A PHYSICALIST RELATIONIST THEORY OF COLOR

by

ELIEZER MINTZ

A dissertation submitted to the

Graduate School–New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Philosophy

Written under the direction of Professor Alvin Goldman

And approved by

New Brunswick, New Jersey

October, 2009

ABSTRACT OF THE DISSERTATION

A Physicalist Relationist Theory of Color

By ELIEZER MINTZ

Dissertation Director:

Professor Alvin Goldman

The nature of color is an open philosophical and scientific question. In this work I develop a physicalist relationist theory of color. So far, attempts to identify color as a physical property of objects have not been convincing because no physical property used by scientists seems to be well correlated with color sensations. I define a new physical property which I call transformance and show that transformance is 100% correlated with color sensations. Intuitively, transformance is a very general abstract physical property that describes how a system transforms or modifies light or information that characterizes light. It turns out that transformance is a relational property of objects like velocity and weight. Transformance is related to surface reflectance as weight is related to mass.

Transformance is a much better candidate to be color than surface reflectance because it is 100% correlated with color sensations, precisely models surround effects and fully explains all issues that relate to perceptual variation or the lack thereof (metamers). Several concrete examples are provided that show how the transformance of different systems should be modeled. After defining transformance, I defend the theory that color is transformance against possible objections while contrasting my theory with the theory that color is surface reflectance. I then discuss the relations between color and color sensation and the more general objections to any physicalist theory of color. The final part of the thesis deals with epistemological issues related to my theory of color including two apparent paradoxes that I believe every theory of color must answer: 1) How is it that human beings get along so well with an erroneous theory of color? 2) How can science, which is based on the common sense theory of color, conclude that the common sense theory is wrong? I show that my theory provides answers to these questions.

Acknowledgements

If it weren't for Prof. Barry Loewer's willingness to take a chance and accommodate a middle aged businessman with an interest in philosophy, you would not be reading this thesis. I am forever grateful to Barry for giving me the chance to pursue a graduate degree in philosophy at Rutgers. I am also grateful to Barry for recommending that one of the first seminars I take is Prof. Alvin Goldman's metaphysics seminar. That was the first time that I encountered the problem of color and have been obsessed with it ever since.

I would like to thank Prof. Alvin Goldman for agreeing to be my advisor and for his patience in teaching an old dog new tricks. Alvin's steady guidance and support were critical in maintaining the philosophical nature of this thesis. I would also like to thank the rest of my thesis committee, Brian McLaughlin, Randy Gallistel and Alex Byrne for their comments and for helpful discussions.

Last but not least, I would like to thank my wife Liat for her support and understanding. She never once asked what a degree in philosophy would be useful for.

Dedication

I dedicate this work to Yonatan, Idan and Michael.

I can't imagine anyone else was ever asked to contemplate the nature of color at such an early age and I hope I don't have to apologize for this later in your life.

Table of Contents

Chapter 1

Color, What is the Problem?

Tomatoes are red and the sky is blue. Isn't this clear to everybody? What is there more to say? Surprisingly, there are many interesting things still left to say. Let's begin by asking a few simple questions. These questions are not new and philosophers have been asking them and answering them for centuries. In the next few paragraphs I will briefly raise these questions again to demonstrate that understanding color isn't as trivial as most people take it to be.

Where do colors go in the dark? Let's reflect on this seemingly simple question for a moment. A cucumber in the dark still has a certain length, a certain circumference and a certain feel. It even has the same chemical composition in the dark. But, is the cucumber still green when the lights go out? It sure doesn't look green. How do we determine the color of an object? By looking at it of course. According to this simple method then the cucumber is not green in the dark, it is grey. To most people this does not sound right. When asked, "What is the color of the cucumbers in your fridge?" most people would answer "green", even though it is dark inside closed refrigerators.

The first conclusion to be reached from this line of reasoning is that if color is a property of objects, it is a strange one. It is quite different from length, mass and chemical composition. Somehow, light plays an important role in defining color and not in other common properties that we attribute to objects. Another interesting conclusion is that we cannot always tell the color of an object just by looking at it. We have to be sure that we are looking at it under appropriate conditions. What are these appropriate conditions is another problem worth contemplating (it does not have a simple answer either).

Let's reflect on another seemingly simple question: When we see a red car on television, what is it that is colored red? The obvious answer is that the car is colored red. This cannot be true. Clearly there is no car in the television. The thing that is red is a certain part of the television screen. But aren't television screens grey? It seems that the television screen is grey but can emit light in different colors. So is color light? If so, humans have been living under a grand illusion. We believe that it is objects that have color and not light.

Another question: What is it that color blind people see? Do they see color? If two people report a different color for the same object, what is the color of the object? Is color then only in one's mind and not out there in the "real world" since each individual's mind "colors" the world differently?

We use color everyday. When we look around our view of the world is dominated by color. Yet, just some rudimentary reflection shows us just how much we are really on shaky ground. Surely, it is a worthwhile effort trying to understand color better.

