# Explanation and Color Naming Research

Don Dedrick
Department of Philosophy
Department of Psychology
University of Guelph,
Guelph, Ontario, Canada
E-Mail: ddedrick@uoguelph.ca

KEY WORDS: color naming theory; philosophy of science; Actual Sequence Explanations; Robust Process Explanations.

## Abstract

Perhaps one of the central assumptions when one comes to think about scientific explanations-an assumption made by philosophers and scientists alike-is that a causal explanation is an optimal explanation. It seems, after all, that an explanation tells us why something happens, and that to do so is to specify causes. While there is nothing wrong with causal explanation *per se*, many good explanations in science are not in any important sense causal. What I mean by this is that many good explanations in science are compatible with a variety of causal mechanisms and, as such, ignore the details of such mechanisms. I develop this claim in the discussion of color naming research which follows, where I distinguish between explanation types that are (more) close to cauality (Actual Sequence Explanations) and those that are (more) removed from causal details (Robust Process Explanations).

## 1. Introduction

Certain concepts in recent philosophy of science can help us to understand some disagreements and difficulties that arise in research into the universality of color names and into the explanation for such universality. My purpose is to show that such differences can be understood in a novel way via a consideration of the nature of explanation in science.

Perhaps one of the central assumptions when one comes to think about scientific explanations—an assumption made by philosophers and scientists alike—is that a causal explanation is an optimal explanation.[1] It seems, after all, that an explanation tells us why something happens, and that to do so is to specify causes. While there is nothing wrong with causal explanation *per se*, many good explanations in science are not in any important sense causal. What I mean by this is that many good explanations in science are compatible with a variety of causal mechanisms and, as such, ignore the details of such mechanisms. I develop this claim in the discussion of color naming research which follows, where I distinguish between explanation types that are (more) close to cauality (Actual Sequence Explanations) and those that are (more) removed from causal details (Robust Process Explanations).

I will be viewing things from an abstract perspective—that of explanation-types in science. In keeping with this high level of generality it will be useful to have a general characterization of the domain of interest. I will refer to the relevant research, a truly multidisciplinary mixture of results and methods and commitments, as "Color Naming Research" (CNR). So specified, the body of research includes all relevant research.[2] Within CNR there is a set of results that have been linked in such a way as to make a case

for the universality of color naming (Berlin & Kay 1969; Kay & McDaniel 1978; Kay, Berlin & Merrifield 1991; Kay, Berlin, Maffi & Merrifield 1997; Kay and Regier 2003; Hardin 1988; Dedrick 1998a). This I shall call, as I have elsewhere (Dedrick 1998a; 1998b), the "Universalist Tradition in Color Naming Research" (UT). For the purpose at hand, the UT need only subscribe to the following minimal claims: there are physiological and/or psychological constraints on the categorization and naming of colors, within languages, within cultures. Also within CNR one finds work that is critical of the UT. This criticism takes many forms: from obscurantist post-modernism that promotes a naïve and untenable criticism of the UT (e.g. Saunders and van Brakel 1997, 2001), to work that is critical of particular models of color naming (e.g. MacLaury 1997), to detailed ethnographic and historical studies that raise questions as to the fit between the most general color naming theories and color naming practices "on the ground" (e.g. Stanlaw 1997). I shall refer to this critical work—which is clearly a significant part of CNR—as the "Critical Tradition" in color naming research. While there can be no absolute division into universalist and critical traditions—criticism plays a role in any scientific research program—there is a reason for marking the distinction. We are unclear as to how to integrate studies and results that appear *prima facie* at odds with one another. This paper offers an articulation of this problem, and points toward one sort of approach to the problem.

**2. Different kinds of explanation**

With the first set of preliminaries out of the way, I need to introduce a second set of basic concepts. These originate in the philosophy of science. The core distinction I am

going to use is between two different though complementary types of scientific explanation: *Actual Sequence Explanations* and *Robust Process Explanations*. This terminology has been developed by two philosophers of biology, Kim Sterelny and Paul Griffiths (1999), though the distinction exemplified originates in work by the philosophers Frank Jackson and Philip Pettit (1992), writing about general issues in social science. I will follow Sterelny and Griffiths' articulation (1999, p.84), with some modification.

[1] *Actual Sequence Explanations* seek to explain the nuances of the causal history of the world we find ourselves in. They explain the contrasts between actual histories of interest, as well as related histories. Actual Sequence Explanations aim to tell us what actually happened or happens in some restricted domain of a population.

[2] *Robust process explanations* (Robust Process Explanations) reveal the insensitivity of a particular state of affairs to some feature of its actual history, aiming at generalizable features that may be invisible to actual sequence explanations. Robust Process Explanations aim to capture robust regularities that pertain to populations.

Sterelny and Griffiths are, as mentioned, philosophers of biology, and we can draw an example from biology to illustrate the point of this distinction. Questions as to whether some phenotypic trait is an adaptation are of interest to evolutionary biologists yet it is not clear that any trait is an adaptation for every member of a species. Sterelny and Griffiths provide a striking example. (1999, p. 84) Albinism is typically deleterious in nature. And yet there are zoos which encourage the flourishing of albino animals, due to their rarity. In the ecology of the zoo, albinism is an adaptation. This may seem a far-fetched example—far from the natural world—but it illustrates the general point quite

well. In so far as the environments that a type of animal inhabits vary, selective pressures vary. This being so, it makes no sense to say that a given trait is or is not an adaptation in some absolute sense.[3]

The distinction between populations and individuals (and other sub-population groups) is crucial to evolutionary biology. Critics have argued, for example, that natural selection based accounts of evolution—Darwinist accounts, I shall call them—are unfalsifiable. Since the fittest organisms survive, any organism that survives is fitter than any which does not. If you survive, you are fitter; if you are fitter you survive. This is said to be a tautological truth which makes Darwinism unfalsifiable and hence trivial (Koestler 1978; Hitching 1982). In fact, Darwinism does not subscribe to the fitter = survivor equation. The view, rather, is that fitter organisms are more likely to survive. This subtle difference makes the principle of natural selection an empirical principle (we can look to see if, statistically, organisms which possess a trait tend to survive). It also helps to clarify the distinction between Actual Sequence Explanations and Robust Process Explanations..

