribute, the relation of hue and saturation to the physical character of the stimulus is less straightforward, due to the metameric character of human color perception (cf. section 4.2). It holds only under quite specific conditions that hue is roughly correlated with the dominant wavelength of the light that is momentarily affecting the retina, and saturation with relative intensity of the light at this dominant wavelenth.

Basic colors, focal colors, color prototypes, perceptual categories of color, unique hues and *primary colors* are terms tied to various theoretical attempts to establish certain points or regions of the color space as in some sense privileged. (Cf. Saunders and van Brakel, 1997a, p. 167.) It has been sometimes assumed that some of these notions single out an identical set of colors; therefore these labels have at times been mixed up in use. *Basic colors* are whatever is referred to by *basic color terms*. These are defined in Berlin and Kay (1969) by linguistic and mild psychological criteria (cf. section 3.2.3), so that they pick 11 color terms for English, including orange, purple and gray, but excluding violet, turquoise or crimson. *Focal colors* are, in the B&K tradition, the points of the color space that have been explicitly marked by speakers of a language as best examples of basic colors in that language. Both basic colors and focal colors are thus primarily defined in relation to particular languages, although this has not been emphasized in the universalist spirit of the B&K program. *Color prototypes* are color points or regions that have been regarded as universally privileged in perception based on psychological evidence such as accuracy in memorizing colors or reaction times in various cognitive tasks. This notion is linked to universalism of Eleanor Rosch; in contemporary psychological research on color it is usually replaced by the more neutral notion of *perceptual categories*, which is analyzed in chapter 6.

Unique hues refers to the set of colors that are supposed to be phenomenologically primary (constitutive of color appearance), as well as linguistically salient in the sense that other colors can be described in their terms. Usually, this notion is assumed to select four colors: focal red, yellow, green, and blue. It is critically examined in chapter 5. The notion of *primary colors* is mainly a technical one: a set of primaries is any minimal set of colors of light (*e.g.*, red-green-blue) such that mixtures of these colors are able to match any perceivable color (cf. section 4.2).

Chapter 3

The WCS data. Critique and defense

3.1 The World Color Survey and the data

The World Color Survey (Kay et al., 2009; Kay and Cook, in press) investigated linguistic categorization of color in 110 unwritten languages of 45 language families.[8] On average there were 24 informants consulted per language and the modal number was 25; however for some languages the number of informants was as low as 6 (Náhuatl, Western Tarahumara). Apart from consulting more representative amounts of preferably monolingual speakers living in their own community, the methodology of the WCS research departed from that of Berlin and Kay (1969) in another important respect. Berlin and Kay (1969) first elicited color terms for a language in question, determined the basic ones, and then asked the informant(s) to draw the extension and focus of each basic color term on the Munsell color array. In the WCS, to the contrary, the 330 color chips were presented to the informants successively, in a fixed random order, and the informants were asked for a short name in response to each. The researchers subsequently asked the informants to determine best examples for salient color terms, and only in this task the full Munsell array was displayed.[9] For further details on the WCS methodology, see

[8]Listed also in Regier et al. (2005). The field data were gathered mostly in late 1970s, whereas the subsequent processing drew out over three decades.

[9]The basis of the employed color array is Lenneberg and Roberts' (1956) array of 40 hues (columns) at 8 levels (rows) of lightness, where each chip is on the maximal level of saturation that is available, for that particular hue-lightness combination, in the Munsell color order system (cf. section 4.1). The Munsell system is construed so that in any single of the 3 dimensions, the perceptual distance between any two neighboring chips is the same. There is however no determinate relation between perceptual distances in different dimensions. At the same time, the chips in the Lenneberg and Roberts array are on various levels of saturation. It follows that

Kay and Berlin (1997); Kay et al. (2009); Kay and Cook (in press).

Various summaries and interpretations of the WCS results have been published, usually in support of the universalist positions, notably Kay et al. (1991, 1997); Kay and Maffi (1999); Kay and Regier (2003); Regier et al. (2005); Kay (2005); Regier et al. (2007). For the present thesis it is more important that in Kay et al. (2009) and the on-line available WCS data archives (Cook et al.), the gathered data for all 110 languages are presented in great detail and in a considerably raw state, prior to further interpretation with possible universalist biases. That makes them a good candidate for the empirical base of evaluation for any model of linguistic categorization of color – but only as far as these data can be defended from the relativist critique concerning, ultimately, their representativeness with respect to the phenomenon at hand. The agenda of the present chapter is to examine this criticism, in order to delimit a subdomain of the WCS data that could stand the relativist objections, and therefore could be considered representative of the general phenomenon.

Let me first briefly describe the organization of the detailed WCS data that were made available. The WCS data archives (Cook et al.) offer virtually all primary data in the "atomic" form; that is, they allow one to track how a particular speaker (say, #22) of a particular language (say, Mazahua) named a particular chip (say, #287) of the Munsell array, and also what chip(s) she marked as the best example(s) for that particular name. Kay et al. (2009) release more digestible generalizations of these data for each of the examined languages.[10] From the presented figures, especially informative are the *aggregate naming arrays* (ANAs). Given a particular language, the array graphically displays with which basic terms the 330 color chips were generally labeled. There are such arrays for various levels of interindividual agreement: only those chips that reached the given level of agreement among speakers are marked. For each language, two of the possible ANAs are naturally salient among others that are on arbitrary levels of agreement. The first is the "modal agreement array", which is on the highest level of agreement such that the whole array is still covered.[11] The second is the "full agreement

even within individual rows or columns of the 320 chips array, neighboring chips can only be perceptually equidistant by and large. To the Lenneberg and Roberts array, the WCS adds a column of 10 achromatic chips from white to black. The reader may get an idea of the resulting array from an image that is available at the on-line WCS data archives (Cook et al.), with the reservation that faithfulness to the standardized physical set of Munsell chips is generally not guaranteed on screening devices.

[10]For the present purposes we can disregard the fact that with support in these data they characterize each language as fitting a specific slot in their evolutionary scheme of types in stages. That is, arguably, a highly interpretive and reductive step. The present chapter is only concerned with the raw WCS data, not with universalist theories and interpretations accompanying them.

[11]That is equivalent to the following: a modal agreement array is one that for each of the 330 chips gives the term with which the chip was most often named. Modal agreement arrays

array", which marks only the chips upon which all informants agreed (and so, in most cases, a substantial part of the array remains empty). The possibility of explicit attention to the degrees of interindividual agreement on color term use will be important for my later arguments.

The available WCS results also offer a picture of categorical focality, gained via asking the informants to mark the best examples of basic colors. In the following, though, I will have no role for this part of the data. Even in the B&K line, there is progressively less emphasis on focal colors, as Berlin and Kay's (1969) original notion of universal focal points proves untenable. Kay et al. (2009) note that in the WCS, individual best example choices were unexpectedly dispersed, although there are still clear tendencies, and they do not support anything in their characterization of individual languages with the data. For further discussion of focal colors in the WCS, see MacLaury (1997b); Dedrick (1998); Saunders and van Brakel (1997b); Jameson (2010). For me, the main reason to disregard focal choices is that the extension of a category on the Munsell chart is more informative of that category than its mere focus,[12] and that the latter is obviously not independent of the former. If a model can successfully predict categorical extensions, then placing foci simply in the middle of the predicted categories to some extent guarantees success in prediction of foci. I therefore suggest that models of color categorization be evaluated only against data on categorical extensions.

3.2 Criticism and defense of the WCS data

3.2.1 Can there be universal patterns of color categorization?

Anna Wierzbicka's (1999; 2006) general objection to the B&K program, including the WCS, is the following: many languages lack a word for "color"; so color is not a universal concept;[13] therefore "the idea of 'color universals' – conceived of as universals of language and thought – is self-contradictory. There can be no universals in how people think about color given that in many languages people do not talk about color at all." (Wierzbicka, 2006, p. 2.) This protest is serious, as it implies that the WCS imposes on speakers a universal domain of color where

precisely correspond to the "modal maps" that are employed in Regier et al. (2007) and Regier et al. (2009).

[12]Cf. Dedrick (1998). A category is defined by specifying membership. Stating extensions on the Munsell chart provides, in a sense, exhaustive information about categorial membership (although it abstracts from certain semantic phenomena, such as vagueness and prototypical effects). This is not the case if we merely specify foci, unless we also provide principles for partitioning of the non-focal chips.

[13]This is presented as a strong probabilistic conclusion, which seems plausible. Wierzbicka does not claim that absence of a word in a language strictly implies absence of a corresponding concept.

there in fact *cannot* be any. But I find Wierzbicka's argument fallacious. Absence of a general concept of color in a language is not a reason to conclude that its speakers do not talk about color, or rather *colors*. If they have a specific set of words and concepts with similar function to, for instance, "red" in English, then they *do* talk about colors. True, they could not confirm that they do, since they lack the concept of color. But that they do is a claim in English (i.e., a claim that we make, using English, which happens to have a term for the general concept of color), and a correct one in this case. Thus also the claim that there are universals in how people talk about color(s) is not self-contradictory given the assumption that some people lack the concept of color.

That is not to say that it is a correct claim. Of the peoples who lack a general concept of color, some may very well also lack specific expressions anywhere close to the color terms we know from the Western languages. Then it is appropriate to conclude, again in English, that those speakers do not linguistically categorize color. (Even then we can go on talking of regularities of color categorization, if we merely substitute more restrained terms like "cross-linguistic tendencies" or "universal patterns" for "universals".) Whether such is the case of any particular language is an empirical question, and I want to argue that the WCS is reasonably suited to provide an answer. "To K&K [Kay and Kuehni], Warlpiri visual descriptors are semantically 'colour terms', because they want them to be, semantically, 'colour terms' [...] To Warlpiri-speakers, however, they are not colour terms because Warlpiri people do not (did not) think about the visual world in terms of 'colours'." (Wierzbicka, 2008.). It seems clear to me that neither of the two described opinions is relevant for the problem of color terms in Warlpiri; the question is whether it is *appropriate* to state in English that Warlpiri have color terms. Indeed, the data on Warlpiri presented in Kay et al. (2009) do not seem to support a strong affirmative answer. But what justifies Wierzbicka's assessment of Warlpiri is precisely such data, and *not* the fact that the Warlpiri would not understand the question.

Contrary to what one might expect, Wierzbicka is in fact a proponent of semantic universals, not a general relativist. (Cf. criticism in Saunders and van Brakel, 1999.) Her position on color follows from the fact that in her set of roughly 60 semantic universals (such as *you, body, good, see, die, because* or *above*; Wierzbicka, 1996, 2006), gained through empirical cross-linguistic work, neither the general nor any specific concepts of color are represented. While I respect her empirical work in cross-linguistic semantics, I strongly disapprove of her metaphysics of meaning with its apparent reifications of structuralist semantic component analysis, and the resulting artificial opposition between semantic universals and the rest of language semantics. The basic confusion seems to lie in the redescription of the empirically attested semantic universals as semantic *primes* or *primitives*, that is, "elementary units of meaning out of which all complex and culture-specific meanings are built"; and it is aggravated by Wierzbicka's unfortunate adoption of "chemical" terminology of semantic "atoms" and molecules". It is quite unclear why a concept that is universally attested should automatically qualify for a building block of meaning, and indeed, what exactly would that mean. Yet this picture leads Wierzbicka to obscure assertions regarding semantic componency[14] and to drawing a thick line between some 60 "primitives", which are supposed to be universally shared and identical across languages, and all other, strictly language-particular, complex meanings.

From the perspective of the conception of meaning to which I subscribed in section 2.5, this is not plausible. If a concept, or the meaning of an expression in a language, is determined by the whole of social norms governing its referential and inferential use, then a concept can be at best *by and large* identical across languages and cultures. For illustration, take one of Wierzbicka's universals, *die*. While we may trust her claim that this notion is a considerably universal one, it simply *cannot* be an identical concept for two communities that substantially diverge, *e.g.*, in their ideas of one's posthumous fate. For my purposes, the point of deconstructing the absolute notion of a universal concept is the following. Speakers of a language that has no general concept of color cannot have a notion of red *identical* to our own notion of red, since their potential notion of red cannot be co-defined by red being a color. Nonetheless, other similarities, particularly in referential application of corresponding terms, can be sufficient for us to conclude that the language in question after all *does* have a reasonably comparable term for red; or that it has terms which we recognize as color terms although they have no direct equivalents in English. Again, whether a particular language is like that is an empirical issue, and so far nothing prevents us from appealing to the WCS data for decision.

To sum up my position contra Wierzbicka: despite the lack of a general concept of color in some languages, it is still an empirical question whether there are universals or universal tendencies in linguistic categorization of color. This question should be decided on the basis of the WCS. At the same time, it should not be understood in Wierzbicka's sense as one of presence or absence of identical semantic atoms in languages, but as one of *by and large* identity of concepts, as determined by the whole of norms governing their use.

[14]"'Size' is not a semantic component of 'big' and 'small'; on the contrary, 'big' and 'small' are both semantic components of 'size' (just as 'mother' and 'father' are both semantic components of 'parent', rather than the other way round). 'Colour', on the other hand, is indeed a semantic component of the English word *blue*." (Wierzbicka, 2008, p. 888.)

3.2.2 Are there such patterns?

The relativist critics of the B&K program often question the use in data collecting of the Munsell color chips, which are only diversified in hue and lightness and, unsystematically, saturation. They are inclined to see it as imposing a universal domain of color on speakers for whom color is not an autonomous linguistic and conceptual domain in our sense. One classical reference is Conklin (1955), who, having attempted to describe color categories in Hanunóo, concluded that in this language, categories of visual experience intermingle color with aspects of freshness and moisture. Lyons (1995) points to similar tendencies in classical Greek; in other languages, color is blended in categorization with ripeness or evaluation (Saunders and van Brakel, 1997a). From reported experiences of field linguists working on color naming it is quite clear that "colour as hue is not everybody's interest" (Gage, 1995, p. 188). "Many [informants] would simply stare at the array. [...] Several attempted to provide a different name for each perceptually different chip, employing terms that later proved to be names of trees, plant dyes, and parrot feathers. One informant, when asked to show where all the red chips were, took the pen and very carefully circled the entire board." (Saunders and van Brakel, 1997a, p. 174). "It is noteworthy that in the WCS, fieldworkers were regularly driven to comment that consultants had 'eye disorders', were 'colour blind', 'problematic', 'messy', or in one case, behaved like a cretin." (Saunders, 2000, p. 95.)

Apparently, not all languages have something we could plausibly call color categories. But the relativist critique is wrong when it suggests that the WCS methodology shapes the data in such a way that they enforce a conclusion that all languages do. (That is a way we could explicate the accusation that the Munsell color array imposes universal domain of color on the natives.) The key is to pay attention to the level of interindividual agreement in color naming. If speakers of a particular language are not accustomed to more or less abstract naming of colors, to use of color related terms that are not restricted to specific contexts,[15] they will certainly not reach a high level of agreement in individual naming of the color chips. The WCS elicitation procedure, which demands that the informants name hundreds of color chips, may be unusually frustrating for such speakers; but the outcome, if carefully interpreted, will not lead us to the conclusion that there are color categories in a language in which there are in fact none.

The importance of the level of agreement has been neglected in both the relativist and the universalist tradition. In Regier et al. (2007) and Regier et al. (2009), languages of the WCS are represented with "modal maps" of their categorical systems, which precisely correspond to the modal agreement arrays described in section 3.1. On one hand, this is a substantial advance when compared to the highly theory-driven representation of languages by their type and stage in

[15] As, for example, specialized cattle color terminology is.

the B&K evolutionary scheme. On the other hand, modal agreement arrays do exactly what the relativists accuse the entire WCS of: they by definition present any language as "standardly" categorizing color. That is because for any language, the modal agreement array is on a level of agreement *as low as needed* in order to have at least that level of agreement for each of the 330 chips. Yet in case of some languages (Gunu, Halbi, Ifugao), as little as 12 % or 8 % of speakers (3 or 2 out of 25) agreed on names for particular chips (Kay et al., 2009). Actual systems of color categorization in languages cannot be represented by modal agreement arrays; in order to represent them (and evaluate models against them) one needs some of the intermediary arrays between modal and full agreement, on a reasonable agreement level.[16] For languages such as Warlpiri, this will preserve the possibility to manifest that they in fact do *not* categorize color at all, namely by leaving most of the higher-agreement array empty; or to display that they have *some* established color categories but are nowhere near full categorical partitioning of the array.[17]

As to languages such as Hanunóo, Lucy (1997) is right that the WCS outcomes leave no chance whatsoever to retrieve other possible aspects besides hue, lightness and saturation, such as freshness, that may be reflected in their categorization of visual experience. Therefore, the WCS is indisputably reductive in the sense that it cannot provide us with all there is to know about basic visually grounded categorization in languages. But again, appropriate attention to levels of agreement secures that, even if we do not record everything there is to record, we will not observe color categories where there are none. The reasoning above can be repeated. If, for some speakers, color aspects are inseparably bound in categorization with other visual or material aspects, these speakers are not likely to achieve high agreement in abstract naming of color chips. If they do score high in agreement, it can be taken as evidence that color is, in a linguistic sense, a considerably autonomous domain for them.

3.2.3 Linguistic criticism

The opponents of the B&K program have also raised objections of more narrowly linguistic nature. First, they have often criticized the a priori restriction of the

[16]Here, I wilfully leave this "reasonable" level of agreement unspecified. While it seems obvious that 15 % will not do and 95 % is way too strict, I do not see any conclusive reason to set the limit at, *e.g.*, 70 % rather than 55 %. Neither do I believe that there is one universal solution. The question whether a particular rule constitutes a social (specifically, semantic) norm of the community at hand depends not only on the percentage of the population that acts in compliance with this rule, but also on the social status of the followers.