Physicalist Color Realism

Are objects colored? As the short discussion above shows, the answer isn't as trivial as it seems. Color realists who are also physicalists answer this question in the affirmative and argue that color is a physical property of objects. In this work I will argue for a realist and physicalist notion of color.

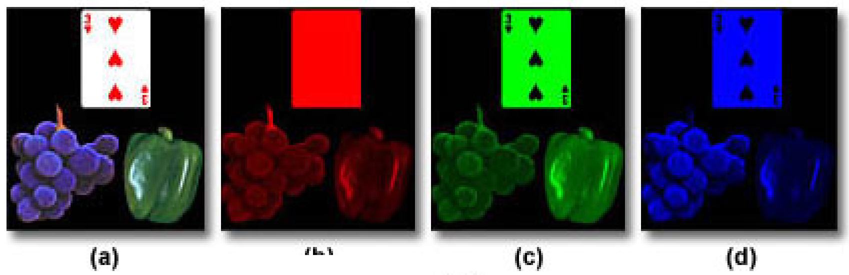
More precisely, the challenge for color realism of the physicalist type is to answer the question: “Which physical property (that is not a disjunction), if any, is correlated with color perceptions?” Once this physical property has been identified then color can be defined as *that* physical property. The reason this is the case is because color sensations are cognitively impenetrable. Our beliefs or what we know cannot influence the color sensations we have. If we look at a ball that we know is blue but is illuminated solely by a red light, we will see the ball as red even though we know and believe that the ball is blue. Color sensations influence beliefs but beliefs do not influence color sensations. For a physicalist, color sensations are like the output of an un-calibrated measurement device over which humans have no control. They are systemic representations, to use Dretske’s terminology (Dretske 95), and not acquired representations. They are non-conceptual and are just the outcome of a complex causal process which starts with color and ends with color sensations. The aim of the physicalist project as I see it is to reverse engineer this device and figure out what is the simplest and most basic physical property that is correlated with color sensations.

This interpretation of the physicalist project is also supported by Dretske: “[speaking about a pressure gauge] When I try to find out what the gauge is representing *k* as, I have to look at both the gauge and the world – at the gauge to tell that it is working right, and the world (i.e *k*) to find out what the world ***is*** like when the gauge is in state P.” (Dretske 95) Assuming that the parts of the brain or mind that are responsible for color sensations are working well, the question becomes whether there is a state of the world *k* that can be described using a non-disjunctive physical property and is highly correlated with color sensations.

Once the elusive physical property that is correlated with color sensations has been identified, it needs to be defended against scientific and philosophical challenges. For example, the theory that surface reflectance is color does not pass muster with most scientists and many philosophers because of the problem of perceptual variation. The essence of the problem is that correlation exists between our sensation of color or color experiences and surface reflectance only when very specific conditions are present. Tomatoes do not look red in the dark even though the surface reflectance of tomatoes is the same in the light as it is in the dark. Conversely, objects with very different surface reflectances look the same to us. There is also significant variance between the color experiences of viewers in identical viewing conditions. In addition, there is the problem of contrast colors such as brown and black that are dependent on the object's environment and not only its surface reflectance. For example, with a change in background brown objects begin to look orange.

Among color physicalists, surface reflectance has been the leading candidate for being color. Because of surface reflectance's inadequacies, many philosophers and color scientists have become color eliminativists. Scientists and philosophers holding this view believe there is no actual property in the world that correlates well with color experiences. Color is only a psychological property of the mind. It is just an illusion, a useful human invention with no corresponding entity in the world. Galileo summarized this view well about 400 years ago: "I think that tastes, odors, colors, and so on are no more than mere names so far as the object in which we place them is concerned, and that they reside only in the consciousness. Hence, if the living creatures were removed, all these qualities would be wiped away and annihilated." (Drake 1957)

Color Variations with Illumination



re 3

The same scene illuminated by (a) white (b) red (c) green and (d) blue light

This work is a defense of physicalist color realism but not of surface reflectance. I will argue that the jump from surface reflectance's inadequacy to eliminativism is too hasty. Couldn't there be another physical property that is more appropriate? Such a property has not been identified by scientists, but this of course is not proof that such a property does not exist. In this work I define a physical property that is highly correlated with our color experience. I call this property transformance. While a full understanding of transformance is not possible without the aid of mathematical formulation, transformance can intuitively be described as a very general abstract physical property that describes how a system transforms or modifies light or information that characterizes light. It is an abstract physical property that can be attributed to any physical system. It is related to surface reflectance but not identical to it. As a starting point for understanding, one can think of the relation between transformance and reflectance as the relation between weight and mass. Once transformance is described, it will become clear why scientist have not bothered to define this physical property. Because of the way science is divided into fields and specialties, transformance is not useful for any practicing scientist. It is however I believe very useful for gaining a philosophical understanding of color.