No one thinks that the existence of albino tigers in zoos "falsifies" the claim that albinism is deleterious for tigers. Why not? If fitness is defined classically in terms of number of offspring (a similar argument can be advanced for inclusive fitness), we have a counterexample. Yet, as I say, no one takes this proposed refutation seriously. One reason this is so has got to do with the largely statistical nature of explanation in the life and behavioral sciences. According to this line of thought, it is unreasonable to expect unbroken regularities of the sort one might expect in physics. And anyway, researchers in

the life and behavioral sciences do not consider "laws" to be a desideratum for their theories.

While I agree with this standard argument, having advanced it myself with respect to certain critics of UT (Dedrick 1998a), there is a clear sense in which it will not do. We are, that is, owed an explanation for why many such theories have the statistical character that they do.

**3. Contingency**

Let us consider CNR. This research isolates various components that are widely believed by CNR researchers to be relevant to the explanation of color naming among humans. The set of components, which is neither static nor clearly demarcated, includes: human genetic structure, relatively large-scale properties of brains (and, of course, small-scale properties as well), functional models of color vision, color discrimination (as tested via a number of experimental paradigms ranging from infant habituation to various color matching tasks), cognitive-psychological regularities (ranging from prototypicality to categorical perception), social-psychological regularities, ethnographic facts, linguistic facts. In some cases, color naming researchers develop accounts of individual components and then try to show how those components hook up with one another. UT models which view many such components as ordered in a hierarchy with explanatory links between "levels" offer one such account (see Dedrick 1998, Section I for a discussion). *No matter how one specifies the relevant components*, they interact with other components and this interaction is far too complex to be captured in a single model. There is too much *contingency* in the interactions, ranging from contingency in the

genetically developmental environment, to contingency in the life-cycle of the individual, to social, ecological, and historical contingencies in the group a given individual belongs to. None of this is mysterious—the presence of a color term for "green or yellow" might be explained the following way: there is a lot of greenish-yellow in a specific environment—the range of color is ecologically salient![4] If it is so explained, there will be no mystery as to why other relevant components—a functional model of color vision, for example—are not explanatory of the existence for such a term *in some given language*. We cannot predict what sorts of contingencies count, and in what ways, for particular cases. In this sense then, Actual Sequence Explanations stand in very uncertain relationship to Robust Process Explanations. While Actual Sequence Explanations inform us of the actual contingencies that matter in specific cases, Robust Process Explanations are not intended to capture the nuances relevant to a particular language or speaker and it is not, therefore, a failure when they do not do so (as it is not a failure when the evolutionary biologist's population-level claims about albinism fail for the inhabitants of zoos). We have given a reason why this is so. It's not just because of some general methodological commitment to statistical methods. The methodological commitment is in fact grounded in the way that the world actually is, i.e., contingent with respect to the interaction of the many components relevant to CNR (and the albino tiger's continued existence). With respect to the distinction between Actual Sequence Explanations and Robust Process Explanations, it is the former that will capture contingent facts about color vocabularies most precisely. It is the latter that capture population level facts.

**4. Compatibility**

A crucial point, noted by Sterelny and Griffiths, is that Actual Sequence Explanations and Robust Process Explanations are not in competition with one another. What is meant, when saying this, is that the specificity of Actual Sequence Explanations is, purposefully, not a feature of Robust Process Explanations. When it comes to CNR, this claim is most relevant to the following style of argument: Because color names may have particular histories and origins, because color terms have connotations specific to those histories and origins, it is an error to view them as a result of a universal process (e.g. McNeill 1972; Ratner 1989; Simpson 1991; Lucy 1997; Roberson, Davies and Davidoff 2000).There are a number of problems with claims like this, but I focus on just one. This argument assumes that because there is an Actual Sequence Explanation for some set of color words, a Robust Process Explanation is irrelevant or even falsified by the Actual Sequence Explanations. This I propose, is a mistake. The sort of detail that is relevant to an Actual Sequence Explanations is simply not appropriate to a Robust Process Explanation. Histories can vary—perhaps on a number of dimensions—without such variance being inconsistent with robust claims.

In a paper critical of my own work on color naming, Saunders and van Brakel (2001) take up the question of color naming in Bellonese. In a well-known article on this subject the anthropologists Kushel and Monberg (1974) argue for two claims of interest here: (1) That the Bellonese have a sophisticated *color-relevant* vocabulary that is very different from the Western system of abstract color terms (terms that refer entirely to the chromatic properties of a stimulus and thus abstract away from non-chromatic dimensions of real objects in their ecological and social context). The Bellonese color-relevant words, we could say, are multi-modal. Their meaning cannot be reduced to the

designation of this or that abstracted color. (2) The Bellonese have three of Berlin and Kay's basic color terms, glossed as red, black, and white in English.