[17]To illustrate how diverse the WCS languages are in this respect, we can consider Cree, Chumburu, and Guarijío, as captured by aggregate naming arrays in Kay et al. (2009). In Cree, mere 7 chips out of 330 (that is, less than 3 %) reach the 64% level of agreement; in Chumburu it is 158 chips (48 %) for the same agreement level; in Guarijío it is 267 chips (81 %) at 60% agreement.

research, from Berlin and Kay (1969) on, to the "basic color terms" (BCTs). (Cf. Lyons, 1995; Lucy, 1997; Wierzbicka, 2006.) These are defined as color related expressions that are monolexic, not hyponymous to other color terms (such as "crimson" is to "red"), not contextually restricted (such as "blond", "rubicund", "bay" are to, respectively, hair, faces, and horses), and that are "psychologically salient" (with further subcriteria). It is quite trivially correct to say that one cannot describe the entire complexity of color categorization in languages of the world taking into account only the terms that pass the criteria for being labeled "basic". But that does not make this restriction arbitrary.[18] It seems quite natural that those who are after cross-linguistic patterns in color categorization should first seek them in the domain of color notions that are most central in conceptual systems of their languages; and that is what BCTs by definition are.

Lucy (1997) implies that the WCS data are valueless due to the authors' complete ignorance of grammatical structure of any single language involved. In particular, there is no reflection of grammatical status of the terms the speakers use when naming the color chips. Thus, what is referred to as the sets of BCTs of individual languages is actually an uncontrolled mixture of nouns, adjectives, participles or even verbs.[19] Although this is true, it is hardly a devastating objection.

For one thing, Lucy's critique originates in a fundamentally different background that that of the WCS authors, namely in descriptive typological linguistics of Boasian ancestry. For him, doing linguistics primarily involves detailed systematic description of individual languages. In particular, such a description cannot by any means avoid an in-depth investigation into the language's grammar via distributional analysis. It is quite natural that when measured by this ideal, the WCS appears to Lucy as a bad parody of proper linguistics (Lucy, 1997, p. 330). But that is of course a misunderstanding of what the ambitions of the WCS are, or in any case what they should be. No one claims that a WCS outcome for a particular language comprehensively depicts the entire complex system of color-related expressions and concepts of that language. Only the kind of linguistic work to which Lucy appeals can provide something like that. On the other hand, such exhaustive treatments are not necessary in order to reliably reveal some prominent universal patterns of color categorization. The WCS has some indisputable flaws, but its essential advantage consists in handling the 110 languages in a uniform way, thus greatly facilitating generalizations. In Lucy's tradition, hundreds

[18] Wierzbicka, (2006, p. 20): "Shweder and Bourne (1984: 160) note that, by choosing, *arbitrarily*, this particular set of criteria, Berlin and Kay have excluded from their field of vision ninety-five percent of the relevant data. In my view, this *excellent observation* [...]" (Italics are mine.)

[19] What is more, the authors, according to Lucy, constrain basic color terms by an accidental feature of the grammar of English, namely that they be monolexic. In my opinion, though, this constraint is justified by the fact that there is an undisputed universal linguistic tendency towards lexicalization of salient notions.

of languages have been comprehensively described until now, but so far nobody has attempted to span the multiplicity of the involved methodologies to formulate a generalization about color categorization on the basis of such "proper" linguistic work.

To address Lucy's objection regarding ignorance of grammar in the WCS more specifically, I would like to defend the following principle. In the WCS methodology there are certainly factors by which the resulting picture of cross-linguistic patterns of color categorization is distorted to some extent. Some of these factors are relatively harmless, namely those which cannot be shown to work towards any particular bias of the resulting picture. While grammatical status of basic color terms is indeed ignored in the WCS, there seems to be no evidence for the claim that, *e.g.*, adjectival BCTs in general categorize color differently from nominal BCTs, or that any observed patterns would have to be reassessed if grammatical distinctions were taken into consideration. On the contrary, what we should primarily force out from the WCS results is the influence of the factors that bring about a specific, unequivocal bias. Using modal agreement arrays to represent individual languages, a practice that I criticized above, is one of such factors, for it incorrectly presents languages as though color was fully categorized in all of them.

3.2.4 Impact of culturally dominant languages

There is another factor with a clear impact on the resultant picture of the WCS which requires consideration and bracketing, namely the influence of culturally dominant languages, particularly (but not exclusively) of the languages that were once engaged in the Western colonialism (English, Spanish, French, Portuguese etc.). Although this is a matter of the cross-linguistic reality captured in the WCS, rather of the WCS methodology, it will be immediately made clear why it is appropriate to discuss it here. Western influence on some of the WCS languages has been noted by both sides of the debate. Since it is clearly a factor in support of cultural relativism rather than of perceptuo-biological universalism, it has been naturally emphasized by the former side of the dispute and downplayed by the latter. (Cf. Saunders and van Brakel, 1997a; Kay and Berlin, 1997). But saying that "some", "many" or "not all" WCS languages are impacted by culturally dominant languages is just presenting different sides of the same facts. While genetic-wise, the languages involved in the WCS are respectably diverse, the sample was not construed so as to filter out effects of language contact, in particular the Western influences.[20] At the same time it is relatively clear towards what patterns of

[20]Kay and Berlin, 1997, p. 201: "The selection of the WCS sample from the full population of the world's unwritten languages was determined primarily by the presence or absence of a Summer Institute of Linguistics missionary linguist in the field area." Saunders and van Brakel

categorization this factor works in general, because the languages that are culturally dominant are in general characterized by well-established and comparatively fine-grained color categorization, unlike some of the languages under their impact.

Now, it is perfectly true that such influences constitute the present-day reality of color categorization in languages of the world, with whatever universal patterns it displays. Nevertheless, in the modeling perspective which I elaborate in the rest of the present thesis, largely on a perceptuo-biological basis, I will have little to say about this factor; and I would claim that nothing can be done about it in modeling except for *ad hoc* integrating some kind of representation of the known history of language contact. Therefore, my concern is strictly speaking not with modeling of the contemporary cross-linguistic reality of color categorization, but rather with modeling what this reality *would have looked like* if languages had been allowed to develop independently of each other.[21]

As the basis of evaluation, consequently, we should select data that represent something which comes, to our knowledge, as close as possible to such an ideal reality. That would necessarily involve a serious reduction of the WCS sample, beginning with languages like Kriol, where most of the informants "exhibit Stage VII color terminology with eleven basic terms all derived from English". (Saunders and van Brakel, 1997b, p. 217.) How far this weeding should go, that depends ultimately on the trade-off between a desirable level of developmental "purity" of a language and the minimal size of a sample that is necessary to evaluate anything at all.

3.2.5 Circularity in the WCS?

At some point of their critiques, both Lucy (1997) and Saunders (2000; 2007; Saunders and van Brakel, 1997b) accuse the B&K program of some sort of circularity or self-confirmation. This criticism is aimed, without clear differentiation, both at the procedures of collecting and processing data, and at explanatory ideas of the universalists, especially their conception of visual perception. Here I only discuss relevance of such objections for assessment of the WCS data; for the rest, see footnote 26.

I have already dealt with the relativist suspicion that the very use of the Munsell color array in the experiment implies that universal engagement of languages

(1997b, p. 217) quote field comments from the WCS that indicate influence of English, Spanish, French or Hindi on Agta, Carib, Halbi, Kriol, Mazatec, Papago and Zapotec.

[21] A successful model in these lines will thus not univocally support one side of the universalism-relativism opposition as far as the contemporary cross-linguistic reality of color categorization is concerned. That is not a problem, since this opposition does not present, as I have argued, a meaningful either-or question. Instead, the model is meant to specify the role of various factors emphasized by either of the sides.

in color categorization will be "found". The conclusion was that with appropriate attention to levels of agreement the color array does not impose anything like that. For any particular language, it allows either a refined specification of its color-categorical system, or conclusion to the absence thereof. In that sense, the array should be seen as an *ally* against universalist interpretative pigeonholing of languages into 9 types (such as [W, R+Y, G+B, Bk]) in 5 evolutionary stages (cf. sections 2.1 and 2.2), not as a means of reading universal order into empirical data on languages. Lucy (1997, p. 334) condenses a good deal of criticism into the following, seemingly devastating remark: "The procedure strictly limits each speaker by rigidly defining what will be labeled [1], which labels will count [2], and how they will be interpreted [3]. To use a political metaphor, it is as if one political party were entitled to dictate what you would vote on, to count the votes, and to report what the results meant." But under a scrutiny, the objection falls apart. (1), that is the choice of the Munsell chips, was defended above (3.2.2) and found justified; so was (2), the restriction to basic color terms (3.2.3). (3) is groundless provided that we represent languages with aggregate naming arrays, without further interpretation of their color categories (in case they have any) as "red", "green-blue", etc., as has been usually done in the B&K program.[22]

But the relativists are not only worried that the WCS methodology ensures *that* universals will be found; they also suggest that it somehow generates the specific patterns they grant there to be in the data. "After all, as apologists for this tradition often note, it works! These color systems are there! [...] Well, I agree that something is there, but exactly what? [...] This approach [...] not only seeks universals, but sets up a procedure which guarantees both their discovery and their form." (Lucy, 1997, p. 331.) "There is plenty of order in the WCS. This order, however, is partly apparent and partly real, and the part that is real is "real" in different senses. Part of the order is created by the method used (the Munsell system) [...] Second, order is created by data processing (Cf. Table 6)." (Saunders and van Brakel, 1997b, p. 218.) As we can see, there is no disagreement about the fact that in the WCS data there are some manifest patterns. To mention some that are not intuitive from the western perspective, these patterns include a cross-linguistic tendency to merge (what we call) green and blue into one category, and to a lesser extent also red and yellow.[23] The relativists do not deny such strong tendencies in the data, but they suggest that they are an artifact of the chosen

[22]Cf. Lucy, 1997, p. 334: "So when a category is identified now [on the Munsell color array], it is really the investigator who decides which "color" (or "composite color") it will count as."

[23]Especially the first of these generalizations seems to be so strong that it can be safely expected to survive the reduction of the empirical WCS basis to its "reliable core" which is being proposed in the present chapter. Hence, I occasionally mention merging green and blue in a single category as an example of a salient cross-linguistic pattern of color categorization, without further qualifications.

method.

However, very little justification is given for such a strong claim. Virtually no argument is offered that would specify how exactly is the WCS methodology supposed, for instance, to encourage categorical merging of blue with green and to suppress merging of blue with red. This is not to deny that, in the data processing stage, there may have been individual, "atomic" cases where a problematic data point was treated in compliance with universalist preconceptions. For instance, if there were two or more names for a particular chip equally agreed on, the researcher might have inclined to record in the aggregate naming array that name which best fitted the contemporary universalist picture of color categorization. (Cf. explicit tables of "ties" in Kay et al., 2009.) But the room that was present in the WCS for distortion of this kind is nothing but marginal compared to robustness of the patterns in the data (which, to remind, reflect over half a million "atomic" acts of naming a color chip). Saunders' (2007, p. 472) complaint that "The 'data' [...] contain countless examples of the influence of prior expectations, colonialism and global standards, as also of 'outliers' discarded because they did not fit the preconceptions of the investigators" can be rejected: in part as exaggerating rhetoric, in part as referring to the practices of the universalist interpretation of the WCS data which we discard as well.

Saunders and van Brakel's (1997b) Table 6, which presents a collection of potentially subversive field comments from the WCS, seems to be the most concrete indication of how the relativists suppose that universal patterns are created during the data processing. Some of the included comments only illustrate the influence of culturally dominant languages, which I approvingly discussed above. But the rest fails to support the relativist point, since nearly all of these comments are disarmed once we subscribe to representing languages by means of the higher agreement aggregate naming arrays, not modal agreement (see section 3.2.2). For instance, comments are quoted which state that in Kalam "[t]he naming of black is surprisingly inconsistent. The most common black term (S) is only used by eleven informants. Seven other terms are used for black including the white term (T) once, the green term (K) twice and the blue term (M) four times", or that in Tifal, "There are six words for red with none especially more prominent than the others". In perfect compliance with these observations, the aggregate naming array for Kalam on the 52% level of agreement displays no marks of an established category in the black region, and the array for Tifal on the same agreement level displays no marks of an established category in red.

3.2.6 Unsystematic flaws

Only few of the included field comments indicate substantial methodological problems, and it is without dispute that the influence of those should be filtered out from the WCS data in order to arrive at the “reliable core”. The most flagrant case is Karajá, where the field comment notes that “the responses may be more regular than normal as the chips were shown around the room and their responses were taped. Some may have given the same response as the previous one because it was more convenient not to disagree.” That is also reflected in the nonsensically high coverage of the Karajá full agreement aggregate naming array. Such violation of the standard procedure, needless to say, flies in the face of my appeal to levels of interindividual agreement, and the Karajá data thus cannot be of any use for our purposes. This is an illustration of the last component of the reduction of the WCS data to their “reliable core” that I propose. To gain an empirical basis of evaluation that is maximally representative of the phenomenon under scrutiny, we cannot avoid going through reports which indicate what the data provided by the WCS for particular languages really stand for. These data are not sacred, and should more cases like Karajá show up, there is no point in keeping such flawed parts in place.

3.3 Conclusion of chapter 3

In this chapter, I have first argued that the lack of a general concept of color in some languages does not imply that there cannot be any universal patterns of color categorization. Further, I have shown that the methodological problems of the World Color Survey, presented as fatal by the relativist side of the color term debate, are by far not serious enough to render the WCS data on color categorization useless. However, a considered reduction of the overall WCS outcomes is necessary in order to reveal the “reliable core” of the WCS results, which could be reasonably taken for the empirical basis of evaluation of any models of the (idealized) phenomenon of language categorization. I argued that three principles of such a reduction are desirable. First, color categorization systems of the world’s languages can only be represented by the higher agreement aggregate naming arrays resulting from the WCS, as opposed to the modal agreement arrays. Second, data for languages which are heavily influenced by other, culturally dominant languages, in particular languages of the Western colonialism, need to be omitted. Third, available reports on factual genesis of the data for particular languages should be inspected and flawed data resulting from unsystematic failures in the procedure are to be eliminated.

Chapter 4

“Color spaces” and “the perceptual color space”

It seems clear that any serious model of linguistic categorization of color is bound to involve, as a crucial component, some sort of representation of topological relations among colors, based on perceptual identity, similarity and difference, as experienced by normal, that is standard trichromatic, observers.

Normal color perception is (somewhat misleadingly) called “trichromatic” because it is mediated by three types of cones on the retina, each type having its peak of sensitivity at a different wavelength of light. In some observers, genetic mutations produce a different number of cone types and/or shifts of their sensitivity functions. Various known types of “color-blindness” have been explained as more or less serious deviations from standard trichromacy. About 8 % of males and less than 1 % of females are in this sense non-standard observers, whose topology of color perception in various ways differs from that of normal trichromats. (Baylor, 1995; Mollon, 1995; Fairchild, 2005; Jameson and Komarova, 2009b.) Most of this chapter deals with various attempts to capture the color topology of a standard trichromatic observer. Possible impacts of heterogeneity in population on development of linguistic color categories are discussed in section 4.9.

In the literature on color categorization, the term “color space” is often used quite loosely, denoting many different topological representations related in some way to color, which are built on various bases and for manifold objectives. The following overview should clarify which of the multiplicity of “color spaces” are the most relevant for our modeling purposes, as well as what exactly can be expected of them.

4.1 Color order systems

The Munsell color system, first published in 1905, with last revision in 1940, is the basis of the color array most commonly used in color categorization research and discussed in previous chapters. It is an example of a color order system, rather than a *stricto sensu* color space. (Cf. Brill, 1997.) Basically, a color order system is an ordered and labeled collection of physically (colorimetrically) specified color samples. Unlike some other color order systems, the Munsell system is intended to precisely capture certain features of normal human color perception. More specifically, it aims at perceptually uniform spacing between neighboring samples in each of the three assumed dimensions of color perception: lightness (called *value* in the Munsell system), hue, and chroma.[24] In a 3D representation, these correspond to cylindrical, not Cartesian, coordinates. There are 10 degrees of lightness from white to black, 100 steps in the hue circle (with focal red, yellow, green, blue and purple in regular intervals of 20 steps) and a variable number of chroma degrees available for particular combinations of lightness and hue.

The uniform spacing along each dimension is based on extensive psychophysical[25] experimentation. On the other hand, no explicit relation is given for spacing in one and another dimension; therefore, there is no general metric of color similarity across dimensions. The restricted uniformity of spacing is guaranteed for highly specific conditions: for a normal trichromat observing the samples on a uniform middle-gray background, under 2° of visual angle, under a standardized, daylight-like illumination. And one more qualification is in order: contrary to what is usually assumed, the uniform spacing in the hue dimension may arguably hold only *within* each of the five main regions marked out by the red, yellow, green, blue and purple point. My reason for this claim is that it is an *assumption* of the Munsell system, rather than a psychophysical finding, that it is precisely these five hues which regularly divide the hue circle. If the five hues are placed at regular intervals *a priori*, it is most unlikely, although not impossible, that the sizes of the hue steps between each adjacent two of them will come out perceptually equal.

For other perceptually relevant color order systems, such as the Natural Color System and the OSA system, as well as for additional details on the Munsell system, cf. Jameson and D'Andrade (1997) and Fairchild (2005, ch. 5).

By definition, color order systems are discrete. The perceptual character of a sample, reflected in its position in the system (*e.g.*, red-purple 6, lightness 3, chroma 4) is linked to its colorimetric characteristic only via a lookup table. No general mathematical relation between physical attributes of a sample and its perceptual attributes (under the given conditions of observation) is provided.