Physicalist **Relationist** ***Color Realism***

My definition is not only physicalist but also relational since transformance turns out to be a relational physical property of objects. Many physical properties are relational and can only be defined relative or in relation to another physical entity[1]. For example, the velocity of an object is relative to a frame of reference. In the case of a person sitting in a moving train, his velocity relative to the train is zero while his velocity relative to the platform could be 100 km/h. In fact, since there are an infinite number of frames of reference, a person can be attributed an infinite number of different velocity values all at the same time. Each of these values is associated with a different property of the person, velocity relative to the train, velocity relative to the platform, velocity relative to the sun, etc. Since velocity is relational, so are several well known properties that are derived from it such as momentum and kinetic energy.

Another example of a relational physical property is weight. Strictly speaking, when we specify the weight of an object, we need to specify in which gravitational field the object is. The weight of the object on earth is different that its weight on the moon. Weight is a physical property of an object that is relative to the gravitational field the object is in.

In the same way, it turns out that transformance is a relational physical property *if it is attributed to an object*. To explain what I mean, consider again the case of the person

[1] Throughout the text, when I write that "x is relative to y" I mean that x can only be defined as a relation to y, not that the truth value of x is relative to situation y.

sitting in the train. Instead of attributing the velocity of 100 Km/h to the person, we can attribute it to the system made up of the person and the platform. In the same way we can attribute a velocity of zero to the system consisting of the person and the train. When attributed to a system, the property is not relational anymore. Simplistically stated, velocity is intrinsic for some systems while it is relational for objects. Transformance also can be attributed either to objects or to systems. Whichever path one chooses makes no scientific difference, but by choosing to attribute color to objects, a bridge can be built between our common sense view of color and the scientific view of color, and this of course is philosophically advantageous.

The idea that color is a relational property is not novel. Specifically, Brian McLaughlin has argued that color is a relational physical property if there is a physical property that is common to all objects that have the same color (McLaughlin 2003). I will argue that transformance is general enough to play this role and that McLaughlin's optimism that such a property can be found was warranted.

I will argue that in attributing color to objects humans make the mistake of not realizing that color is a relational physical property of objects. Our case is similar to that of the person who measures the pressure in his auto tires once at noon and once at night and is surprised to find a large discrepancy. He assumes that pressure reflects the amount of air in the tire and is non-relational. He doesn't realize that air pressure depends on temperature and volume and is a relational property. The change in temperature between day and night is what causes the discrepancy. However, because the person possesses an erroneous theory of pressure, he concludes that something was wrong with the pressure gauge or the way the measurement was done in one case and therefore one measurement was veridical while the other was not. This is a small error that shouldn't be categorized anywhere near a "grand illusion".

This is different from the view of color relationists (sometimes also called color relativists) found in the literature that view our concept of color as being that of a relational property (Jackson and Pargetter 1987, Cohen 2003a, 2003b, 2004, Jakab and McLaughlin 2003, McLaughlin 2003). My view is that our concept of color is that of an intrinsic property of objects but since color is really a relational property of objects, we make mistakes when we speak of colors, not in sensing them. When we say that the color of the shirt was different in the store, we don't mean that relative to the light in the store the color was different. Rather, I believe our common sense view is that in the store we were mistaken about the intrinsic property of the shirt, namely its color. We are implying that our senses were not functioning well on one occasion. The situation though is that our senses are functioning perfectly well in both cases and the reason we believe we are mistaken in one of the cases is that we don't realize that color is a relational property.

Vision Science, Philosophy and Terminology

Understanding color is a multidisciplinary endeavor in which vision scientists and philosophers are involved. Since Larry Hardin published "Color for Philosophers" (Hardin 1982/1988) it has become widely accepted that any philosophical theory of color needs to be constrained by and evaluated against the scientific data. I would even strengthen this constraint and assert that any physicalist philosophical theory of color must be accepted as plausible by vision scientists and be compatible with their own theories. In addition, any philosophical theory must be translatable to the terminology of vision scientists. The onus is on philosophers to make their terminology as clear as possible to vision scientists.

The problem from the vision scientist's angle is described by Davida Teller (Teller 2003): "The key to understanding the perspective of vision scientists is that our goal is to unite three interestingly diverse kinds of entities: Visual stimuli (e.g., physical objects and their properties); neural states (the states of ensembles of neurons at many processing stages within the visual system); and conscious perceptual states (our visual perceptions of particular physical stimuli). We wish to discover and understand the regularities, or mapping rules, between physical states and perceptual states, between physical states and neural states, and between neural states and perceptual states. The first two kinds of mapping rules are the domain of visual science; the third kind has remained largely in the realm of philosophy."