For Saunders and van Brakel, the second claim is suspicious and most likely to be the result of the Bellonese being "goaded/encouraged/forced" (2001, p. 545) to use the terms of interest to the anthropologists (who, themselves, had some doubts about the status of the basic terms in Bellonese). In commenting on Bellonese color relevant words (Dedrick 1998a; 1998b) I argued that these two claims are compatible. One could, in fact, make the same multi-modal claim about many other color-relevant vocabularies, given a suitably fine-grained analysis, such as the one we find in (Casson 1997) for English.

The current terminology provides a more satisfactory account of the difference between (1) and (2). The first claim is part of an Actual Sequence Explanation for Bellonese color relevant words. The second is part of a Robust Process Explanation designed to capture relevant, highly regimented regularities about color naming and it is, as I have said, "insensitive" to certain features of Bellonese, as it is insensitive to certain features of all particular cases. Indeed, there is no reason to think that the regularities relevant to the UT must even be visible from the "ground level" analysis of a language. Neither of these kind of explanation eliminates the other.

That there are robust regularities is admitted by Saunders and van Brakel, "Nobody disputes that the majority of languages contain a word that in some contexts can be translated as "red" and that this has something to do with what all ecological systems "humans-environment" have in common" (2001, p. 543). The authors do not think that these and other undisputable facts (as they call them) constitute the basis for any interesting science. My view is that they do, and that they are the objects of a Robust

Process Explanation. As for Kuschel and Monberg and their doubts as to the existence of basic color terms in Bellonese, such doubts are entirely appropriate for they mark the imprecise boundary between the different explanation types I have described.

One anthropologist who comments, implicitly, on the distinction between Actual Sequence Explanations and Robust Process Explanation is James Stanlaw (1997). In an article on color and culture-contact (Stanlaw is writing about loan words in Japanese) the author is, in many respects, critical of the UT. It is the detail of his account, its fine-grain, that gives his criticism its heft. Yet Stanlaw implicitly appreciates the distinction between his own careful analysis of Japanese, and the Robust Process Explanation that analysis is compatible with. He writes that there are

> … interesting universalist and particularist interactions when colors and cultures come into contact. General universalist properties—like the Berlin and Kay encoding sequence—are found for the Japanese data, but we also saw how specific social and linguistic situations—such as borrowing—modified them. That is, the universalist arguments of Berlin and Kay do not necessarily refute all Whorfian considerations under all conditions. Languages can certainly vary semantically, but obviously not without constraint; people cannot just call anything anything, after all. These constraints, however, are often a complex interface of both human cognitive universals, and the particulars of the cultures and languages in contact. It is on this edge that much of the linguistic and social interaction takes place. (Stanlaw 1997, 258-9)

Stanlaw talks about the "refutation" of claims within the UT. In this sense he is in the grip of a traditional characterization of scientific practice: A theory is like an argument which is refuted when a general proposition—or "law"—in such an argument is subjected to counterexample (Hempel 1966). And yet Stanlaw resists the proposed refutation he

himself considers. Instead, he refers to the complexity of the interaction between the particular and the universal. While I agree entirely with Stanlaw that it is on the edge between cognitive universals and particular cultures where most of the linguistic and social action takes place, there is another correlated edge we need to be aware of: that between the different kinds of explanation which are themselves interacting in the *account* of Japanese color nomenclature. Both are necessary to capture the interaction which is the object of these explanatory strategies: one is Stanlaw's highly detailed Actual Sequence Explanation for Japanese color words, the other a Robust Process Explanation grounded in the UT. This distinction between explanatory types gives us a good idea as to why UT is not refuted by the careful, fine-grained analysis: as noted, at the beginning of this section, Actual Sequence Explanations and Robust Process Explanations are not in competition.

**5. The relationship between Actual Sequence Explanations and Robust Process Explanations (1) : Contingency, compatibility and Jameson's argument for a cognitive version of UT**

Recent work by Kimberly Jameson and her colleagues provides an interesting example of how different explanations and explanation types interact. In work on differences of color experience among human subjects Jameson and colleagues (Jameson, Highnote, and Wasserman 1998, 2001; Jameson unpublished) have produced evidence illustrating that the well-known gender-linked genetic basis of color discrimination may give rise to observers who are functionally tetrachromatic. This research indicates that such subjects (identified genetically) may express retinal phenotypes which allow them to

distinguish more color bands within a spectral display than subjects identified as trichromatic (male or female). Thus the hypothesis of a "richer color experience" for the putative tetrachromatic subject. It is not my point to evaluate this research but to note one of Jameson's own concerns as to its implications. Suppose functional tetrachromacy exists, and this explains richer color experience for some subjects. Yet these subjects have typically not been identified using standard psychophysical techniques (and perhaps cannot be). And further—here is the point of interest to me—Jameson notes that such subjects—those with richer color experience— need not express such experience linguistically. They conform to social norms regarding color naming. As a consequence of this fact, Jameson suggests that universality in naming must be grounded somewhere else other than simply in color experience itself. Jameson and d'Andrade's (1997) Interpoint Distance Model (IDM) is an attempt to specify that grounding. Without going into the details of the IDM, I should like to provide a partial characterization of the reasoning that motivates the IDM, adapted to the current discussion.

> [1] There is Robust Process Explanation which connects the genetic basis for color discrimination with behavioral data concerning the ability to distinguish bands of color in a spectral display. Supposed tetrachromats (identified genetically) identify more colored bands than trichromats. Thus there is evidence for functional tetrachromacy which is insensitive to multiple (possible) genetic bases[5] for that behavioral capacity, a fact that can be determined only by Actual Sequence Explanations for individual genotypes.