[24] Munsell chroma is closely related to saturation; see the discussion at 4.8

[25] In general, psychophysical research investigates relations between physical and perceptual characteristics of stimuli.

4.2 The CIE 1931 system

A rudimentary mathematical formulation of the relation between the physical and the perceptual is presented by the CIE (Comission Internationale de l'Éclairage) XYZ system of tristimulus colorimetry of 1931. To understand its function, it is important to appreciate how the standard trichromatic physiology of human color perception works.

The only human receptors for color vision are the three types of cones. That is, the only color information that is available at higher levels of visual processing is given by the levels of their activation. In each of the cone types, which are characterized by different response functions over the range of wavelengths of the visible light, the same level of activation can be achieved by various combinations of wavelength and intensity of light. Hence, if two stimuli differing in spectral distribution of their emitted or reflected light cause the same pattern of activation in the three types of cones, they will appear identical despite their physical difference. This is called *metamerism* or *metameric matching.* Three monochromatic lights ("primaries") that are sufficiently apart from each other, such as the commonly used red, green and blue, can match in appearance any spectral distribution of light when they are mixed in an appropriate proportion. (Baylor, 1995; Mollon, 1995; Fairchild, 2005.)

Given the colorimetric characterization of a stimulus, the matching functions of the CIE 1931 system provide the intensity of each of three oversaturated primaries XYZ that is needed to match the stimulus in appearance. In this way, each tristimulus value XYZ sets apart the stimuli of corresponding color appearance from all other stimuli, thus giving an account of metameric matching. This amounts to a most rudimentary general perceptual characterization of physically defined stimuli: for any two physically different stimuli, the system allows to decide, via identity or non-identity of their XYZ values, whether they match or not. Note that it is much less than to determine *how* the stimuli appear, or at least which one is lighter, more saturated or closer to green. Unlike the ordering of samples in the Munsell system, the topology of the XYZ space has a very limited perceptual relevance. What is shared by both these systems, though, are the narrowly defined observing conditions under which the predicted perceptual relations hold. (Fairchild, 2005, ch. 3.)

4.3 Perceptual color spaces *stricto sensu*

Perceptual color spaces in the narrow sense are constructs which combine the strong sides both of perceptually relevant color order systems and of the CIE 1931 system. They provide more informative characterization of perceptual properties and relations of physically specified stimuli, like color order systems do, but in contrast to them they provide it not only for finite collections of samples, but for arbitrary stimuli on a functional basis.

The most widely used perceptual color space nowadays is the CIELAB (or CIE L*a*b*) color space from 1976. For a tristimulus XYZ characterization of a sample, given by the CIE 1931 system, plus the XYZ value of the referential white, the defining equations of CIELAB (see Fairchild, 2005, ch. 10) give the position of the color in the space of three Cartesian dimensions L*, a* and b*. The cylindrical coordinates of this 3D space correspond to lightness, hue and chroma, with lightness represented by L*. CIELAB, like other perceptual color spaces, is intended to go beyond the partially uniform perceptual spacing of the Munsell system, following the ideal of full perceptual uniformity. The Euclidean distance of any two color points in the space (not just along each dimension separately) should be proportional to the perceived color difference between them. (Fairchild, 2005, ch. 3; 10.)

Thus, CIELAB can be seen as a direct attempt to represent the full, ideal topology of a standard observer's color perception determined by the relations of identity and relative similarity or difference of perceived colors. It is this topology that is occasionally referred to as "the perceptual color space" or "the internal color space" (*e.g.*, Hardin, 1997). One should keep in mind that rather than some kind of independent psychological entity, it is a theoretical abstraction built on the basis of perceptual relations among colors. (Cf. Jameson and D'Andrade, 1997; Saunders and van Brakel, 1997a; Dedrick, 1998.) For construction of perceptual color spaces such as CIELAB, judgments of perceptual relations are gained by means of psychophysical experiments. In these, the role of language is severely restricted, such as to judgments of perceptual identity and difference, as in matching and threshold experiments. (Cf. Fairchild, 2005, ch. 2.) Thus it seems relatively safe to assume that perceptual color spaces approximate, however imperfectly, a non-linguistic and universal reality of low-level (see section 4.4) color perception or sensation (as far as normal trichromats are concerned, and detaching from minor interindividual variation, which can be linked to minor physiological differences). That qualifies them, in principle, as plausible components for models that aim to explain existing patterns of linguistic categorization of color. This is not the case with many other color spaces (in the broad sense) which are, in their topology, to various degrees dependent on the established color categorization of particular languages. An explanation based on such color spaces would be in danger of circularity.

The other official CIE perceptual color space adopted in 1976 is CIELUV (or CIE L*u*v*). Since then it has proved less perceptually adequate than CIELAB and "[a]t this time there appears to be no reason to use CIELUV over CIELAB" (Fairchild, 2005, p. 80; p. 195).

In perceptual color spaces, one essential limitation is inherited from color order systems and the CIE 1931 system. They are constructed to reflect appearance of stimuli under a fixed set of specific, "standard" viewing conditions. As in the previous constructs, a direct (foveal) gaze of a normal trichromatic observer is assumed. For non-standard observers as well as for perception in visual periphery, the captured perceptual relations do not hold. (Fairchild, 2005, p. 34.) Furthermore, the instructions for use of the CIE color spaces, quoted in Fairchild (2005, p. 194), state: "These spaces are intended to apply to comparisons of differences between object colours of the same size and shape, viewed in identical white to middle-grey surroundings, by an observer photopically adapted to a field of chromaticity not too different from that of average daylight." For some purposes, particularly construction of imaging devices, such restricted specification of color appearances is not sufficient.

4.4 Digression: Rich perception aside

The notion of "low-level" color perception or sensation, which is quite crucial for this and the following chapter, is meant as opposed to "rich", cross-modal and more cognitively involved color perception, uncovered notably in Merleau-Ponty's *Phenomenology of perception* (2005). A particular color may be perceived not just as green of a middle lightness and a rather high saturation, but also as the green of a long gone woolly carpet in one's grandparents' house, or the green of a juicy Granny Smith apple. A related voice is Saunders (2006) and Saunders (2007), with her notion of "seeing color" as a socio-historical institution and a cultural skill, rather than a biological given. (Cf. also Costall, 1997; Dubois, 1997.) I believe that non-reductive phenomenological and socio-historical reflection is of great importance for general understanding of color. Nevertheless, the low-level physiology of color perception has been described fairly well and one can hardly deny that, *e.g.*, mediation by the three types of cones is a *sine qua non* for even the most phenomenologically complex color experiences. For some purposes, as that of ours, considering color perception restricted to the sensational, pre-cognitive level seems to make a good sense; contra Saunders (2007, p. 475), who pleads for leaving the "sensory core" behind in color categorization research. In case of "rich" (unlike the "low-level") color perception, the contemporary color science does not offer models that could be readily used in modeling of linguistic categorization. For this practical reason, more cognitively involved phenomena of color perception are set aside in the present thesis.[26]

[26] I believe Saunders (2000; 2006; 2007) rightly points out the danger of replacing richer notions of color and overwriting the life world (in Husserl's sense) of color experience with scientific accounts of the low-level color perception. It is nonetheless hard to assess the power of contemporary color vision science to cause, via color technology, general changes in human color perception, as suggested by Saunders (2006; 2007), Saunders and van Brakel (1997a), Brill (1997), Simpson

4.5 Color appearance models

Color appearance models, such as the actual official CIE model CIECAM02, present the most recent type of technical models of human color perception. They in various ways generalize beyond the standard viewing conditions assumed in perceptual color spaces. They aim to account for some of the many effects of color perception by which, under specific viewing conditions, color appearances of individual stimuli are shifted in unexpected ways. These include, *e.g.*, effects of simultaneous contrast, of chromatic adaptation, of non-standard illumination, of spatial structure of the visual input, which can be demonstrated by various kinds of optical illusions.[27] Thus, these models need substantially more input than merely the XYZ characteristic of the stimulus. They also work with colorimetric specifications of the stimulus background and its broader surround, as well as with specification of the illumination and of the mode of viewing (such as observing an illuminant, or an object, or a projected figure). Other spatial and temporal characteristics of the visual field may be included as parameters. (Fairchild, 2005, ch. 6 to 16.)

While predicting color appearance over a variety of viewing conditions is necessary in many technical applications, in color categorization modeling we can arguably do without it, and therefore also without general color appearance models. Considering only perceptual relations under specified conditions of observation, as captured in perceptual color spaces, is a simplification, but a justified one. Rather than being a mere marginal subset of all possible viewing conditions, those conditions seem reasonably standard and representative of everyday cognitive dealing with colored objects. The perceptual relations with respect to foveal perception of similar-sized objects on a neutral background by daylight are surely a more plausible explanatory basis for linguistic categorization than relations corresponding to any extreme deviation from these conditions. Minor deviations are not likely to cause serious distortion in perceived color differences. On the other hand, radical

(1997). This is another sense in which Saunders accuses the universalist line of color categorization research of circularity (cf. section 3.2.5). According to her (2007, p. 475), experimental research, including the WCS, only reveals structures that have been imposed on perception by the global power of color science and technology. To me, this seems like a heavily exaggerated claim.

[27]Apart from low-level perception effects and related to the "rich color perception" discussed in the previous section, unconstrained color appearance is influenced by cognitive effects which make it dependent on knowledge concerning the observed object and the illumination conditions. Known objects tend to retain their appearance to some extent across viewing conditions and unusual illumination is cognitively discounted if noticed. Thus, a sheet of paper and a banana tend to appear, respectively, white and yellow even under a dim or violet illumination: an example of "color constancy" of objects. These cognitive effects are particularly hard to account for formally and are generally not covered by color appearance models. (Cf. Mollon, 1995; Fairchild, 2005, ch. 6.)

deviations can change appearances drastically, but are rare in practice, especially in pre-technological societies, and it is hard to imagine that the should have any consistent impact on color categorization.[28]

I therefore conclude that the innovations brought by color appearance models are rather dispensable for our purposes. In the following, I will propose modeling of color categorization on the basis of the CIELAB color space.

4.6 Other "color spaces"

The color models described above appear to be, for the present purposes, the most relevant of all available color spaces (in the broad sense). For the sake of completeness, we can mention other "color spaces", which are arguably less suitable as an explanatory basis in color categorization research, although they are also used in some of the literature to represent color relations. Such are, for instance, the "cone-opponent" space of colors derived from physiological findings on early stages of color processing (*e.g.*, Wuerger and Parkes, 2011), or "hue circles" of various, sometimes fairly traditional origin.[29] Jameson and Komarova (2009a) and Jameson and Komarova (2009b) build their categorization models upon the basis of a hue circle that is derived from the Munsell color system. The representation of the hue topology by this circle might not be completely adequate, due to imperfect uniformity of spacing in the Munsell hue dimension (see section 4.1). Another topological representation of color is the familiar one-dimensional space locating the perceived hue relatively to the wavelength of the observed monochromatic light. The categorization models of Baronchelli et al. (2010) and Loreto et al. (2012) are built on the basis of this physically defined color space. That again seems to be an inferior choice compared to more perceptually relevant color topologies.

4.7 The perceptual basis for explanation

From what was said so far, the CIELAB space comes out as a plausible perceptual basis for a model of color categorization. One practical choice is to project the 330 Munsell color chips into CIELAB and consider specifically their partitioning based on the perceptual distances. Color categories in languages tend towards compactness, or optimality in minimizing within-categorical and maximizing inter-categorical distances; cf. Regier et al. (2007); Jäger (2009). Partitioning of the

[28]Consider, for instance, perception of two objects on a crimson vs. an orange background, in a dim blue light, with the observer adapted to a strong red light.

[29]An exhaustive art-historical overview of the past opinions regarding color relations is offered by Gage (1999).

Munsell chips in CIELAB is implemented in Regier et al. (2007) and Correia and Ocelák (2013).[30] Nevertheless, I want to argue that this is not an adequate treatment of the color categorization problem.

Jameson and D'Andrade's (1997) idea of explaining tendencies in color categorization in terms of optimal division of the irregularly shaped color space is increasingly popular (cf. Regier et al., 2007, Correia and Ocelák, 2013), yet it seems wrong in an important sense. Sure, it may be the case that in the (ideal) perceptual color space, some salient colors are located in a way that is less regular than expected. But the color space as such *cannot* be irregular, and that is by definition of the space as perceptually homogeneous, uniformly spaced for a standard observer. What *is* irregular, or "bumped", is the figure that results when the 330 chips of the Munsell array are projected in the space – and that is also what is being partitioned in the models by Regier et al. (2007) and Correia and Ocelák (2013). But there is *no good reason* to consider partitions of an irregular set of points in the color space: the more irregular such a set is, the less it can be regarded as a representative substitute for the whole space, and the more irrelevant its optimal partitions are for the problem of color categorization. There *are* good reasons to consider optimal partitions of the color space as such – but these will not lead to a small set of optimal solutions but to indefinitely many, as no irregularity will prevent arbitrary rotations of an optimal partitioning solution from being optimal as well. (See also section 4.9.) Had the 330 chips of the Munsell array been chosen so as to form a perfect, regular sphere in the perceptual color space, that would have qualified optimal partitions of the figure as relevant for the color categorization issue – but at the same time, it would have led to a great many of them being equally optimal.

The conclusion is clear, even if surprising: partitioning optimality cannot be sufficient to explain cross-linguistic tendencies of color categorization. We should consider partitions of the color space (practically, CIELAB) as such, not of the irregular Munsell figure in it; but additional principles will be needed to suppress arbitrary rotation of the optimal partitioning solutions within the color space. Possible candidates for this role will be discussed in section 4.9 and chapters 6 and 7. The 330 Munsell chips should not be employed in the model itself, but only in evaluation: they are the only color points for which reliable empirical cross-linguistic data regarding categorization are available (cf. chapter 3). A model providing independent partitions of the color space should be evaluated against these categorization data.

[30] The question of a reasonable, realistic mechanism of such partitioning is not addressed in this thesis, as it goes beyond the perceptual basis considered here. Basically, I assume a solution in terms of agent-based evolutionary game-theoretic interaction, as in most of the recent color categorization models; cf. Steels and Belpaeme (2005); Jäger and van Rooij (2007); Baronchelli et al. (2010); Loreto et al. (2012); Correia and Ocelák (2013).

One more qualification is in order. In fact, CIELAB, since its introduction, has turned out to lag behind its ambition of perceptual uniformity. (Fairchild, 2005, p. 189f.) While the original color difference formula simply measured the Euclidean distance between two color points, the CIE has adopted two revised formulas since, namely the color difference equation of 1994 and the yet more involved formula CIEDE2000. (Cf. Sharma et al., 2005.) Similarly to the original formula, these more recent formulas also work upon CIELAB. Fairchild (2005, p. 82) and Kuehni (2002) express doubts about whether the improvement in adequacy brought by the latter to most practical applications compensate for the added complexity. But CIEDE2000 adjusts, among others, CIELAB's inaccurate hue spacing in the blue and green region (Luo et al., 2002, MacEvoy at *handprint.com*). This might prove important for the present purposes, since Correia and Ocelák (2013) had difficulties accounting with the Euclidean CIELAB metric for the strong cross-linguistic tendency to categorically merge green and blue.[31]

It is hard to estimate in advance how strong impact these improved color difference metrics can have on the artificial partitioning of the color space, especially if the latter results from a game-theoretic interaction. For computational simulations, there will certainly be a practical trade-off between perceptual adequacy and complexity of the solutions ordered from the simple CIELAB Euclidean metric to CIEDE2000.

4.8 Are there indeed three dimensions to color?

Until now, I have taken for granted the more or less mainstream notion of color as characterized by exactly three distinct perceptual attributes, which people find intuitively plausible and are able to distinguish with a little training: hue, lightness and saturation. There is, however, not a unanimous consensus on such a notion within color vision science and color categorization research.

Fairchild (2005, ch. 4) distinguishes 6 "perceptual" or "appearance" attributes of color: hue, brightness, lightness, colorfulness, chroma and saturation. He further claims (p. 91; p. 145) that the first five are necessary for a full specification of a perceived color (whereas hue, lightness and chroma are sufficient for most practical applications dealing with related object colors). Brightness is to lightness what colorfulness is to chroma. The first attribute of either pair is supposed to capture an "absolute" perceptual property, which generally grows with increasing luminance. The second should capture a "relative" one, measuring the respective "absolute" attribute by brightness of a white area observed under the same illumi-

[31] A different option would be to compute color differences based on CIECAM02 (cf. Xue, 2009), where spacing in green and blue is also condensed in comparison with CIELAB (MacEvoy at *handprint.com*).

nation, and so remaining roughly invariant across luminance changes. While these may be intuitively appealing distinctions, it seems in order to ask what the status of such attributes is; whether they are phenomenologically obvious, or rather advanced theoretical constructs (embodying also the idea of color constancy of objects; see footnote 27).

Sokolov (1997) claims the perceptual color space to be four-dimensional. Saunders and van Brakel (1997a) refer to various evidence against the 3D space of—cylindrically—hue, brightness and saturation, they point to interdependence of the latter two attributes, and they quote a suggestion that the perceptual color space should be conceived as 6- to 8-dimensional. Kuehni (2002) casts doubts on the very ideal that differences in various perceptual attributes should add up to overall color difference in a Euclidean fashion.