She continues: "Now, as far as I can see, color realism is the view that of the vision scientist's three entities – surface spectral reflectance, neural signals, and perceived color – one is color, and the other two are not. But if you ask a color scientist which of the three entities is color, she will answer that the question is ill-posed. We need all three concepts, and we need a conceptual framework and a terminology that makes it easy to separate the three, so that we can talk about the mappings among them. Color physicalists can call surface spectral reflectance physical color if they want to, although surface spectral reflectance is a more precise term. But to call it color (unmodified) is just confusing and counterproductive, because for us the physical properties of stimuli stand as only one of three coequal entities."

Because physicalist color realists have so far mostly equated color with surface reflectance, vision scientists rightly believe that color realism and surface reflectance go hand in hand. I ask vision scientists to give color realism another chance and entertain the possibility that there could be another physical property that does fit the bill. I will argue that this property is transformance, a property that involves light, objects and neurons.

My definition of color will not surprise most visual scientists since it satisfies Hurvich's observation (Hurvich 1997): "It should be clear by now that object color is not physical light radiation itself, that it is not something that inheres in objects, having to do exclusively with the chemical makeup of the object, nor is it only the nervous excitation that occurs in the eye and brain of an observer." Color, according to my definition is a property that inheres in the combination of the light source, the object and the brain. It cannot be isolated in any one of them and can only be found in their combination.

Teller also makes a suggestion with which I wholeheartedly agree: "It would help if the argument were framed in a neutral terminology, and within (or in an instructive relationship to) the vision science worldview."

In order to follow Teller's advice regarding terminology, I will attempt to use the following definitions rigorously.

Color – the physical property (if it exists) that is highly correlated with color experience or color sensations. That is, whenever this physical property is a certain value, the color sensation associated with it is of a specific kind and only that kind.

Color Experiences, Color Sensations – that difficult to describe conscious or perceptual state that one has when one looks at a tomato. It is what some philosophers call "what it is like to see color". Color sensations are only in the mind.

Physical Color – this is a term I will not use. Instead I will talk of actual physical properties such as surface reflectance, emmitance and transformance.

Color and Color Experiences

Understanding color does not require understanding color experiences. As the definitions above show, from a color realist point of view, color and color experiences are completely different things that are correlated through an apparently causal chain. Light hits objects and then reflects to the eye and is analyzed by the visual system. Somehow at the end of this chain or perhaps in a way ancillary to it, color sensation occurs. Not knowing how this happens does not hamper the understanding of color much in the same way that not understanding how a digital thermometer works does not necessarily hamper understanding of temperature nor not understanding how an Ampere meter works hampers the understanding of electrical current. One can understand that temperature is the average kinetic energy of molecules without understanding how a digital thermometer measures temperature. All one needs to know is that the number displayed by the thermometer is well correlated with temperature. Furthermore, specifically in the case of temperature, our initial thermometers were our hot-cold sensations. We still don't understand hot-cold sensation but we do understand temperature. I will argue that this can also be the case with color. We need not understand color sensations to understand color. What is sufficient for a realist understanding of color is that color (a physical property) is well correlated with color sensation (a psychological property which is perhaps reducible to a physical property).

The leading argument against color physicalism is that the relations between color experiences pose a constraint on the definition of color and that there are no physical properties that stand in the appropriate relations. For example, most people would accept that red is more similar to orange than to blue and that this tells us something about red, orange and blue. I will argue that this is a mistake that is caused initially by conflating color with color sensation or by assuming that relations between representations of physical properties constrain the relations between the physical properties. For example, a person not familiar with Arabic numerals may look at the display of a digital thermometer and decide that the shape “6” is more similar to the shape “9” than to the shape “7” and therefore deduce that 6°C is closer to 9°C than to 7°C. I will argue that in the same way that the shape of numbers has no relevance to physical closeness of temperature values, similarities between color sensations puts no constraints on similarities between different values of color.

In order to show that the brain does not only represent color using color sensations, I will discuss color blindsight as well as recent experimental evidence that shows that completely healthy adults can discriminate between colors without having any color sensations (Boyer et al. 2005, Lamme 2006). Once it is clear that color sensation is just one way of representing color in the brain, it will become even clearer that color sensation does not constrain what color is.

Structure of the Work

In the next chapter of this work, transformance will be defined. Several examples will also be given of how the transformance of a system should be modeled. In the third chapter I will defend the theory that color is transformance against possible objections while contrasting my theory with the theory that color is surface reflectance. In the fourth chapter I will discuss the relation between color and color sensation and the more general objections to any physicalist theory of color. The fifth chapter will deal with epistemological issues related to my theory of color.

Literature Survey

The scientific and philosophical literature about color is vast. In this survey I will focus mainly on philosophical publications from the last thirty or so years. Galileo, Locke and Russell had very interesting things to say about color but they lacked the scientific perspective that we are fortunate to now have. For example, it is obvious to current day practitioners that any theory of color must address head on the issue of contrast colors and simultaneous contrast and that this is not a minor detail that can be glossed over. After all, any surface can be made to look any color by changing its surround. Every theory of color needs to be evaluated against this and many other salient scientific facts about color.