[2] There is a Robust Process Explanation which identifies color naming regularities that are robust across languages and explains those regularities in terms of the nature and structure of visual experience. It is insensitive to various group differences, such as those revealed in particular ethnographic and/or experimental studies, and to individual differences revealed by Actual Sequence Explanations for particular subjects.

[3] There is a Robust Process Explanation which is in competition with [2] in the sense that it specifies a cognitive explanation (computation of psychological relational structure, or 'distances,' between color sample exemplars —the IDM) for the color naming regularities specified in [2]. This Robust Process Explanation is insensitive to group differences such as those revealed in particular ethnographic and/or experimental studies, and to individual differences at the level of visual experience.

For Jameson and her colleagues, Robust Process Explanation [2] is not optimal. Some of the differences that [2] is insensitive to in fact matter to an adequate explanation of color naming. Robust Process Explanation [2] makes the assumption that trichromatic vision is not only the statistical norm, but that it is properties of trichromatic vision that explain robust color naming practices across languages. [3], however, is to be preferred over [2]. If it is the case that varieties of color experience can generate the color naming regularities that both [2] and [3] acknowledge as given, then an account such as [2] which appeals to a normal form of visual experience will not be sufficient to explain the

regularities in cases where the normal form is absent. Thus [3] which proposes a different sort of process—interpoint distance computations indexing relational structure among exemplars—has the right sort of independence from visual experience: it is insensitive not only to differences between individuals (that we might construct Actual Sequence Explanations for) but to the robust properties of other models. The reason [2] fails—assume it does, for the sake of argument—is because it identifies the wrong set of robust properties, and this can be seen via consideration of [1], a Robust Process Explanation which is suitably independent from [3].

Jameson's argument proposes much more than a cognitive model for color naming. It requires us to sort the various Robust Process Explanations and Actual Sequence Explanations into those which count, those which do not count, and those which count in some restricted sense. Consider trichromacy. According to [2], above, it is a more or less uniform vision system that generates more or less uniform naming practices. Jameson does not deny this but, instead, reinterprets the claim: trichromacy establishes a *social* norm on the basis of its prevalence in typical populations. (Jameson unpublished) In this sense, trichromacy is a constraint on color naming, but one that, for at least some subjects (e.g. functional tetrachromats), is primarily social in nature rather than perceptual. Functional tetrachromats must learn, Jameson suggests, to interpret other (trichromatic) speakers' claims about color matches in such a way that they are compatible with their own judgments grounded, as they are, in a richer color experience. They must do this for reasons that have to do with the pragmatics of linguistic communication, a pragmatics that requires some adjustment between an individual's perceptual experience and his or her social model of color categorization.[6]

This example—the reasoning which leads to the IDM—illustrates the themes I have been discussing. Two themes have been proposed. The first is *contingency*: there is too much complexity in the interaction of the components specified in CNR for a model to capture all the relevant interactions. The IDM allows for there to be various actual differences at the level of individuals—and even groups of individuals (sub-population level Robust Process Explanations for, say, functional tetrachromats)—which, while they constrain color naming, may do so in various, causally incompatible ways (the trichromat makes different adjustments between the personal and the social cognition of color than does the tetrachromat). The IDM is in fact a Robust Process Explanation which allows a number of degrees of freedom when it comes to the generation of actual color categorization and naming so that, for example, cultural interests and information can account for some cross-cultural variance in the actual color category names expressed in a language. How much variance? One cannot answer this question in a general way (that's the point of contingency) though one may be able to answer it via the construction of more fine grained accounts (Actual Sequence Explanations for particular speakers; sub-population level Robust Process Explanations for particular groups).

The second theme is *compatibility*: Actual Sequence Explanations and Robust Process Explanation perform different though compatible tasks. There are numerous actual and possible Actual Sequence Explanations that are irrelevant to the robust process explanation proposed in the IDM (those which specify genetic equivalence classes for functional tetrachromacy; those which identify a given subject as trichromatic or tetrachromatic; those which identify culturally significant factors in a group's set of color category names).

## 6. The relationship between Actual Sequence Explanations and Robust Process Explanations (2) : Conflating Robust Process Explanations and Actual Sequence Explanations

In a remarkably detailed and theoretically sophisticated work in CNR, the cognitive anthropologist Robert MacLaury attempts to develop a strong version of interaction (between the Robust Process Explanation and the Actual Sequence Explanation strategies). For MacLaury, and this is my characterization,[7] The Robust Process Explanation and the Actual Sequence Explanation merge. By this I mean that MacLaury seeks an explanation for human color naming that is capable of capturing all the involved judgments. One way to think of this project is as follows: MacLaury takes the color naming practices of every individual he studies to exemplify a property worthy of a unified explanation. In the ideal case this would lead to an account that can incorporate any judgment that any individual makes under any circumstance. MacLaury's theory of color categorization is based on his "vantage theory," developed in detail in (MacLaury 1997). Of this theory he writes that

> Although vantage theory may not offer the final word on the findings, any account that superseded it must explain all the observations more cogently and economically and should accommodate at least a few more that no other theory can address. (1997, 392-3)

What are "all the observations"? Minimally, they are constituted by a list of 100 statements MacLaury provides in his "Appendix VII" (449-457). The list contains statements that are not observations at all, but universal generalizations. Two examples:

14. Foci never "float." They show patterned differences between categories, stages of evolution, and languages.

20. The warm category divides before the cool category.

Claims like this are exactly what one would expect to find in a Robust Process Explanation. But there are other claims, much more particular. Two examples:

17. An individual who focused on an elemental hue while at home focused between unique hues while away from home and during analytical training.