Saunders and van Brakel (1997a, p. 175) also include a list of color-related attributes that are, in their opinion, unjustly eliminated from the standard three-dimensional scheme: "such features as duration, size, texture, glossiness, lustre, fluctuation, flicker, sparkle, glitter, shape, insistence, pronouncedness, brilliance, fluorescence, glow, iridescence, colourfulness, nuance, background or surround colour". That is, however, a rather odd collection of attributes, if they are meant as candidates for additional dimensions of the perceptual color space. It includes conditions of observation, which are considered as parameters in color appearance models and are hardly attributes of color in themselves (duration, background and surround color, flicker, size, shape...); groups of terms that are more or less synonymous (sparkle and glitter; luster, brilliance, glow and fluorescence); and attributes that can arguably be reduced, as far as visual perception is concerned, to spatial patterns of color (texture, glossiness).

While hue, lightness and saturation may not suffice to characterize the color of an object exhaustively, one can hardly doubt that they do it much better than duration, size and flicker, or than luster, glitter and brilliance. Those who defend the familiar notion of a three-dimensional color space do not claim that it perfectly fits the perceptual reality, but appeal to its practical utility and reasonable correspondence to psychophysical findings. According to them, multidimensional scaling of color similarity judgments supports a 3D space of hue, brightness and saturation, with approximately Euclidean perceptual relations, at least locally. (Hardin, 1997; Jameson, 1997; Mausfeld, 1997.)

The issue of perceptual attributes or dimensions of color is both empirically and conceptually intricate. Without contributing more arguments or taking sides, I conclude that there does not appear to be a unanimous opposition to general perceptual relevance of 3D representations of color in terms of lightness/brightness, hue and chroma/saturation (as cylindrical coordinates) that would prevent us from relying on CIELAB as the perceptual basis for explanation of color categorization.[32]

4.9 The role of color vision deficiencies

So far, I have only considered the character of color perception by standard trichromatic observers. However, every population has a proportion of color-deficient speakers of various types. The simulation results in Jameson and Komarova (2009a) and Jameson and Komarova (2009b) indicate that a realistic proportion of the more common color deficiencies (types of dichromacy: protanopy, deuteranopy, and types of anomal trichromacy: protanomaly, deuteranomaly[33]) in the population is able to bias the dynamic of agent-based interaction in favor of some categorical solutions and against others. In other words, the appropriate perceptual basis for color categorization modeling need not be exhausted by the perceptual makeup of a standard observer; instead, the perceptual heterogeneity of the population is likely to play a role.

For color-deficient observers, the similarity-based topology of color differs from the standard. To my knowledge, presently there are no artificial color spaces attempting to capture the character of color perception by these non-standard observers, which would of course be desirable for our modeling purposes. Jameson and Komarova (2009a) and Jameson and Komarova (2009b) represent each of the four non-standard topologies in a simplified way, by means of a Munsell-based hue circle, where for each two points of the circle there is certain empirically adjusted probability that the observer confuses these points. As linguistic color categories are primarily differentiated in the dimension of hue, rather than lightness and saturation, this seems to be a reasonable simplification of the problem which can be readily used in further modeling experiments.

Jameson and Komarova (2009b) report an important result of their experiments with perceptually heterogeneous populations: a realistic proportion of color-deficient agents cancels the effect of rotational arbitrarity that is characteristic of a homogeneous population of standard observers; that is, the fact that the interaction in the homogeneous population leads to categorical solutions which share a basic shape but are arbitrarily rotated over the hue circle.[34] Dichromatic and

[32] I do not think that such a worry is out of question by the mere fact of using one of the more recent, non-Euclidean metrics of color difference upon CIELAB, as these metrics are only supposed to adjust the perceptual relations captured in CIELAB by way of reweighting dimensions etc.

[33] The share of each of the first three deficiencies in men of European Caucasian descent is between 1 and 2 %, in women it is two orders of magnitude lower. For deuteranomaly, it is about 5 % in men and about 10 times less in women. (Fairchild, 2005; Jameson and Komarova, 2009b.) These numbers can differ for other populations; *e.g.*, in men of Chinese and Japanese descent the overall share of these four deficiencies is between 4 and 6.5 % (Birch, 2012).

[34] Rotational arbitrarity is not a mysterious specific of (simulated) populations of standard

anomal trichromatic perception is characterized as involving regions of the hue circle within which the hue points are more likely to be confused. At the population level, a proportion of deficient agents leads to the effect that categorical boundaries are repelled by these regions. "[E]volved systems tend to minimize the likelihood that colors perceptually confusable by *some individuals* in a population will tend to be classified by the entire population into different color categories" (Jameson and Komarova, 2009b, p. 1432). Admittedly, this is a result of artifical categorization, which also depends on other non-trivial assumptions concerning the form of interaction between agents. Still, it suggests a principle which might illuminate why there is more cross-linguistic regularity in categorization than can be explained simply by optimal partitioning of the color space. (Cf. section 4.7.) The idea is promising because even if the proportion of each color vision deficency somewhat varies in different populations, it is always the proportion of the same *type* of observer, and therefore the same kind of impulse for weakening the rotational arbitrarity effect. Other conceivable explanations for lack of rotational arbitrarity in empirical categorical systems will be discussed in chapters 6 and 7.

Another significant result of Jameson and Komarova (2009b) is that once a realistic proportion of *protanopic* and *deuteranopic* observers is added to the population of standard trichromatic agents, further addition of a realistic proportion of *protanomalous* and *deuteranomalous* agents does not have a qualitative impact on the categorical solutions preferred by the population, apart from strengthening the tendencies triggered by the protanopic and deuteranopic agents.

4.10 Conclusion of chapter 4

In this chapter I have examined the issue of a suitable basis for color categorization modeling which would realistically reflect human low-level color perception. I have outlined several types of topological representation of color that are usually employed in the color categorization literature, and I have discussed their relevance for the present modeling purposes. On that basis, I have suggested the following modeling strategy. Game-theoretically induced partitions of the CIELAB color space (as opposed to partitions of the Munsell array projected into CIELAB) should be considered, based on the color differences determined by the CIEDE2000 color difference formula. It is, however, desirable to include a realistic proportion of color-deficient agents in the simulated populations, as in Jameson and Komarova (2009b); for this is one of the possible mechanisms to suppress rotational arbitrarity and generate particular categorization tendencies. While the 330 color points

trichromatic agents, but rather a trivial consequence of the fact that the hue circle is a topology of hue where the spacing is uniform precisely with respect to this type of observer. Therefore, rotational arbitrarity of categorization by standard trichromats is to be expected also when divisions of the full color space (not just the hue circle) are considered.

of the Munsell array have no role to play in the model as such, they should be used for evaluation of its ability to produce realistic partitions of the color space, as they are the only color points for which reliable data are available, as far as actual color categorization in the world's languages is concerned.

Chapter 5

"Unique hues"

5.1 History and position of the concept

It is a common notion in the contemporary color science that there are certain hues, labeled "unique hues", that are perceptually unlike all other hues. Namely, these hues are phenomenally pure or unmixed and all other chromatic hues can be described in their terms since they are perceptually composed of two of them in a specific proportion. It is typically claimed that there are 4 such unique hues: pure red, yellow, green, and blue.[35]

The modern idea of 4 unique hues in perception dates back to the 19th century Prague and Leipzig physiologist Ewald Hering. Hering (1878, 1964 in English) coined his theory of color perception in terms of red-green and yellow-blue opponent mechanisms, in disagreement with the contemporary trichromatic theory by Young and Helmholtz (cf. Baylor, 1995). His ideas were revived in the influential opponent-process theory of color developed in the 1950s by the psychophysicists L. M. Hurvich and D. Jameson (1957). The opponent-process theory describes perceived color as a combined output of three psychophysical channels organized in an opponent fashion: red-green, yellow-blue, and an achromatic black-white channel. These three channels had been identified on the basis of the assumed phenomenal uniqueness of the respective colors, which was also confirmed by the subjects' consistent performance in experiments. Hurvich and Jameson's psychophysical findings then gained strong support from physiologists who described certain patterns of opponent color-coding (that is, activation of a neural channel by one color and inhibition of the same channel by another) in post-receptoral neurophysiological processing. (Valois et al., 1966; Valois and Valois, 1975.) The

[35] Occasionally, only the intension of the term "unique hue" is acknowledged, and the extension examined, as in Logvinenko (2012).

physiological primacy of four (or 6, including also white and black[36]) unique hues was subsequently adopted as a powerful explanatory principle in color categorization research (Kay and McDaniel's Fundamental Neural Response categories, 1978, see 2.1 and 2.2; Hardin, 1988).

However, it turned out that the response patterns observed in neurophysiological channels of color processing do not fit to the three primary perceptual oppositions that had been examined by psychophysicists. Despite continual efforts in neurophysiology, there is a wide consensus in the present-day color science that we are still lacking a plausible physiological explanation for existence of four perceptually unique hues interrelated in a double-opponent way. (Cf. Krauskopf et al., 1986; Mollon, 1995; Jameson and D'Andrade, 1997; Saunders and van Brakel, 1997a; Dedrick, 1998; Mollon, 2009; Jameson, 2010; Broackes, 2011; Wuerger and Parkes, 2011)

The rhetorical situation is remarkable. The past deconstruction of the proposed link between the physiological and the perceptual has led almost everyone in the field to the following conclusion: further neurophysiological research is needed in order to account for the well-known fact of there being four unique hues in human low-level color perception, namely (the focal or pure) red, yellow, green and blue (hereafter, RYGB). This basic, intuitively quite plausible description of the perceptual phenomena, surviving without a change from Hering's time, is taken for granted. That places the burden of connecting the perceptual and the physiological fully on neurophysiology. The perceptual primacy of RYGB is a common unquestioned assumption in neurophysiological studies on color as well as in color categorization literature. (Cf. Mollon, 1995; van Laar, 1997; Jameson, 1997; Dedrick, 1998; Stoughton and Conway, 2008; Mollon, 2009; Panorgias et al., 2010; Broackes, 2011; Wuerger and Parkes, 2011.) Broackes (2011) calls the four perceptually unique hues "striking phenomena to be explained". Valberg (2001) titles them "an old problem for a new generation". According to Mollon, a leading neurophysiologist, "the special phenomenal status of the four pure hues is perhaps the chief unsolved mystery of colour science" (Mollon, 1995, p. 146; cf. also Mollon, 2009).

In this chapter, I will argue that rather than a central problem that neurophysiology should solve, the perceptual uniqueness of red, yellow, green and blue is a central chimera of the contemporary color science. Neurophysiologists have not managed to find specific mechanisms for the so-called unique hues, and they will not, because there is nothing to be found; there are no perceptually unique hues. My arguments do not in the least depend on any kind of physiological evidence. I aim my criticism solely at the received wisdom concerning the phenomenal, by

[36] Since Hering and throughout the discussion, black and white have kept an ambivalent status, usually mentioned in connection with the assumed achromatic perceptual channel but omitted from the list of unique hues.

which neurophysiological research has been commonly driven. I will try to demonstrate that the notion of RYGB as perceptually unique hues lives on folk intuitions and on conceptual confusion caused by the misleading powers of language, and that it falls apart under scrutiny. One after another, I will present and discard the reasons which have been suggested for granting the privileged status in color perception to red, yellow, green and blue. My criticism is not completely new in any of the points (in particular, cf. Saunders and van Brakel, 1997a; Jameson and D'Andrade, 1997; Jameson, 2010), but to my knowledge, as yet nobody has formulated a comprehensive critique of the unique hue concept which it deserves, given its position in current neurophysiology of color. Of course, rejecting all proposed reasons for thinking that there are perceptually unique hues does not quite imply non-existence of unique hues; but if my point is correct, then the burden of proof is thrown upon the opponents.

To prevent misunderstandings, I would like to precede the discussion with an important qualification: I do not deny that red, green, yellow and blue are cognitively privileged colors for speakers of English, of Western languages, and perhaps of most languages in the world. But even if the last were true, one could still imagine multiple explanations. What I reject is that the primacy be conceived as an established universal fact of prelinguistic, low-level[37] color perception, one that could be sensibly accounted for in neurophysiological terms. My criticism of the unique hue concept should be read with this in mind.

Insofar as this chapter challenges a central assumption of the current neurophysiological research on color, it could be seen as a digression from the flow of the present thesis. However, deconstruction of red, yellow, green and blue *qua* perceptually eminent hues is relevant to its focal issue, linguistic categorization of color, in at least three ways. First, unique hues (together with white and black called "the Hering primaries") have been appealed to in explanation of color categorization in languages. (Cf. Jameson, 2010.) Admittedly, the Hering colors seem to have been given up as *explanans* in the B&K research line once the alleged correspondence between perceptual and physiological channels proved untenable. Still, one could, with less reductionist ambitions, take perceptual uniqueness for a plausible explanatory principle, regardless of physiology: were there any such uniqueness. Second, the opponent organization of the assumed four unique hues is sometimes placed as a constraint on construction of perceptually relevant color spaces (in the broad sense), with the demand that the red-green and the yellow-blue opposition coincide with two axes of the space. That is, for instance, the case in the Swedish NCS color order system. Once the notion of opponent unique hues is challenged, so should be perceptual relevance of color topologies which rest on this theoretical preconception, rather than merely on psychophysical results

[37]See section 4.4.

regarding color difference. Third, the significance of which will become patent in the subsequent chapter: the assumption of four unique, prelinguistically salient hues is likely to have had a serious impact on the design of experiments in existing research of "categorical perception" of color and on the interpretation of results.

5.2 Criticism

5.2.1 Unique hues as referents of necessary and sufficient descriptors

One of the two major motivations usually given for describing red, yellow, green and blue as perceptually unique is straightforwardly linguistic. Unique hues are claimed to be those corresponding to the set of terms that are, individually or in combinations of two, necessary and sufficient for the description of any point of the hue space (circle). A slightly different definition to the same effect (namely, to selecting the four familiar hues) is that a hue is unique if and only if it cannot be described by names other than its own; this is how unique hues are officially defined by CIE (1987). In both these closely related senses, the uniqueness of RYGB has been repeatedly corroborated by experiments on forced naming of colors, beginning with Sternheim and Boynton (1966). This "linguistic" notion of a unique hue (in either variant) is employed by, for instance, Miller (1999); Werner and Bieber (1997); Hardin (2005); Stoughton and Conway (2008); Panorgias et al. (2010); Broackes (2011).

It is rather obvious that according to this definition, unique hues are language-relative. (Cf. criticism in Saunders and van Brakel, 1997a, and Jameson, 2010.) Usual concise introductory claims about universality of unique hues in this sense, with mechanical reference to Berlin and Kay (1969), are simply false. A great deal of the world's languages do not have a set of color terms that would correspond in meaning to the English "red", "yellow", "green" and "blue". That much can be inferred even from Berlin and Kay (1969), not to speak of more recent cross-linguistic evidence (Kay et al., 2009). Some languages arguably have no established color terms at all. (Cf. chapter 3, in particular sections 3.2.2 and 3.2.5.) It is clear that in none of those languages the four familiar hues can come out as referents of necessary and sufficient descriptors, or as exactly the set of hues that cannot be described by other names then their own.

So, the privileged status of the RYGB hues in relation to linguistic descriptors is a fact of particular languages, and does not lead to any conclusions regarding their primacy in prelinguistic, low-level color perception.

5.2.2 Unique hues as phenomenologically pure color experiences

Another major motivation for setting RYGB apart from all other hues, sometimes proposed explicitly in contrast to the CIE linguistic definition, is their alleged phenomenologically unique character. The focal red, yellow, green and blue are claimed to be "pure" in a sense in which other hues are not. A whole variety of expressions is employed in the literature to express this phenomenological observation. Unique hues, like red, are perceptually *simple*, *unitary*, *unmixed*, do not *contain* any other hue; unique yellow *appears neither* red *nor* green; unique red does not *seem/look* in any way yellow*ish* or blu*ish*. Other hues, such as orange, have unique hues *in* them as *constituents*, they *consist* or are *composed* of them, are perceived as *mixtures* or *blends* of the unique hues (possibly in a specific *percentage*, such as 72 % of red and 28 % of yellow); they also *share* hue qualities with other non-unique hues (*e.g.*, orange and lime share a yellow component). (Cf. Mollon, 1995; Jameson and D'Andrade, 1997; Byrne and Hilbert, 1997; Dedrick, 1998; Hardin, 2005; Bornstein, 2006; Mollon, 2009; Wuerger and Parkes, 2011; Panorgias et al., 2010; Broackes, 2011; Xiao et al., 2011; Logvinenko, 2012.)

The core of my criticism of this uniqueness notion consists in noting that the assumed independence of these phenomenological observations of a particular language is illusory. Sure, there is no red or green in focal yellow, while there is some yellow and some red in focal orange; and focal orange is reddish and yellowish and can be called reddish yellow, while focal yellow cannot be called "limeish" and "orangeish" and cannot be labeled "orangeish lime". These statements are true, but they are true *both* because of what human color perception is like *and* because *this is how "yellow", "red", "orange" and "lime" are properly used in English.*

The actual interdependence of the "linguistic" and the "phenomenological" notion of a unique hue is nicely illustrated by frequent cases of unintended blending of both:

- "What is the minimal set of component hues sufficient to specify colour as perceived by normal trichromats?" (Logvinenko, 2012.)
- "Experiments have shown that terms such as orange and purple are not necessary, but can be reduced to yellowish-reds and reddish-blues, respectively, whereas red, green, yellow, and blue cannot be reduced to any other hues." (Werner and Bieber, 1997.)
- "All colors can be described in terms of four non-reducible 'unique' hues: red, green, yellow, and blue". (Stoughton and Conway, 2008.)