I believe it is correct to say that the issue of contrast colors and simultaneous contrast was not consistently addressed by philosophers until the late 1980's (or maybe even later) when the implications of Hardin's "Color for Philosophers: Unweaving the

Rainbow" began to sink in (Hardin 83/88). In his book, Hardin reviewed the scientific knowledge about color and based on this review, advanced a strong argument for eliminativism, the view that colors do not exist (or that nothing is colored). Just as importantly, Hardin framed the philosophical discussion in such a way that it is practically a given that any philosophical discussion about color needs to be constrained by the scientific knowledge about color.

The different theories of color prevalent in the literature can be categorized based on several different criteria. The major fault line though is between physicalist theories and eliminativist theories. According to physicalist theories colors are physical properties of some kind or another. Eliminativists argue that either non-physical objects are colored or that nothing is colored. The reason that the fault line in the argument is the physicalist-eliminativist divide is that color scientists as scientists are implicitly physicalist. Therefore, any discussion that includes them also implicitly assumes that if color exists as a property in our world, it must be a physical property of some kind.

Physicalist theories of color can themselves be categorized according to different dimensions:

1) Is the theory relationist or absolutist (is the physical property a relational one or an absolute/intrinsic one)?

2) What is the physical property or properties that should be equated with color?

3) Is color one property or several?

4) Should color be defined anthropocentrically or not (that is, relative to human experience)?

An initial "bare bones" theory that colors should be equated with micro-physical properties of bodies emerged from the discussion between Smart, Armstrong and their critics (Armstrong 68, 87, Smart 75). Jackson and Pargetter extend this theory in both a relational and dispositional way (Jackson and Pargetter 87). They make the theory relational by defining color relative to viewers and viewing circumstances. Their theory also has a dispositional bent because it defines green as the property that under standard conditions would cause an object having the property to look green. More recently, Lewis has also supported a view that colors are microphysical properties (Lewis 1997).

The next step forward for physicalist theories was to view color as an abstract physical property such as surface reflectance. Hilbert initially proposed such a theory (Hilbert 1987). Subsequently Byrne and Hilbert elaborated and defended this view in several essays (Byrne and Hilbert 1997b, 2003, 2004). Byrne and Hilbert's "Color Realism and Color Science" (Byrne and Hilbert 2003) is of special interest because it is followed by numerous commentaries, both by philosophers and scientists, and a response by the authors. There are several other philosophers who have also argued that color is surface reflectance (Dretske 1995, Matthen 1988, Tye 1995). Hardin has been the main critic of surface reflectance as color (Hardin 1988, 2003, 2005).

In order mainly to solve the problems associated with perceptual variance, relational theories of color were introduced (Jackson and Pargetter 87, Cohen 2003a, 2003b, 2004, Jakab and McLaughlin 2003, McLaughlin 2003). Strictly speaking, relational theories do not have to be physicalist but except for Cohen who is noncommittal, all the theories in the literature are physicalist. A variant of the relational view is the ecological view of Thompson (Thompson 1992, 1995, 1995a). According to this view, colors need to be analyzed as a relation between the perceiver and his environment.

Mohan Matthen has argued for a physicalist view that he calls pluralistic realism (Matthen 1999, 2001, 2004). According to Matthen, in different organisms, colors should be associated with different properties.

Dispositionalism is the view that colors are powers that induce certain visual experiences in perceivers, usually in certain conditions (Johnston 1992, McGinn 1983, Peacocke 1984). Dispositionalism may also be of the physicalist kind if the dispositions are equated with physical properties of objects.

Primitivism is the view that colors are sui generis properties that are transparent to us (Campbell 1993, Stroud 2000, Byrne and Hilbert 2007). The discourse about primitivism is the only modern one that rejects the idea that science constrains the philosophic discussion of color.

The prevalent opposing view to physicalism is eliminativism. This is the view that everyday objects do not have color. Hardin is of the view that nothing is colored (Hardin 83/88). Boghossian and Velleman are projectivist and view some mental elements are colored (Boghossian and Velleman 1989, 1991). Most color scientist can be read as supporting eliminativism. I think such reading is not accurate. Color scientists are eliminativists about the colors of common sense while philosophers are eliminativist about color in general.

Byrne and Hilbert have compiled a very useful anthology of scientific and philosophical essays about color (Byrne and Hilbert 1997a, 1997b).

Chapter 2

Transformance Defined

In this chapter a physical property, transformance, will be defined. Transformance is highly correlated with color sensation for all perceivers under all viewing circumstances and thus is the answer to the initial challenge of eliminativism. My aim is to provide a relatively succinct mathematical/scientific definition of transformance and how it correlates with color sensation in this chapter and to present more rigorous argumentation in support of equating color with transformance in the next two chapters.