82. In some languages that focus the cool category in green and blue, such as Tzeltal, no dual foci were volunteered after many interviews, 52 for Tzeltal. In such languages, rare individuals canceled a focus on one hue while choosing a final focus on an opposite hue.

These are, admittedly, extreme cases by the exemplified standards of MacLaury's list. But consider the following:

15. Most individuals add to a mapping when asked to include more colors of X-name.

83. In some cool categories, the naming range of a single term may be skewed toward one hue and its mapping skewed toward the other. The focus may land anywhere.

Many of the statements in MacLaury's inventory are like this. Thus his set of "observations," contains, among other things, universal generalizations, claims specific to a single speaker or a small group of speakers, unquantified statistical claims.

In his attempt to model color naming via vantage theory MacLaury has conflated the highly specific virtues of Actual Sequence Explanations with the "population thinking" of Robust Process Explanation. Is it really the case that a theory of color categorization/naming ought to account for one individual's naming behavior as exhibited in the morning versus the afternoon? If we need to

explain why "most individuals" will name further samples when asked, do we also need to explain why some do not? Is MacLaury's theory (or the Robust Process Explanation which is a part of it) "falsified" if empirical research discovers a case where the cool category divides before the warm? I, for one, say no. MacLaury, I propose, ought to differentiate between the Robust Process Explanation his data support (an interesting Robust Process Explanation indeed) and the actual histories that his careful fieldwork uncovers. Some of the "observations" he cites are not just grist for a Robust Process Explanation they form part of a Robust Process Explanation(the universal generalizations). Other observations belong to Actual Sequence Explanations for specific speakers, or Robust Process Explanations for specific cultures. Other of his claims, the unquantified statistical claims, can be understood as measuring the extent to which a Robust Process Explanation—a model—applies to or "fits" the world.

On this reading of MacLaury's project it is a distinct part of that project to assign the various relevant claims (the "observations" as he calls them) to these different roles. Do they count towards the robust explanation, which in MacLaury's case would include general claims about the transition from brightness classification to hue classification? (MacLaury 1992; 1997) Are they culture-centric claims due to the particulars of a culture and language? Part of the point of calling them *particulars* is to suggest that they are not, or need not be, robust features captured by a general model. Finally, the unquantified statistical claims. I should say that I am not objecting to the lack of quantification which can, of course, be done. The point is that their statistical nature can be taken to

indicate a number of interesting features of the explanatory project as a whole and, as noted in the last section, they may reveal important facts about the structure of the population of interest. They might, for instance, indicate particularities that are properly aspects of more fine-grained Actual Sequence Explanations. They might, as well, indicate some general lack of robustness in the Robust Process Explanation. Let me juxtapose MacLaury's ideas with those of Jameson and the IDM, discussed in the last section.

In her work on tetrachromacy and color naming, Jameson argues that the population structure of humans is such that we cannot ignore functional differences in color perception if we are to get a grip on the basis (i.e. explanation) for universality. One might dispute this claim, but it is worth pointing out that Jameson makes a rough attempt to quantify her claim that the range of variation is significant enough to be a factor that cannot be ignored when it comes to the explanation for universality (Jameson unpublished). MacLaury, on the other hand, simply presents us with the fact that there are, for example, color naming anomalies (for a single speaker) that pertain to time of day. I think we need an argument that time-of-day, say, is likely to be robust as opposed to one of many contingent factors that are operating on any occasion a speaker of a language is asked to name her colors. To put this in terms of my preferred nomenclature: we ought not to expect our robust explanations to model the exacting and contingent facts of actual token causal processes–these being the target of Actual Sequence Explanations.

I hope that the reader is clear that I am not so much criticizing MacLaury as suggesting the constraints on his model—any model—are less onerous than he imagines. Indeed, I suggest that, because of contingency, a good model should not have as a desideratum, precise judgments for all actual cases (there is, for example, no such model in physics for all physical systems of a given type[8]). Rather, it should tell us how, in a general way, people are able to make the kinds of color judgments that they do. In providing such an account, one is forced to ask questions about, and to make decisions about, what is to count towards the adjustment of a Robust Process Explanationas it applies to a population (in the case at hand, a set of languages, speakers, cultures) and what that Robust Process Explanation is *willing to be insensitive to*. Of course, scientists do make such decisions. I suggest—and MacLaury's work seems a good example here—that a distinction between different explanatory types is a useful tool when it comes to deciding what a given observation is to count for—whether it is to count at all.[9]

**7. The relationship between Actual Sequence Explanations and Robust Process Explanations (3) : color categorization is (or is not) universal?**

According to the Roberson, Davies, and Davidoff (2000), color categories are not universal. Indeed, "Color categories are not universal" is the title of a recent paper by those authors in the *Journal of Experimental Psychology* (Roberson, *et al* 2000). Kay and Reiger (2003) appear to disagree, for while they are cognizant of the Roberson paper they state their own recent conclusions as follows:

> (1) There are clear cross-linguistic statistical tendencies for named color categories to cluster at certain privileged points in perceptual color space; (2) these privileged points are similar for the unwritten languages of non-industrialized communities and the written languages of industrialized societies; and (3) these privileged points tend to lie near, although not always at, those colors named "red", "yellow", "green", "blue", "purple", "brown", "orange", "pink", "black", "white", and "gray", in English. (Kay and Reiger 2003, p. 9089)

Since these claims pretty much recapitulate the original universalist findings of Berlin and Kay (1969), we have to wonder as to the basis for the radically different assertions of Roberson and her colleagues, and those of Kay and his. It would seem that at least one of them must be mistaken. Color categories either are or are not universal, one would think.