One cannot describe a color by means of other colors; one has to use *words* for that. Provided that the phenomenological description is presented as strictly language-independent, sloppy use such as the quoted is highly suspicious. Broackes has the following: "There are certain features that make a particular hue [...] count as [...] *unique* yellow: *(a)* its *looking maximally unmixed* [...] and *(b)* its forming one of a collection of hues (...) that can be said to be 'in' other hues and which together are sufficient for characterizing all hues whatever." (Broackes, 2011, p. 617, original italics.) Here we have "being in other hues" on the linguistic side, or vaguely in between, rather then on the phenomenological side. But what a surprise; the hues that can be properly said to be in other hues, themselves containing no other hues, are identical to the hues the names of which are necessary and sufficient to describe any other hue.

The overall story seems to be as follows: There are four unique hues in low-level color perception, completely independent of acquired language skills. By a weird chance (or is it cultural superiority?), English and other Western languages, unlike many (possibly most[38]) languages of the world, have four privileged color terms that focally refer exactly to these perceptually unique hues. I do not deny that there *might* be such a coincidence. But it is hard to suppress the feeling that the proponents of unique hues are getting it way too cheaply; especially if authors do not even agree whether red's being "in" orange is evidence for the perceptual primacy of red, or rather for the linguistic primacy of the term "red". It seems much more likely that the coincidence is not established via independent examination of the perceptual and the linguistic, but achieved by a straightforward projection of the latter to the former. Hardin (2005) unwittingly provides a telling example: "Names for the Hering elementary colors are necessary and sufficient for naming all of the colors, a fact that justifies singling them out as perceptually elementary."

A large part of the vocabulary employed to express the alleged perceptual uniqueness of RYGB primarily relates to physical composition of objects and substances: *unmixed, contain, in, mixture, consist*, etc. If one is engaged in mixing pigments, these may be taken literally: orange will often consist of red and yellow, and it is well known that one can get green as a mixture of blue and yellow. But in describing perceptual experience, there is no point in interpreting them literally, as referring to physical or psychological componency which one should further trace, maybe in neurophysiological terms. Orange can be said to perceptually contain red and yellow, yet *literally* there is as little of red in orange as there is of beauty in the eye of the beholder.

The perceptual reality as far as red, yellow and orange are concerned is not that orange in any literal manner consists of the former two, but that orange lies between red and yellow in terms of relations of perceptual similarity and difference, as captured in perceptual color spaces. The RYGB terms in English

[38]According to Kay and Maffi, 1999, of the 110 languages examined in the World Color Survey, 23 have distinct terms for red, yellow, green an blue, as opposed to 41 that merge (what we call) green and blue in a single category.

have a privileged position in referring to regions of the perceptual color space; this position also involves the appropriateness of expressing a certain subset of color relations in this space in terms of “containing blue”, “being mixture of yellow and red” or “appearing reddish”. That is a fact of English, from which nothing can be concluded about the character of prelinguistic, low-level color perception. It is therefore a fallacy and a mere reification of idiosyncratic language patterns when color perception is described as discontinuous with reference to unique hues, as in Conway and Stoughton (2009, p. R442): “the familiar color circle, composed of a continuous series of colors, is perceived as discontinuous, punctuated by four unique hues – red, green, blue, and yellow.”

The case of unique hues is not strengthened by the fact that unique yellow, green etc. have been repeatedly located in the spectrum of monochromatic light by means of precise psychophysical experiments. Indeed, subjects in general have little problems with adjusting, *e.g.*, a color in the yellow region of the spectrum when instructed that the result should be pure yellow without any hint of red or green. Moreover, they place their subjectively pure colors in the physical spectrum with considerable interpersonal agreement (for qualifications cf. Saunders and van Brakel, 1997a; Jameson, 2010; Broackes, 2011). However, in all these psychophysical experiments it is an assumption, embodied in the instruction to the subjects, that there *are* unique hues, namely one in yellow, one in green etc. (Saunders and van Brakel, 1997a; Jameson, 2010; Wuerger and Parkes, 2011.) Sufficient interpersonal agreement in performance is then supposed to reveal their precise location. But whatever interpersonal agreement of normal trichromatic observers there is on locating “pure yellow, without a hint of red or green” at a specific wavelength, it need not reflect primacy of the respective hue in the low-level color perception. The only appropriate conclusion is that it reflects the observers’ command of English, in particular of how “yellow”, “red” and “green” are properly used.

Let me finally comment on the position of two authors who have defended the “phenomenological” notion of a unique hue explicitly against the objection of relativity to a particular language.

Broackes (1997), first, finds it intuitively very hard to think in terms of a different set of unique hues than RYGB and to see, for instance, an orange component in red instead of the other way round. In my opinion, intuitions of a Western speaker are simply irrelevant here. In Western societies, RYGB are highly cognitively salient colors which can be for many reasons considered “basic”. They are, on the whole, the first chromatic color categories to be learned, and other colors are often presented to children in terms of the previous four. (For the query “toy blocks”, Google returns images where red, yellow, green and blue blocks heavily prevail over blocks of all other chromatic colors.) Such a strong cultural salience of RYGB makes any intuitions in favor of their *perceptual* primacy unreliable.

Second, Broackes wonders if there are some languages that exercise sets of unique hues different from the RYGB set. But suppose there turned out to be none, and, at the same time, reducing colors to RYGB proved to be more than our parochial cultural practice. That would still imply nothing to the effect of primacy of these hues in low-level color perception, for different explanations would be conceivable. In particular, one could try to argue that the significant status of RYGB categories across languages follows both from other perceptual constraints (as elaborated in this thesis) and from environmental factors (see concluding remarks in chapter 7).

Ingling (1997) defends the phenomenological, language-independent primacy of unique hues very emphatically: "Over the wavelength range, roughly between 520 and 570 nm, the observer will notice no marked transition in hue. The colors seen are various shades of yellow-green. For wavelengths longer than, say, 590 nm, although yellowness persists, there is no green. Upon crossing a wavelength around 575 nm, something has happened. The color changes from greenish to reddish as a point called unique yellow is crossed. It is sheer obstinacy to deny that this transition is not qualitatively different from crossing a wavelength of, say, 550 nm. [...] There are objective transition points in the spectrum that have properties not dependent on language. The fact that discontinuities can be described by language does not mean that languages cause discontinuities."

But slamming down one's hand and saying "really!" does not help. To repeat my point, the statement that a light at 575 nm of wavelength is purely yellow without a hint of red or green is correct, and it is correct because that is how "yellow", "red" and "green" are properly used in English. Rather than "the (non-linguistic) experience of yellow" being pure at that wavelength, it is the case that the (non-linguistic) experience at that wavelength defines the correct use of "(pure) yellow" in English. The perceptual change that occurs when the wavelength is changed from, roughly, 570 to 580 nm is appropriately described as a change from greenish yellow to reddish yellow via pure yellow. The change from 545 to 555 nm is not. Whether these two changes are perceptually comparable or not, that is up to a perceptual color space to decide.[39] I can ascribe no other sense to Ingling's "qualitative difference". The reader may decide whether it is more than ethnocentric naivety, as well as an instance of the "Augustinian conception" of language as criticized by Wittgenstein (1967), to claim that English belongs to a minority of languages of the world that faithfully mirror perception as they apply special terms *just* to the colors that *really are* special.

[39]The latter difference will be probably found smaller, but that is a fact of different resolution abilities in different parts of the spectrum. It is not relevant for Ingling's case: at other wavelengths one can as well find, on one hand, a pair of hues both describable as greenish yellow, and, on the other hand, a pair consisting of a reddish yellow and a greenish yellow hue, such that the perceptual difference (captured in a color space) within the first pair is greater that within the second.

5.2.3 Minor motivations for unique hues

Several other motivations have been proposed for the idea of four perceptually unique hues. (Cf. Fairchild, 2005, ch. 1; Werner and Bieber, 1997.) Sometimes they are cautiously presented as Hering's original impulses for his opponent-processing theory, to which the authors do not explicitly subscribe. Each of these motivations clearly fails to support the notion of the focal RYGB as perceptually unique hues. Here I recapitulate them, as well as their flaws, in order to forestall the possible impression that everything, by and large, points to perceptual uniqueness of RYGB, or that this uniqueness provides a useful account for a whole bunch of perceptual phenomena. Many bad reasons do not constitute a good one.[40]

Color exclusions It has been stated on countless occasions in the literature that red-green and yellow-blue are opposite in that they do not mix; one cannot see a reddish green or a yellowish blue.[41] Let me avoid repeating the qualification that this is true because it respects the appropriate use of "red" etc. in English (cf. Wittgenstein, 1977); for there is an independent problem. It must have been only by force of tradition that only these two pairs were repeatedly noted to be exclusive. There is no more conceptual or perceptual exclusion between (focal) red and green, yellow and blue, than there is between red and turquoise, red and lime, orange and violet, and indefinitely many other hue pairs that are sufficiently remote in the color space.

Color complementarity A usual motivation for schematizing two hues as opposite in "the" hue circle is that the mixture of corresponding lights in equal proportions appears achromatic; that is, the two colors "cancel" to gray.[42] First, indefinitely many hue pairs are like that. Second, red-green and yellow-blue are not. A mixture of red and green light has a yellow tint, mixture of yellow and blue, green. It takes, roughly, turquoise to cancel red to gray and bluish purple to cancel yellow.

Negative afterimages The idea of RYGB color opponency has been also related to the effect of afterimages seen on a white background after adaptation (approximately one minute is enough) to a saturated color stimulus. But the situation is similar to the case of color complementarity. All saturated hues produce

[40]On purpose, I mostly omit references to the literature where these motivations are mentioned, as it is not clear that the authors would want to defend them explicitly. If a particular motivation turns out not to be seriously defended by anyone, the better for my case.

[41]But cf. Crane and Piantida (1983).

[42]Note that this is not necessarily the case for "opposite" hues if the color space is defined on the basis of perceptual difference judgements.

afterimages. Moreover, the afterimage of focal red does not appear green but turquoise, and the afterimage of focal yellow is bluish purple rather than blue.

Simultaneous contrast It has been pointed out that green appearance of a stimulus is supported by red background, yellow appearance by blue background, etc. But although effects of simultaneous contrast are accounted for in modern color appearance models, nobody has, to my knowledge, shown that the perceptual oppositions red-green and yellow-blue come out from these effects as privileged in any way.

Invariant hues Unique hues are occasionally identified with the "invariant hues", that is the hues corresponding to the wavelengths of light that are not affected by the Bezold-Brücke variance of the perceived hue with changes in luminance. Hue invariance seems to be a well-defined concept which can in principle select a small amount of hues as in a sense perceptually privileged. But the attested set of invariant hues does not coincide with the set of unique hues identified on the "linguo-phenomenological" basis. (Cf. Saunders and van Brakel, 1997a; Panorgias et al., 2010.) Hence, hue invariance does not support the mainstream notion of RYGB as perceptually unique hues. At best it can define a different—much weaker—notion of perceptual uniqueness. Furthermore, it should be made clearer whether the fact that exactly four invariant hues are regularly reported reflects an independent finding, of follows to some extent from assumptions formed by the idea of four unique hues.

Color blindnesses It belongs to the original motivations for Hering's opponent theory of color vision, and it is still appealed to by Fairchild (2005, ch. 1), that the common types of color blindness can be described as deficiencies in discrimination between red and green or yellow and blue. But that is just a very loose specification by which primacy of RYGB is assumed rather than corroborated. A neutral description would be in terms of discrimination in various regions of the wavelength range. Known variants of color blindness have been accurately explained in terms of genetic mutations causing deviations in the physiology of the retina, namely in the number of cone types and/or in their sensitivity across the range of wavelength (Mollon, 1995; Jameson and Komarova, 2009b). Fairchild's (2005, p. 30) explanation in terms of "the" red-green or yellow-blue psychophysical channel that cannot be constituted due to these physiological changes is completely theory-laden and fails to illuminate anything.[43]

[43]It cannot even explain the difference between protanopic and deuteranopic color perception, caused by absence of one of the standard types of cones, unless it assumes that "the red-green channel" is more developed in one than in the other type of color blindness.

Cross-linguistic focality Lastly, occasional appeals to straightforwardly linguistic findings can be mentioned: especially to the fact that in a substantial part of the world's languages, color categories are focused roughly in the hues that have been identified as "unique" according to the two dominant definitions. The actual strength of this coincidence is disputable (cf. MacLaury, 1997b; Saunders and van Brakel, 1997b; Jameson, 2010). But the main point is that linguistic evidence is here simple irrelevant in principle. Not even complete agreement of languages on the extension and focusing of color categories would allow direct conclusion to primacy of the respective hues in color perception. For cross-linguistic patterns of categorization, multiple explanations are conceivable (and examined in the present thesis). A representation of the human low-level color perception is an indispensable component in any explanation of the linguistic patterns, but as such it must not be influenced by them. Neither can cross-linguistic patterns of uncertain origin be allowed to co-define the universal, low-level color perception if the latter is to be directly explainable in neurophysiological terms.

5.2.4 There are no unique hues

Having rejected all reasons that, to my knowledge, have been proposed for regarding the focal RYGB hues as perceptually unique or privileged, I conclude that, contrary to the mainstream opinion in color science, these four hues do not have any prominent place in low-level color perception whatsoever. Moreover, no other hues seem to have such perceptual primacy, at least in any way as strong as that which is assumed by the mainstream belief. Unless a brand new evidence is provided, we should assume that there are no unique hues.

5.3 "Unique hues" and color spaces

The rejection of unique hues places all the weight of capturing the perceptual relations of colors on perceptual color spaces. It solves puzzles arising from the alleged need to represent color perception as, simultaneously, continuous *and* discontinuous. There is no conceptual place for unique hues in a perceptual color space like CIELAB, and that is perfectly fine if there are none.

Also, when the uniqueness and fundamental opponency of RYGB is dismissed, so can be all demands to the effect that in a perceptual color space, the red-green and the yellow-blue opposition should coincide with two orthogonal axes. Hardin (2005) presents it as a puzzle that the World Color Survey data suggest "first, that red and yellow are more like each other than either is like blue or green, and second, that green is more like blue than red is like yellow". But that is no puzzle anymore: without assuming four fundamentally opponent unique hues, there is no reason to expect uniform perceptual spacing of the RYGB hues in the first place.

Jameson and D'Andrade (1997) convincingly argue *against* coincidence of the two dimensions of the hue space with the red-green and yellow-blue oppositions. They imply (p. 311) that uniqueness ("non-reducibility") of RYGB is the *only* reason for keeping this traditional notion, as opposed to many good reasons to the contrary. Once the very idea of perceptually unique hues is rejected, this last reason falls.

5.4 Conclusion of chapter 5

In this chapter, I have examined the notion, widespread in the current color science, that there are certain hues, specifically the focal red, yellow, green and blue, that are unique or privileged in human prelinguistic, low-level color perception. I have successively considered and rejected all kinds of motivation that have been provided for this opinion. First, I dismissed the straightforwardly language-dependent notion of unique hues as referents of necessary and sufficient color descriptors. Second, I argued that the alleged "phenomenological" primacy of red, yellow, green and blue is not independent of facts of particular languages. Finally, I refused a number of minor or historical motivations for considering the RYGB hues unique. I conclude that contrary to the mainstream opinion, there is no good reason to claim that some (in particular, RYGB) hues are unique in color perception in a sense that would allow direct neurophysiological explanation that is often called for. The idea of four perceptually unique hues is flawed and has no relevance for construction of perceptual color spaces, which were proposed in the previous chapter as a suitable perceptual basis for explaining cross-linguistic patterns of color categorization. Neither it is, of course, defensible as an independent explanatory principle.

Chapter 6

"Categorical perception"

A substantial part of the recent research in color perception and categorization has been focused on what is known as *categorical perception of color*. With some simplification, categorical perception of color occurs when discriminability of two color stimuli from different categories is increased compared to a pair of stimuli from the same category, despite equal chromatic differences. In this chapter, I examine the notion of categorical perception, as well as relevant experimental results, in order to assess the suitability of this notion and phenomenon as an explanatory principle for linguistic categorization of color.

I begin the chapter with a consensual picture of the phenomenon in question and of the current state of the field. After that, I question several aspects of the way categorical perception is typically conceived in the literature. This in particular concerns categorical perception as observed in infants and prelinguistic toddlers. Challenging the mainstream views of the phenomenon has direct consequences for the question of the possible role of (pre-linguistic) categorical perception in explaining patterns of color categorization in languages of the world.

6.1 Categorical perception of color: state of the field

In the recent literature on categorical perception of color, the following definition is generally agreed upon. Perception of color samples is categorical if the discrimination of stimuli across linguistic categories is better (faster, more accurate) than the discrimination of stimuli within a category, in spite of equal perceptual spacing of the stimulus pairs. (Cf. Franklin and Davies, 2006; Clifford et al., 2011; Davidoff and Fagot, 2010; Franklin et al., 2009, 2008a,b; Clifford et al., 2009.) Although Harnad (1987), which is a canonical reference, talks of equal *physical* differences within the stimulus pairs, recent works regularly consider equal *perceptual* spacing as measured in some perceptual color space. Also, several studies note that the contemporary notion of categorical perception is more adequate than the earlier one in that it does not require complete lack of discrimination for within-category pairs. Some authors, echoing Harnad's (1987) definition, describe categorical perception in terms of perceptual *similarity* vs. *difference* in, respectively, within- and cross-categorical stimulus pairs (Brown et al., 2011) – it will be later shown that this way of presenting the phenomenon is inaccurate.