Hardin makes the following observation: “What is it about colors that seems to obstruct our understanding? They are given to visual experience along with shapes, yet we have no similar difficulties with shapes. A crucial difference seems to be that the essential character of shapes is amendable to mathematical representation, but the essential nature of colors resist it; the one appears quantitative the other qualitative.” (Hardin 1988)

In this chapter I will attempt to show that colors can be represented mathematically as a first step to better understanding of color.

Some Preliminary Definitions

For our purposes, a radiation source L will be characterized by the energy it produces per unit time in every wave length. L^w will denote the energy of light source L in wavelength w. Unless otherwise stated, I assume that radiation sources are points in space and that they emit the same energy in a uniform way over time. These restrictions can be relaxed without significant change to the definition of transformance. Throughout the text the expressions “light source” and “radiation source” should be taken to have the same meaning.

A transformation T on a light source is defined to be a transformation that takes as an input a radiation source and outputs a radiation source. For every input value the transformation outputs one and only one output value. Loosely speaking, the transformation modifies the input radiation source into the output radiation source. We will see shortly that surface reflectance is an example of a physical property that can be modeled as a transformation. For each wavelength, the transformation allows us to compute a new energy value based on the energy values at each wavelength of the input radiation source. Transformations will either be denoted by their name or by their name followed by empty parenthesis. For example T and T() denote the transformation T.

If T is a transformation and L is a light source then by definition T(L) is also a light source. It is the light source that results from applying the transformation T on light source L. Another way to describe this process is as T() transforming L into T(L). Following our notation above, since T(L) is a light source, $(T(L))^w$ is the energy of T(L) in wavelength w. In order to make this clearer an example involving functions can be given. If the function *f* is defined as *f*(x)=3x+5 then *f*(7) is not a function, it is the number 26. In my parlance *f*() (or just *f*) transforms 7 into 26.

Let’s look at a few examples of transformations of radiation sources:

a) The zero transformation.

T_Z, the zero transformation, is defined as the transformation that for every light source L, and for every wavelength w, $(T_Z(L))^w=0$. Our shorthand notation for this would be: $T_Z(L)=L_0$ where L_0 is the light source with zero energy in every wavelength. The zero transformation applied to any light source transforms it into the zero energy light source.

b) The identity transformation.

The identity transformation, T_I, is the transformation that for each light source L and for each wavelength w, $(T_I(L))^w=L^w$ or more succinctly $T_I(L)=L$. For every light source L, T_I transforms L into itself or in other words, doesn't change L at all.

c) The window transformation.

$T_{w1,w2}$ is a window transformation that for every light source and every wavelength w, if $w<w1$ or $w>w2$ then $(T_{w1,w2}(L))^w=0$ otherwise $(T_{w1,w2}(L))^w=L^w$. Intuitively, a window transformation is the zero transformation outside a given wavelength range and is the identity transformation within the range.

The Reflectance and Emittance Properties of Objects as Transformations

The surface reflectance of objects can be modeled as a transformation. Due to the object's surface properties, the light hitting the object is not the same as the light leaving the object. The surface reflectance of an object specifies how a light source will be changed or transformed by the object. More formally, the surface reflectance of object O is the transformation T_O that has the following characteristics: for every light source L,

$(T_O(L))^w=c^wL^w$

where c^w is the percentage of the energy reflected in wavelength w by object O. In words what the equation above means is that the way the transformation T_O modifies any input light source L's energy in wavelength w is by multiplying the original energy of L in wavelength w by a constant between 0 and 1. This constant which is a real number is designated by c^w. Surface reflectance is usually portrayed in a graph (figure 1). In my notation, the y axis of the graph shows for each w on the x axis its respective c^w. In practice, the c^w for each w is obtained by measuring actual energy values and comparing the incident energy to the reflected energy. Surface reflectance is a multiply realizable property. Two surfaces may have the same reflectance but can be made of different materials.