These authors are, as we might expect, talking about different things. Kay and Regier employ a statistical analysis that has the effect of concealing a great deal of variation in the population in order to determine central tendencies in color naming. Roberson is concerned to replicate the results of Eleanor Rosch (Heider) on color prototypicality (Heider 1972; Heider and Olivier 1972). The different claims can be put as follows:

(A) Kay and Regier: Robust Process Explanation: A very robust, very abstract claim about the statistical tendency of named color categories to cluster in certain predictable areas of color space. The claim: universality explains clustering. This Robust Process Explanation is insensitive to a great deal of individual variation (Reiger, personal communication).

(B) Roberson, *et al* : Actual Sequence Explanation: a language specific claim[10] about the cognition of color. Berinmo subjects do not conform to

expectations, based on the work of Rosch, that color prototypes will be salient independent of language. The claim -- the denial of a Robust Process Explanation: color prototypes are not psychologically salient in the ways that Rosch proposed.

The Actual Sequence Explanation for Berinmo can hardly support a claim to the effect that color categories are not universal, whatever it shows about Berinmo. In a second article Kay (2005) argues that when the naming data collected by Roberson is subjected to the same statistical analysis as their own World Color Survey data, Berinmo actually conforms to and supports the universalist claim! While debate about the methodological assumptions of Kay and Regier are likely to ensue (having been a regular feature of CNR from the beginning, e.g. Hickerson 1971 ) it is worth pointing out that the universalist claim could hardly be thought to stand or fall on one set of experiments, for one language. So what was Roberson thinking when she claimed that color categories are not universal? In order to answer this question we need to look more closely at Roberson's study.

The work in question was explicitly an attempt to replicate Rosch's work on color prototypicality. As is well known, Rosch argued that linguistic salience of certain color words was influenced by the psychological salience of certain prototypical colors, a salience which expressed itself in a variety of tasks (memory, recognition/recall, categorization). Rosch's work provided, for many in CNR, the best *explanation* for the universality observed by Berlin and Kay: the basic color terms (see the quotation at the beginning of this section for a list) are

linguistically salient, because there are prototypical colors that are psychologically salient and which provide a perceptual-cognitive basis for basic color term development. Rosch's work was explicitly cross cultural, comparing the performance of a non-western, aboriginal people (the Dani of Indonesian New-Guinea) with that of North-American English speakers.

Roberson, *et al*, did not study the same language/speakers as Rosch. Nonetheless, the color vocabulary of their non-western informants, the Berinmo, was significantly different from English. They found that virtually none of Rosch's results could be replicated and, of special interest, discovered that there was no cognitive advantage that attached to focal/prototypical colors. Subjects were not, for example, better on a recall task if that task involved prototypical as opposed to non-prototypical colors. Categorical effects that Rosch uncovered and which implicated focal colors as psychologically fundamental were also absent—replaced, in fact, by results more conducive to the relativist view of color categorization that preceded Rosch's ground-breaking work. (Brown and Lenneberg 1954).

It turns out, then, that Roberson's claims do not challenge universality (despite the title of her paper!) so much as the *explanation* for universality. From the fact that there is a universal tendency for named color categories to cluster in certain regions of color space, it does not follow that any specific explanation for that clustering is correct. On the other hand, Roberson and her colleagues give us reason to think that at least one explanation will be problematical. For even if it is the case that Berinmo actually conforms to the universalist statistical claims, as

Kay and Regier assert, it cannot be the case that focality explains Berinmo color nomenclature, in the absence of any significant differentiation in the cognitive salience of focal colors from nonfocals. Kay and Reiger do, in fact, suggest (though they do not assert) that "certain privileged points in color space appear to anchor the color naming systems of the world's languages" (2003, p. 9089). It is difficult to treat this claim as anything other than a metaphor, in the absence of psychological salience attaching to these points.[11] Thus, while claims (A) and (B) turn out to be compatible (as we should expect, given that the former is an Actual Sequence Explanation, the latter a Robust Process Explanation) there is an interesting and I think, open question as to the nature of the explanation for color name universals. Berinmo (and possibly other languages) will need an explanation for their color terms—and it will not be in terms of the "priveliged points" mentioned by Kay and Reiger.

## 8. Conclusion

Explanations in science perform different functions. In some instances they provide detailed information about more or less specific causal processs. On other ocasions they identify robust properties that are relatively insensitive to fine-grained differences in a population. I have argued that one ought not expect all Actual Sequence Explanations to be special cases of some given Robust Process Explanation. Contingency insures that this will often not be the case. Indeed, it is crucial that we view Actual Sequence Explanations as essentially compatible with the robust claims of the Robust Process Explanation we employ—compatible in

the sense that they provide different sorts of information about the structure of a population of interest. On occasion, this information may be so compelling as to suggest the necessity of a revised or alternative Robust Process Explanation (in CNR Jameson's reasoning about role of trichromacy seems such a case). On other occasions, this information may reflect a contingent feature of some segment of the population that is best ignored in constructing more general models (In CNR MacLaury's observations about individual subjects may be like this). It would be good if we could produce some rule or method for determining how some Actual Sequence Explanation is to be regarded with respect to robust process explanations (I think we shall see some interesting debate about this in the context of the disagreement discussed in this section—between Roberson *et al* and Kay *et al*). Such rules are, however, unlikely to be plausible. At best, I propose, we will be able to move forward, methodologically, by paying attention to the different kinds of explanation our theories propose, by examining them closely to identify when and if they are as incompatible as they seem, and by constructing *arguments* as to whether some Robust Process Explanation is beter than another.