In the past decade or so, categorical perception (hereafter, CP) has been subject to intensive research and effects of this kind have been observed in a variety of experimental settings. That includes different age groups of subjects (adults, infants, to a lesser extent toddlers before and after acquisition of color terms), from different language groups (speakers of English, Russian, Greek, Korean, Berinmo or Himba) studied individually and in comparison. Also, it spans various experimental techniques, both behavioral and neurological, suitable to infants (habituation technique, novelty-preference technique), adults (same-different judgement task, odd-one-out judgement task, two-alternative forced choice task), or both infants and adults (visual search task with reaction times measured by means of eye-tracking, measuring event-related potential on the scalp during an "oddball" task). The color samples employed in the experiments are typically differentiated in hue (lightness and saturation remaining constant), with various size of the chromatic difference within the stimulus pairs; the size of the difference within stimulus pairs of the particular research is balanced in a perceptual color space (the Munsell system, CIELUV).

At present, it is generally acknowledged that there is substantial evidence in favor of both *language-induced* categorical perception of color, in a Whorfian sense, and *prelinguistic* color CP. Only few studies (Davidoff et al., 2009; Brown et al., 2011) report failures to find expected CP effects.

On the linguistic side, the evidence is manifold. First, categorical perception occurs for adult speakers only across boundaries of categories that are strongly lexicalized, by means of basic color terms, in the speakers' own language. That is typically shown for particular boundaries by experimental comparison of English speakers with speakers of a language that either subsumes more English categories in one (such as Himba, with a single color term covering green and blue), or splits an English category into more (such as Russian and Greek, with their distinct basic terms for light and dark blue). (Franklin and Davies, 2006; Clifford et al., 2011; Davidoff and Fagot, 2010; Franklin et al., 2009; Jraissati, 2012; Ozturk et al., 2013.) Second, intensive short-term training in artificial categories (such as, training English speakers to split the scale of green into bluish and yellowish green) induces categorical perception effects at the newly learned categorical boundary. (Özgen and Davies, 2002; Clifford et al., 2011; Drivonikou et al., 2011; Clifford et al., 2012.) Third, in explicit confrontation of adult right-eye and left-eye color perception, categorical perception effects have been located predominantly in the right visual field. As the right-eye input is processed in the left hemisphere, which is also thought to be responsible for most of language processing, this finding is taken to suggest dependence of adult color CP on color language. (Clifford et al., 2011; Drivonikou et al., 2011; Clifford et al., 2012; Davidoff and Fagot, 2010; Franklin et al., 2008a,b.)

From the Whorfian perspective, it is an important question whether these effects actually reflect influence of language structures on some kind of "low level" color perception, or whether they can be explained simply by the fact that cognitive performance in the considered tasks is improved by direct recourse to available linguistic labels. Early findings in categorical perception were susceptible to the latter objection, allowing explanation in terms of a "naming strategy". For instance, if one is to decide which of two color samples is identical to a sample that was displayed several seconds or minutes before, it is obviously helpful to have the two colors distinguished by color terms, as the verbal label for the first sample is easier to remember than the particular color itself. However, the above reported evidence for language-induced color CP rests mainly on more recent experimental techniques, in particular on the visual search task and event-related potential (ERP) measuring. These techniques minimize the role of memory and effectively rule out the possibility of improving the task performance via conscious labeling. That is the least problematic sense we can give to the statement that the language-induced CP effects in adults are indeed "perceptual" (cf. Franklin and Davies, 2006; Drivonikou et al., 2011). In this specific sense, it is quite sound to claim that the Sapir-Whorf hypothesis holds; that is, that language affects (color) perception. The term "perceptual", however, is used in various ways in the literature; a more problematic sense of the claim will be discussed in the following section.

Evidence for prelinguistic categorical perception is provided by experiments on infants of 4 to 9 months of age, using various experimental techniques, most recently eye-tracking of the child's visual search as well as ERP measuring. (Bornstein et al., 1976; Franklin and Davies, 2004; Franklin et al., 2005b; Clifford et al., 2009; Franklin et al., 2008a; Ozturk et al., 2013.) The results are supported by a smaller number of studies on toddlers without consistent knowledge of the basic color terms of their language. (Franklin et al., 2005a, 2008b, 2009.) Through assessment of discrimination performance on within- vs. cross-categorical color stimulus pairs, categorical perception has been (claimed to be) observed at several category boundaries. It is regularly reported at the green-blue boundary, and individual studies have found it at the blue-purple (Franklin and Davies, 2004; Ozturk et al., 2013), red-pink (Franklin and Davies, 2004), green-yellow and red-yellow boundary (Bornstein et al., 1976). Lateralization studies by Franklin et al. (2008a) and Franklin et al. (2008b) report a CP effect only in the left visual field (right hemisphere) for infants and prelinguistic toddlers, as opposed to CP effect only in the right visual field (left hemisphere) for adults and competent toddlers. A common conclusion from all these findings is that human prelinguistic color perception is categorical, and that more research is necessary to clarify the relationship between these prelinguistic categories on one hand and linguistic color categories as well as language-induced categorical perception on the other.

In the conceptual examination which follows, I will have more to say on the matter of prelinguistic "categorical perception" than about language-dependent CP in adults. One reason is that, obviously, only the prelinguistic facts of color perception can be appealed to in explaining linguistic categorization of color. Explanation of cross-linguistic patterns of color categorization in terms of structures of color perception that themselves depend on categorical systems of particular languages would be circular. Furthermore, to me the conceptual confusions of the mainstream view of infant "categorical perception" appear somewhat more serious than those in the case of adult CP. The former, unlike the latter, arguably devalue the empirical results of the respective subfield to a considerable degree, for they lead, first, to inadequate conclusions, and second, to less appropriate experimental design in subsequent research. Clearing up the conceptual inadequacies is desirable both for further development in the field and for correct assessment of the possible explanatory role of prelinguistic color CP with respect to cross-linguistic patterns of color categorization.

6.2 How to think of categorical perception

6.2.1 Does categorical perception warp the perceptual color space?

Categorical perception of color is often described (or sometimes even defined) in terms of greater perceptual *similarity* and *difference* between the color samples of, respectively, within- and cross-categorical pairs. (Cf. Harnad, 1987; Brown et al., 2011; Davidoff and Fagot, 2010; Davidoff et al., 2012; Jraissati, 2012; Franklin et al., 2008b). And throughout the field, there seems to be a substantial agreement on the notion that categorical perception effects might be explicable in terms of distortion or "warping" of the perceptual color space, namely by its expansion in some regions and compression in others. Consider the following quotations:

- "It is as though perceptual colour space has been transformed topologically or "warped" [...] The transformation stretches perceptual distances across category boundaries relative to within-category distances." (Franklin and Davies, 2004, p. 351.)
- "Learning colour terms may highlight similarities among colours given the same term and highlight differences among colours given different terms, leading to within-category compression and between-category expansion of

the perceptual colour space, particularly for RVF (LH) [right visual field, left hemisphere] stimuli." (Drivonikou et al., 2011, p. 253.)

- "[T]he results uphold the view that the structure of linguistic categories distorts perception by stretching perceptual distances at category boundaries [...] It would appear that the internal color space [...] is not static; some distances within it are "stretched" or "distorted" by the influence of color labels." (Davidoff and Fagot, 2010, p. 102).

Cf. also Jraissati (2012, p. 441), Clifford et al. (2011, p. 238), and Franklin et al. (2009, p. 243-244). Sometimes, this is more or less explicitly presented as the other option besides explanation in terms of "naming strategy" (see section 6.1; Franklin et al., 2009, p. 242): either the observed CP effects are a consequence of the subject's recourse to conscious verbal labeling, or they reflect the character of the subject's internal color space and its modifications by language. That presents the latter as coinciding with the position that CP effects are indeed perceptual.

I will argue that this is a wrong way of conceiving the phenomenon, in adults as well as in infants. To make my case clear: I do not strongly maintain either a Whorfian or an anti-Whorfian position here. I will only reject the particular interpretation of existing CP findings both in adults[44] and in infants which directly links these findings to the character of the subject's perceptual color space.

Perceptual color spaces, such as CIELAB and CIELUV, and color order systems, such as the Munsell system, are intended to represent the ideal topology of colors ("the perceptual color space") as given by relations of identity, similarity and difference. (See Chapter 4.) They are thus properly built on the basis of standard trichromatic observers' *judgments* of these relations. Per contra, none of the several experimental paradigms supporting adult and infant categorical perception provides us with judgments of color similarity/difference within the employed stimulus pairs. Instead, the experimental techniques measure behavioral or neurological *performance* in discrimination of the paired samples.

The most recent behavioral technique, that is, visual search experiments, does not concern judgments at all, but speed and accuracy in visual detection of a colored target on a background or among distractors that are either within-, or cross-categorically different. Obviously, one cannot *assess* the color difference of a target from its background before one detects the target – but at that point the reaction time is already noted down and that particular trial of the experiment is finished.[45] Also the modern ERP approach in no way deals with color similar-

[44] In adults, the results indeed seem to point in a Whorfian direction quite consistently.

[45] Davidoff et al. (2012) distinguish "perceptual similarity" and "categorical similarity" as two modes of judging similarity of colors, the latter being "default" and manifested in "implicit judgment tasks" such as the visual search task. That is heavily confused, since the authors completely ignore the fact that the visual search task involves *no similarity judgment at all* and

ity judgments and does not allow conclusions regarding similarity and difference within stimulus pairs. It measures neural responses to presentation of deviant ("oddball") color samples among majority of "standard" samples which are, again, either within-, or cross- categorically different from the deviant.

Admittedly, the older experimental techniques such as the same-different judgment task or the two-alternative forced choice task do involve color identity, similarity and difference judgments. However, they do *not* elicit judgments of these color relations within the employed stimulus pairs, such as, "sample A is more similar to B than it is to C". Rather, they establish how *fast and accurate* the observers are in reporting that two color samples (a within-, or cross-categorical pair) are different, or which of two samples is identical to a previously displayed one when the other is either within- or cross-categorically different.[46]

The appropriate conclusion from these sorts of evidence seems to be that discrimination in within-categorical stimulus pairs is significantly slower, more cognitively demanding and more prone to error. However, the conclusion that the stimuli in the within-categorical pairs are more perceptually *similar* (less *different*) than those in the cross-categorical pairs either is fallacious, or involves a tacit redefinition of "similarity". The color identity, similarity and difference relations which constitute the ideal perceptual color space and which are topologically represented in artificial color spaces are consensually revealed via gathering observers' *considered judgments* of these relations. They are *not* defined in terms of neurological response patterns, perceptual performance on color sample pairs or judgment performance under time pressure. All the results of this latter kind are worth the attention they get in the literature, and arguably support a particular reading of the Sapir-Whorf hypothesis (see above), but they must not be confused with the relations that are actually captured in perceptual color spaces. For instance, it is quite possible that there is a shade of yellow-green which is, by considered judgment of an average observer, as different from focal green as it is from focal yellow, yet is more easily detectable (in terms of speed and accuracy) on a focal yellow than on a focal green background. The common talk of the perceptual color space being "warped" in categorical perception (more specifically, compressed in some regions and expanded in others) is not justified by the available CP evidence.

One might object that the "warped space account" is not a conclusion drawn from the existing CP findings, but rather a hypothesis proposed to *explain* these findings. Surely, if two color samples are more perceptually similar (in the usual

they present this task in line with matching-to-sample tasks where similarity judgments are more or less explicitly required (and, not surprisingly, found).

[46]We need not discuss in detail the techniques of earlier research on infant color categorization (that is, the habituation and the novelty-preference paradigm; Bornstein et al., 1976; Franklin and Davies, 2004), since it is even less clear to what extent the results reflect color similarity relations, as opposed to effects of memory and color preference.

sense of considered judgment), their discrimination is likely to be slower or more difficult. Thus, if languages, to some extent, topologically transform the perceptual color space of their speakers, that could provide an explanation for the observed language-induced categorical perception effects.

But that does not work either. Language independence and uniformity of the perceptual color space across (standard trichromatic) speakers of different languages is too fundamental an assumption to be given up. First and foremost, artificial perceptual color spaces are built and used with this assumption. And in the CP research, the color stimuli in within- vs. cross-categorical pairs are regularly chosen so as to even up the within- and cross-categorical chromatic difference, as measured in a color space such as CIELUV or the Munsell color system. Once we accept that languages have impact on the perceptual color spaces of their speakers, the existing evidence regarding color CP becomes worthless. That is because we thereby lose the assumption that the within- and cross-categorical stimulus pairs were of equal chromatic difference (represented, say, by 4 Munsell hue steps) for speakers of any particular language (such as English, Greek or Himba). In other words, in such case there will be a potential explanation, but not much to be explained – no observed categorical perception (according to the consensual definition of CP).

Language either does, or does not influence the perceptual color space. (Independent reasons for the latter were strong enough to make it a fundamental assumption of virtually all color science of the last hundred years or so; cf. Chapter 4.) The existing findings on categorical perception cannot corroborate the former position, since they clearly reflect a different phenomenon; and they cannot be explained by this position, since they all presuppose the contrary. In this sense, there is no point in either explaining or describing categorical perception in terms of a distorted color space.

6.2.2 How categorical is categorical perception?

In one sense, the question "how categorical is color perception?" is simply concerned with whether there indeed are significant CP effects at some or all of the assumed category boundaries. (Cf. Brown et al., 2011; Jraissati, 2012.) This section deals with a different sense of the question: does categorical perception, as documented by the existing research, provide us with anything that can be in a strong sense called (perceptual) *categories*? That is, do CP effects divide the perceptual color space into more or less discrete chunks, in a way comparable with how it is typically partitioned by linguistic categories (notwithstanding vagueness)?

In the previous section I have argued that all the observed "perceptual" CP effects characterize human perceptual performance and processing, and that they must be conceived as distinct from the structure of the perceptual color space. They still belong to "low-level" color perception, but cannot be reduced to the basic, presumably universal relations of similarity and difference. Instead, categorical perception effects should be considered in addition and with reference to the perceptual color space, possibly in the form of an additional dimension. For instance, one can compare the perceptual performance in discrimination of equidistant color points in the red vs. the blue region of the space.[47] Indeed, once we appreciate the distinctness of the performance and processing issues from the core similarity relations among colors, there seems to be *no reason* to expect complete homogeneity of the former with respect to the similarity-based color space. (Of course, drastic discrepancy between similarity of two color points and discrimination performance on them is not likely, but that does not imply any strong logical dependence between these two characteristics. In order to check the strength of the correlation between them, one needs to conceive them as distinct in the first place.)

This conception of low-level color perception, however, makes it apparent that the performance and processing effects can hardly define any "absolute" categories, that is, strictly discrete regions in the color space. They cannot, unless we want to assume that there may be pairs of perceptually different (non-identical) color points which we nonetheless completely fail to discriminate. That seems to be conceptually ruled out.

Still, it is well possible that these effects define reasonably strong "relative" categories. For instance, it might be that for adult English observers, discrimination performance is differentiated across the color space in such a way that discrimination of equidistant color points is consistently and markedly easier *across* than *within* the regions delimited by the English basic color terms "red", "green", "white", "orange", "brown" etc. Such is, at least, a common picture of the functioning of language-induced categorical perception. In my opinion, much needs to be done in order to confirm this picture. The reason is that existing studies in support of categorical perception usually compare *one* cross-categorical with *one* within-categorical stimulus pair (or, at best, several within- and cross-categorical pairs), and typically at the green-blue boundary. Clearly, each examined stimulus pair provides just a tiny fragment of all the evidence that would be necessary in order to conclude that English induces strong and consistent "relative" perceptual categories correspondent to its linguistic categories. Most categorical bound-

[47]Non-trivial differences in perceptual discrimination across the color space might have straightforward physiological causes. Suppose that for some reason, the short-wave cones on the retina react to light somewhat slowlier than the two other types of cones. That would presumably lead to a state where a green-blue sample which is, by considered judgment, as similar to focal green as it is to focal blue is nonetheless more easily discriminable (in terms of speed and accuracy) from focal green than from focal blue.

aries other than green-blue are virtually unexplored. But overall, the research in language-induced categorical perception appears to be on the right track. The stimulus pairs around and across the categorical boundaries of any particular language seem to be especially informative, and also, some attention has been paid to performance differentiation within a category, related to its prototypical structure (Jraissati, 2012; Clifford et al., 2012).

In case of infant color perception, the situation is different and requires substantial conceptual clarification. According to the consensual definition, categorical perception of color occurs if there is a cognitive advantage for discrimination of stimuli from different categories, compared to within-categorical stimuli, despite equal chromatic differences. In case of adults, this definition clearly refers to *linguistic categories*; or that is how it is unanimously interpreted and operationalized in research. For infants, no modification of the categorical perception concept has been proposed and the same (underspecified) definition is explicitly or implicitly applied, regardless of the fact that *no linguistic categories can be assumed* in 4-to-9-month-old infants. In the practice of research, it is still the boundaries between adult linguistic categories (in particular, green-blue), what is being examined in infant CP research. (Cf., among others, Franklin and Davies, 2006; Franklin et al., 2008a; Jraissati, 2012; Ozturk et al., 2013.)

Now, this is reasonable as far as the research question is: does infant color perception manifest CP (-like) effects at boundaries of adult linguistic categories? To answer this question, it is appropriate to compare infants' discrimination performance on a couple of equidistant stimulus pairs within vs. across categories such as blue and green. It is, however, fallacious to infer from positive evidence of this kind that there is *anything* comparable to adult categories in infant color perception. For simplicity, those effects can still be labeled "categorical perception", but this sense of "categorical" is so weak that it actually cannot sustain any self-standing *categories*.[48]

Yet there are plenty examples that such is a usual conclusion from finding infant CP(-like) effects at a particular linguistic boundary:

- "[F]our-month-old infants [...] respond categorically to colour. Furthermore, it was shown that four-month-old infants not only have primary categories such as blue and green, but also have secondary categories such as purple and pink." (Franklin and Davies, 2006, p. 108.)