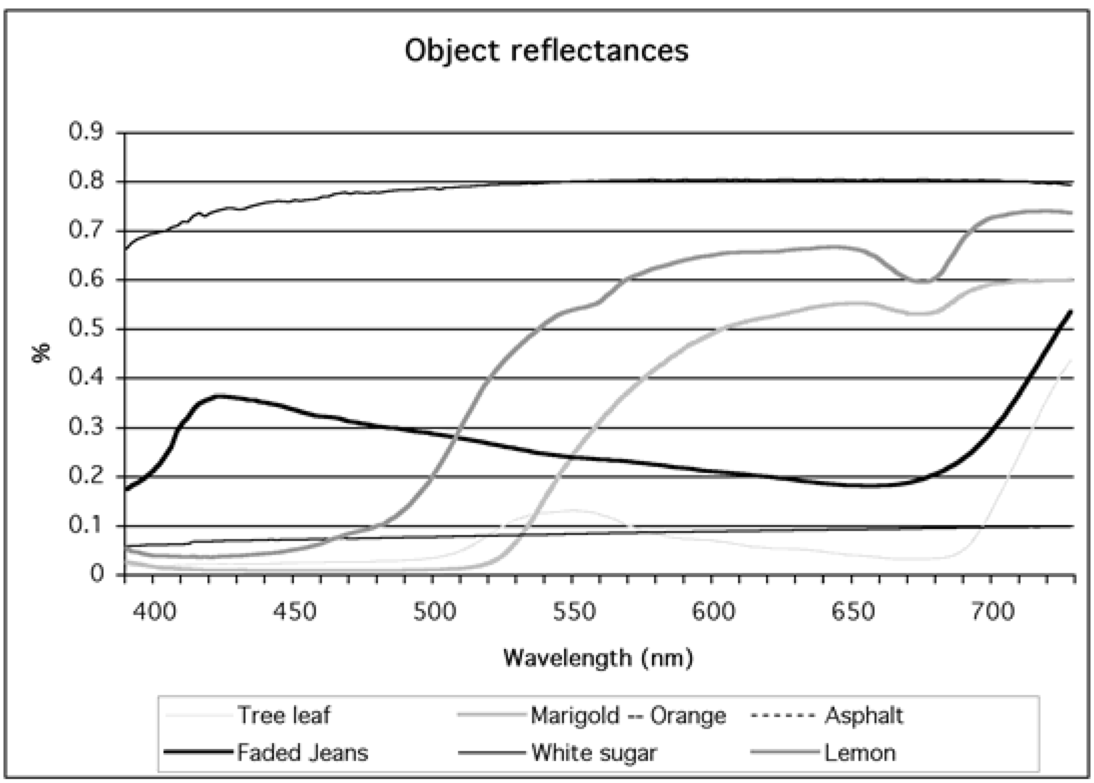


Figure 1 – Taken from Color Realism and Color Science

The reflectance and emittance characteristics of an object, taken together, can also be considered a transformation. Take the transformation T_O that has the following characteristics: for every light source L,

$$(T_O(L))^w = c^w L^w + e^w$$

where c^w is the percentage of the energy reflected in wavelength w by object O and e^w is the absolute energy emitted independently[2] by the object O in wavelength w. Emittance cannot be a transformation by itself because an object always has reflectance characteristics. Full absorbance is mathematically speaking also a kind of reflectance because it means that for every w, c^w is zero i.e. the percentage of the energy reflected at each wavelength is zero.

Light Sources Described as Transformations

In order to aid our mathematical formalism, we would like also to treat light sources as transformations. We do this by mapping light sources to transformations of a standard light source L_S. This is no more than a mathematical trick that will prove convenient later on. Instead of defining a light source L by how much energy it emits in each wavelength we define L by how a standard light source L_S needs to be transformed in order to get to L. In essence, the energy in each wavelength of L is defined relative to the energy in the same wavelength of L_S instead of as a number. For example, if in wavelength 450 nm the energy of L is 80 energy units and the energy of L_S is 40 energy units, then the transformation describing L would state that at 450 nm one has to multiply the energy of L_S by two.

---

[2] This is energy in the form of radiation that the object would emit even if no light is shined on it.

As an example to clarify this move, consider how the real numbers can also be mapped to functions operating on a standard number, say 3 (we can choose any number as the "standard" one). In this case 1 would be mapped to the function that multiplies a number by one third, 2 would be mapped to the function that multiplies a number by two thirds, 3 would be mapped to the identity function and so on:

1 is mapped to $f(x)=(1/3)*x$

2 is mapped to $f(x)=(2/3)*x$

3 is mapped to $f(x)=(3/3)*x=x$

And the general case: r is mapped to $f(x)=(r/3)*x$

Obviously, if the "standard" number is changed, the mapping changes also. If the "standard" number we choose would be 5 instead of 3 then 1 would be mapped to the function that multiplies a number by one fifth, 2 would be mapped to the function that multiplies a number by two fifths, 5 would be mapped to the identity function and so on. The mapping with 5 as the "standard" number would be:

r is mapped to $f(x)=(r/5)*x$

As is in the general case, also in the case of light sources, the standard light source that we choose would be mapped to the identity transformation since $T_I(L_S)=L_S$ by definition.

Using different standard light sources to define the mapping is analogous to rotating and displacing an axis or to using different units of measurement. It merely changes the mapping from transformations to light sources. For example, in the case of temperature, the same number can represent different temperatures depending on which units are used. 20° Celsius is quite different from 20° Fahrenheit. Depending on the arbitrary choice of the standard light source, light will be mapped to different transformations but this does not change what light is and subsequently will not change what color is.