There is a final issue which needs to be addressed. As many readers will have realized, the distinction between an Actual Sequence Explanation and a Robust Process Explanation is, for almost any domain, a relative matter. I think this is not such a serious issue. It is possible to slice one's population ever finer—MacLaury's concern with individual data is an example in CNR.—just as it is possible to find ever more robust properties (think of Kay and Regier's WCS data analysis as an instance of a very robust explanation that divides through a massive

set of cultural differences). That is not to say that such differences in grain must be simply a matter of generality. Because of complexity and contingency CNR will identify various more or less robust regularities that are, themsleves, compatible with one another. The compatibility of UT explanations based on (a) tricromacy and (b) the social imposition of trichromacy on non-trichromatic color categorizers is an example.

## Footnotes

[1] The classic philosophical text in this regard is Hempel and Oppenheim (1948). More recently see Lewis (1986) and Kitcher (1989). For a philosophical discussion of the limits of this model of explanation with specific reference to psychology see Cummins (1983).

[2] I am not interested in ruling any particular research in or out of CNR. If someone thinks some claim or method is relevant then, for the purposes at hand, it is.

[3] Perhaps the clearest evolutionary context where this is so is that of frequency dependent selection, in which the fitness of a trait depends on the proportion of individuals within a population which possess the trait (and the proportions in the population of other competing phenotypes). Frequency dependent selection is a crucial component of game-theoretic accounts of, for example, cooperation. If there are few cooperators in a population, cooperation is likely to be detrimental; if a larger number, cooperators are more likely to thrive via strategies such as tit-for-tat which are beneficial when directed toward other cooperators. Is cooperation an adaptation? It depends. (Maynard-Smith 1982)

[4] I am not claiming that this is a correct explanation for the existence of a yellow/green color term in some language. I am claiming that there is no way to rule such an explanation out, on the basis of some prior commitment.

[5] The idea here is that there may be distinct genetic bases for tetrachromacy. Note that if there are male tetrachromats, as Jameson suggests may be the case, this must be so.

[6] Jameson's proposed model is more complex than this, in the sense that non-trichromats may also possess some meta-knowledge about the fit between their perceptual model of color and their social model for color naming.

[7] For a review of MacLaury's main work (MacLaury 1997) see (Dedrick 2001).

[8] See the discussion in Chapter 5 of (Giere 1999) and Cartwright (1983) for a discussion of this issue in the context of physical science.

[9] I want to emphasize that a natural way of understanding the relationship between Actual Sequence Explanations and Robust Process Explanations—that the latter are simply a "generalization" of the former—is fundamentally mistaken. An account of the peculiarities relevant to MacLaury's single speakers is not what gets "generalized" into that author's claims about (for example) brightness categorization being distinct from hue categorization for the speakers of some languages. Similarily, Jameson's claims about tetrachromats do not, obviously, generalize to trichromats; Stanlaw's claims about the particularities of Japanese are, as Stanlaw points out, not exactly in conformity with the UT and as such can not be thought as a basis for the UT; in my overworked (and I suspect unloved) example, the albino tigers are not the inductive basis for robust explanations concerning the population of tigers as a whole. The Actual Sequence Explanation/Robust Process Explanation distinction is in fact most useful for understanding cases where there are differences among the members of a population. Generalization works in the opposite way, most powerful when the population is uniform with respect to the property of interest. In this sense, the

distinction between Actual Sequence Explanations and Robust Process Explaantions will be most useful when we are concerned with populations that vary. Biological groups satisfy this condition, virtually by definition, but many other groups also satisfy it, including groups which name colors, i.e. all human groups.

[10] This is not exactly right. As with Rosch, Roberson was interested in the comparison of Berinmo to English performance on various tasks. But, again as with Rosch, it is the non-English data which is of the most interest to the researchers. The subtitle of Roberson, *et al* (2000) is "Replications and new Evidence From a Stone-Age culture."

[11] This presents an interesting issue for Kay (Kay and Regier 2003, Kay 2005). For languages with so called composite categories (Kay and McDaniel 1978, Dedrick 1998a) the statistical methods used generate clusters located intermediate between the focal colors that form the composites (e.g. between green and blue; between red and pink; between yellow and red). This is, Kay and Regier say, what we should expect if it is the case that a language does not differentiate between, say, blue and green. It is however not in general the case that individuals or groups as a whole choose such points as their focus for the composite category. They choose, by and large, color samples that are found within the range of one of the composites (Dedrick 1998a Ch. 5). In this sense, and for these cases, it seems unlikely that the "central tendencies" Kay and Regier's method identifies can play a significant role in the explanation for color categorization, since such points are artifacts of the technique, with neither individual nor social salience (but see MacLaury 1997 for cases where individuals locate a color term focus intermediate

between focal green and focal blue). From the point of view of explanation, some of Kay and Regier's results may be *too* robust.

**Author Note**

I would like to thank the editors of this volume, Kimberly Jameson and Nancy Alvarado. Their penetrating comments on an earler draft of this paper resulted in much needed improvement.

## References


Berlin, B. and Kay, P. (1969). *Basic Color Terms: Their universality and evolution*. University of California Press.

Brown, R. and Lenneberg, E. (1954). "A study in language and cognition." *Journal of Abnormal and Social Psychology* 49 : 454-462.