- "Four-month-old infants categorize a range of colours – blue, yellow, green,

[48] My point is not to be reduced to the trivial one that research on, *e.g.*, the green-blue boundary does not in itself license conclusions with respect to other color categories, such as red or orange. Rather, my claim is that a significant infant CP(-like) effect observed in stimulus pairs spanning the boundary between the linguistic categories blue and green does not even allow a conclusion to the effect of there being corresponding, blue and green, categories in infant color perception.

red, purple and pink have been tested so far." (Franklin and Davies, 2006, p. 113.)

- "These studies show that CP effects in infants occur in the LVF [left visual field], while it occurs in the RVF with adults. This suggests that there would indeed be innate categories, independent of language in infants, which would at a later stage be over-ridden by language dependent boundaries [...]" (Jraissati, 2012, p. 444.)
- "Our findings provide independent evidence for the existence of color categories in prelinguistic infants [...]" (Ozturk et al., 2013.) "[L]anguage is not necessary for color categories in humans. [...] What are the color categories that infants begin with?" (Ozturk et al., 2013, p. 114.)
- "The relation between prelinguistic and linguistic CP remains unclear. One possibility is that language makes fairly minor language-specific adjustments to a universal set of prelinguistically available categories. Another possibility is that language carves its categories into cognition *de novo*, without building on prelingustically available categories." (Franklin et al., 2008a, p. 3222.)
- "[W]hat categories are prelinguistically available in the RH? How do these categories compare *extensionally* to linguistic color categories, and is *their extension* governed by similar forces?" (Franklin et al., 2008b, p. 18224, italics are mine and meant to emphasize the strong sense of the infant categories presupposed here.)

The weak point of the mainstream reasoning is that CP(-like) effects at particular adult linguistic boundaries are believed to directly reflect boundaries of infant perceptual categories (presumably in the sense of strong "relative" categories as defined above). This is not justified, since there is no good reason to assume that the regions of the color space[49] where languages place their categorical boundaries are of special salience in prelinguistic color perception.[50] Only if we already presuppose that *either* there are infant perceptual categories corresponding to adult

[49]Here, I fully adopt the somewhat non-trivial assumption made in all existing research on infant categorical perception, that the perceptual color space and its approximations in artificial color spaces are reasonably valid even for infants as young as 4 months.

[50]I believe that the possible impression that there are such good reasons is false. In chapter 5 of the present thesis I reject the opinion that justification for prelinguistic salience of red, yellow, green and blue can be drawn from neurophysiology of color or from language independent color phenomenology. On the psychological side, Eleanor Rosch's influential notion of prelinguistically available, universal color categories has been severely undermined by the cross-cultural research of Debi Roberson and colleagues (Roberson et al., 2000, 2005).

Notice, however, that this does not amount to a complete dismissal of any idea of color categorization in absence of language. A child, or a sparrow, in picking cherries, will go for the dark red ones, and that is a way of behavioral categorization of color in context. But I do linguistic categories, *or* there are no infant perceptual categories at all, we can regard an observed CP-like effect across the green-blue boundary as evidence for there being categories of green and blue in infant color perception.

Once again: upon appreciation that discrimination performance is a distinct issue from the color similarity relations constituting the perceptual color space, there is little reason to expect perfect homogeneity of the former with respect to the latter. On the contrary, we should expect to find CP-like effects on equidistant stimulus pairs from various regions of the color space, irrespective of linguistic categories.[51] It is surely remarkable that these effects *do* occur in infant color perception at certain adult linguistic boundaries, as demonstrated by a handful of studies from the last decade (Franklin and Davies, 2004; Franklin et al., 2005b, 2008a; Clifford et al., 2009; Ozturk et al., 2013; Franklin et al., 2005a, 2008b, 2009). However, these are, to my knowledge, the *only* locations in the color space that have been examined until now. Since we cannot assume linguistic categories to be of any relevance for infant perception, the well-attested CP-like effect at the green-blue boundary does not in any way guarantee that, first, a similar effect takes place, say, at the green-yellow boundary, and second, that a similar effect does *not* take place, for instance, in the middle of the green, blue or yellow linguistic category. Moreover, the green-blue boundary is the only linguistic boundary at which infant CP-like effects are supported by a massive body of evidence. For other boundaries, the evidence is thin (blue-purple) or dubious (red-pink, green-yellow, red-yellow; see 6.3.2).

We are forced to conclude that contrary to the mainstream views of infant "categorical perception", on the basis of the existing results, nearly nothing can be said on the question whether infants perceive colors categorically, and if they do, to what extent their categories coincide with the linguistic color categories of any particular language.

---

not see any chance of generally grounding a privileged prelinguistic position of the linguistically salient boundaries between red, yellow, green, brown, purple, etc. in this sort of categorization.

[51] That is contra Clifford et al. (2009), according to whom "it has been somewhat surprising that prelinguistic infants respond categorically to color on behavioral tasks". Given what CP experiments actually investigate (namely, discrimination performance and processing), it can hardly be too surprising when a CP-like effect is found at any particular location of the color space.

6.3 Infant "categorical perception" in modeling linguistic color categorization

6.3.1 Ideally...

Clearly, not the adult categorical perception, induced by language patterns, but only the infant CP-like effects can be appealed to in explaining the observed cross-linguistic patterns of color categorization. This explanatory strategy is rather obvious and has been suggested (Clifford et al., 2009). However, it does not seem particularly fruitful as long as we assume that infant "categorical perception" amounts to there being strong perceptual categories for infants, most likely coinciding with a set of basic linguistic color categories of English; for there are many languages that do *not* categorize color like English.

Nonetheless, the above proposed conception of infant "categorical perception" fits well in the color categorization model architecture that is being elaborated in the present thesis. Namely, the CP-like effects (read: differentiation of perceptual performance across the color space) can be readily seen as an additional perceptual constraint, besides the very color space, on the game-theoretic interaction of individual agents by which development of linguistic categories can, arguably, be modeled. In this setting, it is not necessary that perceptual categories coincide with the linguistic in order to explain their formation, and there need be no strong perceptual categories at all. Even a feeble performance differentiation over the color space might, in interaction with other perceptual and game-theoretic constraints of the model, lead to the familiar patterns of color categorization observed in the world's languages. Such differentiation is another factor that is, in principle, capable of supressing rotational arbitrarity (see section 4.9) and promoting particular locations for categorical boundaries. As I am apparently the first to consider this as a distinct characteristic of color perception, I cannot refer to any research in favor of its explanatory role with respect to linguistic categorization. My point is rather, first, that a realistic model of an individual human perceiver should include this component, and second, that a proper evaluation (as outlined in chapter 3) will decide whether adding it increases the performance of the color categorization model that is based on game-theoretic interaction of such perceiving agents.

Technically, this factor could be implemented in various manners, also because "discrimination performance over the color space" is a phenomenon too complex to be exhaustively captured by a single value for each pair of color points. CP effects have been reported mostly in the hue dimension.[52] Thus, to start with the simplest, it would be useful to have a curve which would, for each point in the Munsell hue circle, provide some infant performance value (say, reaction time

[52] An exception is Franklin and Davies (2004), who also examine the red-pink boundary.

under particular conditions) characterizing discrimination of this point from the point, say, 2 Munsell hue steps to the left.

6.3.2 In practice...

Unfortunately, nothing as complete is available at the moment. Here, I summarize all the fragmentary evidence in this respect.

As noted above, the only region of the color space that is sufficiently covered is the green-blue boundary, at the middle level of lightness. Franklin and Davies (2004); Franklin et al. (2008a,b); Clifford et al. (2009) report CP-like effects on variously spaced equidistant Munsell stimulus pairs in the boundary's neighborhood.[53] Franklin et al. (2005b) and Ozturk et al. (2013) find similar effects on samples that are equidistant according to the CIELUV color space. On the blue-purple boundary, infant CP-like effects have been observed by Franklin and Davies (2004), using Munsell stimulus pairs, and Ozturk et al. (2013), using stimulus pairs equidistant in CIELUV.[54] For other boundaries, the evidence is rather questionable. Franklin and Davies (2004) find a CP-like effect also at the red-pink boundary, but given their novelty-preference method, this finding may also be a consequence of the fact that infants prefer red to pink.[55] Bornstein et al. (1976) report a CP-like effect at the red-yellow, green-yellow and green-blue boundary, but this finding is not reliable, first, because the employed habituation method does not keep apart perception, preference and memory, and second, because the chromatic differences within the stimulus pairs were balanced in a physical (wavelength), not perceptual space.

More research on infant performance differentiation over the perceptual color space is necessary before perception effects of this kind can be integrated as a full-fledged explanatory component into models of linguistic color categorization.

[53]They use the following ranges of Munsell stimuli, in each case assuming the boundary at 7.5BG: 7.5B - 5B - 2.5B - 10BG - 7.5BG - 5BG - 2.5BG - 10G - 7.5G (Franklin and Davies, 2004); 2B - 2BG - 2G (Franklin et al., 2008a); 10BG - 5BG - 10G (Franklin et al., 2008b, on prelinguistic toddlers); 2.5B - 2.5BG - 2.5G (Clifford et al., 2009).

[54]For chromaticity coordinates of the green-blue and blue-purple stimuli used in Franklin et al. (2005b) and Ozturk et al. (2013), see Table 1 in the latter study. Franklin and Davies (2004) use the blue-purple range 10B - 2.5PB - 5PB - 7.5PB - 10 PB - 2.5P - 5P - 7.5P - 10P (assuming boundary at 10 PB).

[55]As attested in Franklin et al. (2008c). Color preference is another part of broadly conceived infant color perception that would be worth considering as additional explanatory factor for linguistic categorization patterns. Investigating a set of 8 English basic color categories, Franklin et al. (2008c) report infants' significant preference for red and significant dispreference for brown and pink.

6.4 Conclusion of chapter 6

In this chapter, I have examined the possible explanatory role of the phenomenon referred to as "categorical perception of color" with respect to cross-linguistic patterns of color categorization. To properly appreciate this role, some conceptual clarification was necessary. First, I have argued against the mainstream notion of categorical perception as involving a distortion of the perceptual color space. The effects observed in the categorical perception research concern color discrimination performance and processing, not relations of color similarity and difference. Therefore, they need to be conceived independently of the perceptual color space. Second, I have challenged the mainstream opinion that the existing evidence on infant "categorical perception" allows to conclude that infants perceptually categorize color, and in particular, that they have perceptual categories that resemble the basic color categories of English. Such conclusions rest on an unjustified interpretation of the infant "categorical perception" findings in terms of adult linguistic categorical boundaries. Finally, I have shown that the differentiation of infant discrimination performance across the perceptual color space, which is being to some extent revealed by the contemporary categorical perception research, is in principle a suitable explanatory component for a color categorization model along the lines of the present thesis. Notably, this is another possible factor, apart from considering perceptually heterogeneous populations, to suppress rotational arbitrarity and favor a particular placement of categorical boundaries. However, the current body of evidence is so fragmentary that this factor cannot be readily integrated in a working model.

Chapter 7

Conclusion and what is left

This thesis has rather extensively dealt with the problem of color categorization in language and how to provide a plausible explanation, via successful modeling, for its cross-linguistic patterns. Chapter 2 has defined the problem and located it in a historical, disciplinary, and ideological context. Chapter 3 has elaborated on the question what would be a satisfactory solution, or how a color categorization model should be evaluated, by way of addressing the highly non-trivial question of what the cross-linguistic patterns of color categorization actually are. Chapters 4 to 6 have examined several aspects of the appropriate "perceptual basis" for a color categorization model, that is, the assumptions about color perception by individual (normal and color-deficient) agents that such a model should arguably incorporate. While the matter of the chapters 5 and 6 has turned out be of limited practical importance in improving the existing modeling approaches to linguistic categorization, the conceptual investigations of these two chapters have led to conclusions that are far-reaching in that they confute some of the crucial assumptions of the respective subfields of color science.

What has been put forward is, to be sure, not yet a model, but it is not even a full proposal for one. It is rather an attempt to provide a firmer ground for the increasingly popular approach to color categorization using simulated evolutionary game-theoretic interaction of individual, perceiving and communicating agents. In this thesis I have tackled the rather neglected question *what these agents should be like.* Time has not allowed me to cope with two more questions of equal importance.

First, *what should the agents do*? That is, how should the evolutionary interaction proceed, based on what sort of a game and with what kind of dynamic? Various interaction principles have been assumed; basically, either in the Steelsian line, using reinforcement learning upon a "dicrimination game" and a "category game", cf. Steels and Belpaeme (2005); Jameson and Komarova (2009a); Baronchelli et al.

(2010); Loreto et al. (2012), or in a more narrowly game-theoretic framework using similarity maximization games, cf. Jäger and van Rooij (2007); Correia and Ocelák (2013). None of the choices seems to be especially well motivated. Sure, it would be relatively straightforward to pick one of these settings at random and extend the results of the present thesis into a working model. But one has the impression that a satisfactory answer to the question of an appropriate form of interaction for a model of this kind might well grow up to the size of one more thesis.

Second, *what should the agents' input be like*? For the sake of simplicity, the recent models have typically presupposed that the empirical color input, upon which the evolutionary interaction takes place, is uniformly distributed over the perceptual space,[56] and also that the ecological (or cultural) importance of colors does not vary across regions of the space. That nonetheless hardly corresponds to the conceivable origins of human color language. Admittedly, color distribution in the environment as well as diversified ecological significance of particular colors seem less suited as straightforward explanations for the cross-linguistic tendencies of color categorization. That is because both these external constraints on interaction vary substantially across human environments. Still, these two factors are worth more investigation, as they suggest additional possible principles, besides those discussed in chapters 4 and 6, capable of suppressing rotational arbitrarity and favoring a particular placement of categorical boundaries.

The present thesis has not provided any easy formal results. I nonetheless hope that it has, informal through and through, somewhat swept the path towards more thorough ones.

[56]But cf. Yendrikhovskij (2001); Steels and Belpaeme (2005)

Bibliography

A. Baronchelli, T. Gong, A. Puglisi, and V. Loreto. Modeling the emergence of universality in color naming patterns. *Proceedings of the National Academy of Sciences*, 107:2403–2407, 2010.

D. Baylor. Colour mechanisms of the eye. In T. Lamb and J. Bourriau, editors, *Colour: Art and Science*, pages 103–126. Press Syndicate of the University of Cambridge, 1995.

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

C. P. Biggam and C. J. Kay, editors. *Progress in Colour Studies, Vol. I: Language and Culture*. John Benjamins, Amsterdam/Philadelphia, 2006a.

C. P. Biggam and C. J. Kay, editors. *Progress in Colour Studies, Vol. II: Psychological Aspects*. John Benjamins, Amsterdam/Philadelphia, 2006b.

C. P. Biggam, C. A. Hough, C. J. Kay, and D. R. Simmons, editors. *New Directions in Colour Studies*. John Benjamins Publishing Company, Amsterdam/Philadelphia, 2011.

J. Birch. Worldwide prevalence of red-green color deficiency. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, 29:313–320, 2012.

M. Bornstein, W. Kessen, and S. Weiskopf. Color vision and hue categorisation in young human infants. *Journal of Experimental Psychology: Human Perception and Performance*, 2:115–129, 1976.

M. H. Bornstein. Hue categorization and color naming: Physics to sensation to perception. In N. J. Pitchford and C. P. Biggam, editors, *Progress in Colour Studies, Vol. II: Psychological Aspects*, pages 35–68. John Benjamins, Amsterdam/Philadelphia, 2006.

R. Brandom. *Making it explicit: Reasoning, representing and discursive commitment*. Harvard University Press, USA, 1998.

M. H. Brill. When science fails, can technology enforce color categories? *Behavioral and Brain Sciences*, 20:182–183, 1997.

J. Broackes. Could we take lime, purple, orange and teal as unique hues? *Behavioral and Brain Sciences*, 20:183–184, 1997.

J. Broackes. Where do the unique hues come from? *Revue of Philosophy and Psychology*, 2:601–628, 2011.

A. M. Brown, D. T. Lindsey, and K. M. Guckes. Color names, color categories, and color-cued visual search: Sometimes, color perception is not categorical. *Journal of Vision*, 11:1–21, 2011.

R. W. Brown and E. H. Lenneberg. A study in language and cognition. *Journal of Abnormal and Social Psychology*, 49:454–462, 1954.

A. Byrne and D. R. Hilbert. Unique hues. *Behavioral and Brain Sciences*, 20: 184–185, 1997.

CIE. *International lighting vocabulary (Publication CIE No. 17.4)*. Bureau Central de la Comission Electrotechnique Internationale, 1987.

A. Clifford, A. Franklin, I. R. L. Davies, and A. Holmes. Electrophysiological markers of categorical perception of color in 7-month old infants. *Brain and Cognition*, 71:165–172, 2009.

A. Clifford, A. Franklin, A. Holmes, and I. R. L. Davies. Investigating the underlying mechanisms of categorical perception of colour using the event-related potential technique. In C. P. Biggam, C. A. Hough, C. J. Kay, and D. R. Simmons, editors, *New Directions in Colour Studies*, pages 237–249. John Benjamins Publishing Company, Amsterdam/Philadelphia, 2011.

A. Clifford, A. Franklin, A. Holmes, V. G. Drivonikou, E. Özgen, and I. R. L. Davies. Neural correlates of acquired category effects. *Brain and cognition*, 80: 126–143, 2012.

H. C. Conklin. Hanunóo color categories. *Southwestern Journal of Anthropology*, 11:339–344, 1955.

B. R. Conway and C. M. Stoughton. Response: Towards a neural representation for unique hues. *Current Biology*, 19:R442–R443, 2009.

R. Cook, P. Kay, and T. Regier. WCS Data Archives. Available at http://www.icsi.berkeley.edu/wcs/data.html (approached January 2013).

J. P. Correia and R. Ocelák. Modeling basic color term usage with similarity-maximization games. Unpublished, 2013.

A. Costall. "Colour science" and the autonomy of colour. *Behavioral and Brain Sciences*, 20:185, 1997.

H. D. Crane and T. P. Piantida. On seeing reddish green and yellowish blue. *Science*, 221:1078–1080, 1983.

J. Davidoff and J. Fagot. Cross-species assessment of the linguistic origins of color categories. *Comparative cognition and behavior reviews*, 5:100–116, 2010.

J. Davidoff, J. Goldstein, and D. Roberson. Nature vs. nurture: the simple contrast. *Journal of Experimental Child Psychology*, 102:246–250, 2009.

J. Davidoff, J. Goldstein, I. Tharp, E. Wakui, and J. Fagot. Perceptual and categorical judgments of colour similarity. *Journal of cognitive psychology*, pages 1–22, 2012.

D. Dedrick. *Naming the Rainbow: Colour Language, Colour Science, and Culture*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.

D. Dedrick. Explanation(s) and the patterning of basic colour words across languages and speakers. In N. J. Pitchford and C. P. Biggam, editors, *Progress in Colour Studies, Vol. II: Psychological Aspects*, pages 1–11. John Benjamins, Amsterdam/Philadelphia, 2006.

G. Drivonikou, A. Clifford, A. Franklin, and E. Özgen. Category training affects colour discrimination but only in the right visual field. In C. P. Biggam, C. A. Hough, C. J. Kay, and D. R. Simmons, editors, *New Directions in Colour Studies*, pages 251–264. John Benjamins Publishing Company, Amsterdam/Philadelphia, 2011.

D. Dubois. Cultural beliefs as nontrivial constraints on categorization: Evidence from colors and odors. *Behavioral and Brain Sciences*, 20:188, 1997.

M. D. Fairchild. *Color Appearance Models*. The Wiley-IS&T Series in Imaging Science and Technology. Wiley, 2005.

J. Foss. Mad about hue. *Behavioral and Brain Sciences*, 20:189, 1997.

A. Franklin and I. R. L. Davies. New evidence for infant colour categories. *British Journal of Developmental Psychology*, 22:349–377, 2004.

A. Franklin and I. R. L. Davies. Converging evidence for pre-linguistic colour categorization. In N. J. Pitchford and C. P. Biggam, editors, *Progress in Colour Studies, Vol. II: Psychological Aspects*, pages 101–119. John Benjamins, Amsterdam/Philadelphia, 2006.

A. Franklin, A. Clifford, E. Williamson, and I. R. L. Davies. Color term knowledge does not affect categorical perception of color in toddlers. *Journal of experimental child psychology*, 90:114–141, 2005a.

A. Franklin, M. Pilling, and I. R. L. Davies. The nature of infant colour categorisation: Evidence from eye-movements on a target detection task. *Journal of Experimental Child Psychology*, 22:349–377, 2005b.

A. Franklin, G. V. Drivonikou, L. Bevis, I. R. L. Davies, P. Kay, and T. Regier. Categorical perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults. *Proceedings of the National Academy of Sciences*, 105:3221–3225, 2008a.

A. Franklin, G. V. Drivonikou, A. Clifford, P. Kay, T. Regier, and I. R. L. Davies. Lateralization of categorical perception of color changes with color term acquisition. *Proceedings of the National Academy of Sciences*, 105:18221–18225, 2008b.

A. Franklin, N. Pitchford, L. Hart, I. R. L. Davies, S. Clausse, and S. Jennings. Salience of primary and secondary colours in infancy. *British Journal of Developmental Psychology*, 26:471–483, 2008c.

A. Franklin, O. Wright, and I. R. L. Davies. What can we learn from toddlers about categorical perception of color? Comments on Goldstein, Davidoff and Roberson. *Journal of Experimental Child Psychology*, 102:239–245, 2009.

J. Gage. Colour and culture. In T. Lamb and J. Bourriau, editors, *Colour: Art and Science*, pages 175–193. Press Syndicate of the University of Cambridge, 1995.

J. Gage. *Colour and meaning: Art, science and symbolism.* Thames and Hudson, 1999.

C. L. Hardin. *Color for philosophers: Unweaving the rainbow.* Hackett, 1988.

C. L. Hardin. Color order systems: A guide for the perplexed. *Behavioral and Brain Sciences*, 20:190–191, 1997.

C. L. Hardin. Explaining basic color categories. *Cross-Cultural Research*, 39:72–87, 2005.

S. Harnad. Psychophysical and cognitive aspects of categorical perception: A critical overview. In S. Harnad, editor, *Categorical perception: the groundwork of cognition*, pages 535–565. Cambridge University Press, Cambridge, 1987.

E. Hering. *Zur Lehre vom Lichtsinne: sechs Mitteilungen an die Kaiserl. Akademie der Wissenschaften in Wien.* Carl Gerolds Sohn, Wien, 1878.

E. Hering. *Outlines of a theory of the light sense.* Harvard University Press, 1964. Translated by L. M. Hurvich and D. Jameson.

L. M. Hurvich and D. Jameson. An opponent-process theory of color vision. *Psychological Review*, 64:384–404, 1957.

C. R. Ingling. Constraints on the definitions of "unique hues" and "opponent channels". *Behavioral and Brain Sciences*, 20:194–195, 1997.

G. Jäger. Natural color categories are convex sets. In M. Aloni, H. Bastiaanse, T. de Jager, and K. Schulz, editors, *Logic, Language and Meaning - 17th Amsterdam Colloquium, Amsterdam, The Netherlands, December 16-18, 2009, Revised Selected Papers*, pages 11–20, 2009.

G. Jäger and R. van Rooij. Language structure: psychological and social constraints. *Synthese*, 159:99–130, 2007.

K. Jameson. What Saunders and van Brakel chose to ignore in color and cognition research. *Behavioral and Brain Sciences*, 20:195–196, 1997.

K. Jameson. Where in the World Color Survey is the support for the Hering primaries as the basis for color categorization? In J. D. Cohena and M. Matthen, editors, *Color Ontology and Color Science*, pages 179–202. MIT Press, Cambridge, Massachusetts, 2010.

K. Jameson and R. G. D'Andrade. It's not really red, green, yellow, blue: an inquiry into perceptual color space. In C. L. Hardin and L. Maffi, editors, *Color Categories in Thought and Language*, pages 295–319. Cambridge University Press, 1997.

K. J. Jameson and N. L. Komarova. Evolutionary models of color categorization. I. Population categorization systems based on normal and dichromat observers. *J. Opt. Soc. Am.*, 26:1414–1423, 2009a.

K. J. Jameson and N. L. Komarova. Evolutionary models of color categorization. II. Realistic observer models and population heterogeneity. *J. Opt. Soc. Am.*, 26:1424–1436, 2009b.

Y. Jraissati. Categorical perception of color: assessing the role of language. *Croatian Journal of Psychology*, 12:439–462, 2012.

P. Kay. Synchronic variability and diachronic change in basic color terms. *Language in Society*, 4:257–270, 1975.

P. Kay. Color categories are not arbitrary. *Cross Cultural Research*, 39:39–55, 2005.

P. Kay and B. Berlin. Science $\neq$ imperialism: There are nontrivial constraints on color naming. *Behavioral and Brain Sciences*, 20:196–201, 1997.

P. Kay and R. Cook. The World Color Survey (encyclopedia chapter). In R. Luo, editor, *Encyclopedia of Color Science and Technology*. Springer, in press.

P. Kay and L. Maffi. Color appearance and the emergence and evolution of basic color lexicons. *American Anthropologist*, 101:743–760, 1999.

P. Kay and C. K. McDaniel. The linguistic significance of the meanings of basic color terms. *Language*, 54:610–646, 1978.

P. Kay and T. Regier. Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences*, 100:9085–9089, 2003.

P. Kay, B. Berlin, and W. R. Merrifield. Biocultural implications of systems of color naming. *Journal of Linguistic Anthropology*, 1:12–25, 1991.

P. Kay, B. Berlin, L. Maffi, and W. Merrifield. Color naming across languages. In C. L. Hardin and L. Maffi, editors, *Color categories in thought and language*. Cambridge, 1997.

P. Kay, B. Berlin, L. Maffi, W. R. Merrifield, and R. Cook. *The World Color Survey*. Center for the Study of Language and Information, Stanford, 2009.

J. Krauskopf, D. R. Williams, M. B. Mandler, and A. M. Brown. Higher order color mechanisms. *Vision Research*, 26:23–32, 1986.

R. G. Kuehni. CIEDE2000: Milestone, or final answer? *COLOR research and application*, 27:126–127, 2002.

E. H. Lenneberg and J. Roberts. The language of experience: A study in methodology. In *Indiana University Publications in Anthropology and Linguistics, Memoir 13*. Waverly Press, Baltimore, 1956.

A. Logvinenko. A theory of unique hues and colour categories in the human colour vision. *Color Research and Application*, 37:109–116, 2012.

V. Loreto, A. Mukherjee, and F. Tria. On the origin of the hierarchy of color names. *Proceedings of the National Academy of Sciences*, 109:6819–6824, 2012.

J. A. Lucy. The linguistics of "color". In C. L. Hardin and L. Maffi, editors, *Color Categories in Thought and Language*, pages 320–346. Cambridge University Press, 1997.

M. R. Luo, G. Cui, and B. Rigg. Further comments on CIEDE2000. *COLOR research and application*, 27:127–128, 2002.

J. Lyons. Colour in language. In T. Lamb and J. Bourriau, editors, *Colour: Art and Science*, pages 175–193. Press Syndicate of the University of Cambridge, 1995.

B. MacEvoy. http://www.handprint.com. Approached 24-04-2013.

R. E. MacLaury. *Color and Cognition in Mesoamerica: Constructing Categories as Vantages*. University of Texas Press, 1997a.

R. E. MacLaury. Ethnographic evidence of unique hues and elemental colors. *Behavioral and Brain Sciences*, 20:202–203, 1997b.

R. E. MacLaury, G. V. Paramei, and D. Dedrick, editors. *Anthropology of Colour: Interdisciplinary Multilevel Modeling*. John Benjamins Publishing Company, Amsterdam/Philadelphia, 2007.

N. L. Malcolm. Grammars rule O.K. *Behavioral and Brain Sciences*, 22:723–724, 1999.

D. Marconi. *Lexical Competence*. MIT Press, Cambridge, Massachusetts, USA, 1997.

R. J. Mausfeld. Why bother about opponency? Our theoretical ideas on elementary colour coding have changed our language of experience. *Behavioral and Brain Sciences*, 20:2003, 1997.

I. C. McManus. A monochrome view of colour. *Behavioral and Brain Sciences*, 20:204, 1997.

M. Merleau-Ponty. *Phenomenology of Perception*. Taylor & Francis e-Library, 2005.

D. L. Miller. Over the rainbow: The classification of unique hues. *Behavioral and Brain Sciences*, 22:204–205, 1999.

J. Mollon. Seeing colour. In T. Lamb and J. Bourriau, editors, *Colour: Art and Science*, pages 127–150. Press Syndicate of the University of Cambridge, 1995.

J. Mollon. A neural basis for unique hues. *Current Biology*, 19:R441–R442, 2009.

E. Özgen and I. R. L. Davies. Acquisition of categorical colour perception: A perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology: General*, 131:477–493, 2002.

O. Ozturk, S. Shayan, U. Liszkowski, and A. Majid. Language is not necessary for color categories. *Developmental Science*, 16:111–115, 2013.

A. Panorgias, J. J. Kulikowski, N. R. A. Perry, D. J. McKeefry, and I. J. Murray. Naming versus matching and the stability of unique hues. *Ophtalmic and Physiological Optics*, 30:553–559, 2010.

Y. H. Poortinga and F. J. R. van de Vijver. Is there no cross-cultural evidence in colour categories of psychological laws, only of cultural rules? *Behavioral and Brain Sciences*, 20:205–206, 1997.

T. Regier, P. Kay, and R. S. Cook. Focal colors are universal after all. *Proceedings of the National Academy of Sciences*, 102:8386–8391, 2005.

T. Regier, P. Kay, and N. Khetarpal. Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences*, 104:1436–1441, 2007.

T. Regier, P. Kay, and N. Khetarpal. Color naming and the shape of color space. *Language*, 85:884–892, 2009.

D. Roberson, I. Davies, and J. Davidoff. Color categories are not universal: Replications and evidence form a stone-age culture. *Journal of Experimental Psychology: General*, 126:369–398, 2000.

D. Roberson, J. Davidoff, I. R. Davies, and L. R. Shapiro. Color categories: evidence for the cultural relativity hypothesis. *Cognitive psychology*, 50:378–411, 2005.

E. [Heider] Rosch. Universals in color naming and memory. *Journal of Experimental Psychology*, 93:10–20, 1972a.

E. [Heider] Rosch. Probabilities, sampling, and ethnographic method: The case of Dani colour names. *Man*, 7:448–466, 1972b.

E. [Heider] Rosch. On the internal structure of perceptual and semantic categories. In T. E. Moore, editor, *Cognitive Development and the Acquisition of Language*. Academic Press, New York, 1973.

B. Saunders. Revisiting *Basic color terms*. *The Journal of the Royal Anthropological Institute*, 6:81–99, 2000.

B. Saunders. The normativity of colour. In C. P. Biggam and C. J. Kay, editors, *Progress in Colour Studies, Vol. I: Language and Culture*, pages 89–100. John Benjamins, Amsterdam/Philadelphia, 2006.

B. Saunders. Towards a new topology of color. In R. E. MacLaury, G. V. Paramei, and D. Dedrick, editors, *Anthropology of Colour: Interdisciplinary Multilevel Modeling*, pages 467–479. John Benjamins Publishing Company, Amsterdam/Philadelphia, 2007.

B. Saunders and J. van Brakel. Are there nontrivial constraints on colour categorization? *Behavioral and Brain Sciences*, 20:167–228, 1997a.

B. Saunders and J. van Brakel. Colour: An exosomatic organ? *Behavioral and Brain Sciences*, 20:212–220, 1997b.

B. Saunders and J. van Brakel. Colour word trouble. *Behavioral and Brain Sciences*, 22:725–729, 1999.

W. Sellars. *Empiricism and the Philosophy of Mind*. MIT Press, Cambridge, Massachusetts, USA, 1997.

G. Sharma, W. Wu, and E. N. Dalal. The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *COLOR research and application*, 30:21–30, 2005.

C. Simpson. The irrelevance of the psychophysical argument. *Behavioral and Brain Sciences*, 20:207, 1997.

E. N. Sokolov. Four-dimensional color space. *Behavioral and Brain Sciences*, 20: 207–208, 1997.

L. Steels and T. Belpaeme. Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Sciences*, 28:469–529, 2005.

C. E. Sternheim and R. M. Boynton. Uniqueness of perceived hues investigated with a continuous judgmental technique. *Journal of Experimental Psychology*, 72:770–776, 1966.

C. M. Stoughton and B. R. Conway. Neural basis for unique hues. *Current Biology*, 18:R698–R699, 2008.

A. Valberg. Unique hues: an old problem for a new generation. *Vision Research*, 41:1645–1657, 2001.

R. L. De Valois and K. K. De Valois. Neural coding of color. In E. C. Carterette and M. P. Friedman, editors, *Handbook of perception*. Academic Press, 1975.

R. L. De Valois, I. Abramov, and G. H. Jacobs. Analysis of response patterns of LGN cells. *Journal of the Optical Society of America*, 56:966–977, 1966.

N. W. H. van Kruysbergen, A. M. T. Bosman, and C. de Weert. Universal colour perception versus contingent colour naming: A paradox? *Behavioral and Brain Sciences*, 20:209–210, 1997.

D. van Laar. Ekphrasis in colour categorisation: Time for research, or time for revolution? *Behavioral and Brain Sciences*, 20:210, 1997.

J. S. Werner and M. L. Bieber. Hue opponency: A constraint on colour categorization known from experience and experiment. *Behavioral and Brain Sciences*, 20:210–211, 1997.

A. Wierzbicka. *Semantics: Primes and Universals*. Oxford University Press, UK, 1996.

A. Wierzbicka. "Universals of colour" from a linguistic point of view. *Behavioral and Brain Sciences*, 22:724–725, 1999.

A. Wierzbicka. The semantics of colour: A new paradigm. In C. P. Biggam and C. J. Kay, editors, *Progress in Colour Studies, Vol. I: Language and Culture*, pages 1–24. John Benjamins, Amsterdam/Philadelphia, 2006.

A. Wierzbicka. Correspondence. *Journal of the Royal Anthropological Institute*, 14:887–889, 2008.

L. Wittgenstein. *Philosophical investigations*. Blackwell, Oxford, 1967.

L. Wittgenstein. *Remarks on colour - Bemerkungen Über die Farben*. Blackwell, Oxford, 1977.

S. M. Wuerger and L. Parkes. Unique hues: Perception and brain imaging. In C. P. Biggam, C. A. Hough, C. J. Kay, and D. R. Simmons, editors, *New Directions in Colour Studies*, pages 445–455. John Benjamins Publishing Company, Amsterdam/Philadelphia, 2011.

K. Xiao, S. Wuerger, Ch. Fu, and D. Karatzas. Unique hue data for colour appearance models. Part I: Loci of unique hues and hue uniformity. *Color Research and Application*, 36:316–323, 2011.

Y. Xue. Uniform color spaces based on CIECAM02 and IPT color difference equations, 2009. Master thesis, Rochester Institute of Technology.

S. Yendrikhovskij. A computational model of colour categorization. *Color research and application*, 26:S235–S238, 2001.