Cartwright, N. (1983). *How the laws of physics lie*. Oxford University Press.

Casson, R. W. "Color shift : evolution of English color terms from brightness to hue." In *Color categories in thought and language*, edited by C. L. Hardin and L. Maffi, 224-239. Cambridge University Press.

Cummins, R. *The Nature of Psychological Explanation*. MIT Press, 1983.

Dedrick, D. (2001) "Review of C. L. Hardin & L. Maffi, eds*., Color categories in thought and language*, and R. MacLaury, *Color and cognition in mesoamerica: constructing categories as vantages*." *Minds and Machines* 10, 3 : 423-430.

Dedrick, D. (1998a) *Naming the rainbow: Colour language, colour science, and culture*. Kluwer Academic Publishers.

Dedrick, D. (1998b) "On the foundations of the universalist tradition in colour naming (and their supposed refutation)." *Philosophy of the Social Sciences* 28, 2 : 179-204.

Giere, R. (1999). *Science without laws*. University of Chicago Press.

Hardin, C. L. (1988) *Color for philosophers : unweaving the rainbow*. Hackett Publishing.

Hickerson, N. (1971). "Review of Berlin and Kay's *Basic Color Terms*." International Journal of American Linguistics 37, 4 : 257-70.

Heider, E. R. (1972). "Universals in color naming and memory." *Journal of Experimental Psychology* 93 : 10-20.

Heider, E. R. and Olivier, D. (1972). "The structure of the color space in naming and memory for two languages." *Cognitive Psychology* 3, 2 : 337-54

Hempel, C. G. (1966). *Philosophy of natural science*. Prentice Hall.

Jackson, F. and Pettit, P. (1992). "In defense of explanatory ecumenicalism." *Economics and Philosophy* 8 : 1-21.

Jameson, K. A. (Unpublished). *Culture and Cognition: What is Universal about the Representation of Color Experience?* Accepted for publication.

Jameson, K. A. and Andrade, R. G. (1997) "It's not really Red, Green,Yellow, Blue: an inquiry into cognitive color space." In *Color categories in thought and language*, edited by C. L. Hardin and L. Maffi, 295-319. Cambridge University Press.

Jameson, K. A., Highnote, S. M. and Wasserman, L.M. (1998). "Understanding color appearance phenomenology: can variation in photopigment opsin genes give rise to individuals with perceptual tetrachromacy?" *Perception* 27 : 173.

Jameson, K. A., Highnote, S. M. and Wasserman, L.M. (2001) “Richer color experience in observers with multiple photopigment opsin genes.” *Psychonomic Bulletin and Review* 8 (2) : 244-261.

Kay, P. (2005). “Color categories are non-arbitrary.” *Cross-Cultural Research: The Journal for Comparative Social Sciences, 39(1)*.

Kay, P., B. Berlin, L. Maffi, and W. Merrifield. (1997). “Color naming across languages.” In *Color categories in thought and language*, edited by C. L. Hardin and L. Maffi, 21-58. Cambridge University Press.

Kay, P., B. Berlin, and W. R. Merrifield. (1991). “Biocultural implications of color naming.” *Linguistic Anthropology* 1:12-25.

Kay, P., and C. K. McDaniel. (1978). “The linguistic significance of the meanings of basic color terms.” *Language* 54:610-46.

Kay, P. and Regier, T. (2003). **“**Resolving the question of color naming universals.” *Proceedings of the Naional Academy of Science* 100, 9085-9089.

Kitcher, P. (1989) “Explanatory Unification and the Causal Structure of the World", in Philip Kitcher and Wesley Salmon (eds), *Scientific Explanation*, University of Minnesota Press (Minnesota Studies in the Philosophy of Science, Volume XIII), 1989, 410-505.

Lewis, D. (1986) “Causal Explanation,” in LewisPhilosophical Papers Volume II (Oxford: Oxford University Press**),** 1986**.**

Lucy, R. (1997). “The linguistics of ‘color.’” In *Color categories in thought and language*, edited by C. L. Hardin and L. Maffi, 320-346. Cambridge University Press.

Kuschel, R., and T. Monberg. (1974). “We don’t talk much about colour here”: A study of colour semantics on Bellona Island. *Man* 9 : 213-42.

MacLaury, R. M. (1992). “From brightness to hue: an explanatory model of color category evolution.” *Current Anthropology* 33 : 137-186.

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

Maynard-Smith, J. (1982). *Evolution and the theory of games*. Cambridge University Press.

McNeill, N. B. (1972). “Color and color terminology.” *Journal of Linguistics* 8, 1 : 21-33.

Ratner, C. (1989). “A sociohistorical critique of naturalistic theories of color perception.” *Journal of Mind and Behavior* 10, 4 : 361-372.

Roberson, D., Davies, I., and Davidoff, J. (2000). “Color categories are not universal : replications and new evidence from a stone-age culture.” *Journal of Experimental Psychology: General* 126, 3 : 369-398

Saunders, B. A. C. and van Brakel, J. (1997) "Are their non-trivial constraints on color categorization." *Behavioral and Brain Sciences* 20 (2) : 167-179.

Saunders, B. A. C. and van Brakel, J. (2001). Rewriting Colour." *Philosophy of the Social Sciences* 31, 4, : 538-556.

Stanlaw, J. (1997). "Two observations on culture contact and the Japanese color nomenclature system." In *Color categories in thought and language*, edited by C. L. Hardin and L. Maffi, 240-260. Cambridge University Press

Sterelny, K. and Griffitihs, P. (1999). *Sex and Death*. University of Chicago Press.