The Visual System Transforms Light

The claim that the visual system transforms light may seem strange since there is no light in the brain to be transformed. Though there is no light in the brain, the brain can be viewed as encoding light, just as music is encoded on CDs and movies on DVDs. In cognitive science, it is not controversial that the brain encodes external stimuli. This is usually referred to as the brain creating a representation of those stimuli. However, the encoding by the visual system does not maintain 100% of the information about the light hitting the retina. The encoding distorts the information just in the same way that a music signal is distorted when it is transmitted over an AM radio signal. Any listener to an AM music station on the radio will immediately notice that the quality of the music is much less than music transmitted over an FM channel or played directly from a CD. Distorting a signal is just one way of transforming it. The encoding for AM transmission transforms the original music to music that is similar to the original but not the same as the original. During the AM encoding, information about the original signal is lost. Another way of describing this it is to say that AM encoding distorts or degrades music. The same process happens in the visual system. It transforms light hitting the retina in the process of encoding it. In this manner we can talk of the visual system as transforming light and designate this transformation as T_V. Distortion happens when encoding is done in a way that causes information to be lost. In a lossless encoding, the original signal can be reconstructed from the encoded signal. In an encoding with loss, this is not possible. There is not enough information, so to speak, in the encoded signal to reconstruct the original signal unambiguously. Only an approximation of the original signal can be reconstructed.

It is not controversial to say that AM encoding or transmission distorts music, or that music played on an AM channel is of lesser quality than music played on an FM channel. In fact, most if not all people would ascent to these propositions. However, there are no sound waves in the electronics that make up the radio transmitter. It transmits electromagnetic waves (it is in fact a radiation source) and not sound waves. The electromagnetic waves are decoded in our car radio (for example) and translated into electronic pulses that are fed to speakers and cause their membranes to move. It is the movement of the speaker's membranes that generates sound waves.

Had our car radio been disconnected from the car speakers, would it be right to say that the AM transmission did not distort the music? I think not. What is important here is that sounds are encoded in the car radio and with the correct interface could be retrieved. Or assume that our car radio has a special mode in which it only saves music on an internal hard drive but does not send any signals to the speakers. This could be useful if for example we want to take a call on our cell phone and not miss the song currently being played on the radio. Once the call ends, the radio can continue sending signals to the speakers, but based on the music saved on our hard drive and not the music encoded in the current AM transmission (in the realm of television, this is similar to what a TIVO does). But what if once we end the phone call, we do not continue listening to the radio, but turn it off? In my view it would still be the case that the AM transmission distorted the music, even if it never assumed the form of sound waves.

In the case of light and the visual system I believe the situation is similar. The fact that the visual system does not change its encoded light source back into an actual flow of photons or electromagnetic radiation does not matter. What matters is that potentially, this could have been possible. With sound, humans can actually accomplish this feat. We are able to encode air waves in our brain and then reproduce them using our vocal chords. With light, we are more limited, as we are only able to encode light but not reproduce it. But it is quite likely that as our understanding of our brain grows, we would be able to connect electrodes in specific places and connect these to a projector that would reproduce light. The encoded information as to the characteristics of the light source are there, and that is what allows us to speak as if light is present in the visual system.

One part of the transformation light undergoes in our visual system is a window transformation in which the window is the visible range. The visual system initially transforms any radiation source with energy outside this wavelength window into a radiation source with energy only in the visible window. Even more distortion occurs because the human visual system has only three kinds of cones so it at most can provide a three parameter approximation of the spectrum of the light hitting the retina. When a space with infinite dimensions is approximated by three parameters, in most cases information is lost even when the statistics of typical illumination are taken into account (they are presumably innate). The characteristics of the original light hitting the retina cannot be reconstructed from the approximation and therefore the light is distorted.

The way the visual system transforms light is a function of its physical state. Since the physical state of the visual system changes over time, the way it transforms light may also change over time. The transformation associated with the visual system is therefore a physical property of the visual system that may change as the physical state of the visual system changes. One example of such a change is the adaptation of the visual system to different lighting conditions.

How exactly the visual system encodes light sources is not yet known but this is not important for the definition of transformance or T_V. Surface reflectance as a property was used before the exact relation between the microphysical structure of surfaces and their reflectance was understood. In fact, surface reflectance is an abstract property that can be used without any knowledge of the microphysical structure of a surface or how light interacts with matter. In the same way, T_V is an abstract property that can be used without any specific knowledge of how the visual system encodes radiation sources. It may seem counterintuitive to some vision scientists that understanding exactly how the visual system actually processes light is not necessary for defining color. It turns out though that merely understanding that the visual system encodes light sources but does so while losing information about their characteristics, is enough. We shall see this shortly.

Getting to Transformance

Let's begin with a simple system made out of a light source L, a simple object O and a perceiver P (we ignore contrast effects at this stage but will discuss them and more complex systems later). First we agree on a standard light source, L_S. Let L_S be for example typical daylight at some geographic location. The choice is not really important and will only change the units in which color will be measured just as defining the units of temperature as Centigrade or Fahrenheit is not important in understanding temperature.

Based on the discussion above we write:

(1) $L=T_L(L_S)$

where T_L is the transformation that exemplifies the physical characteristics of L given L_S. T_L describes how the light source L_S needs to be modified in order to obtain the light source L.

When the light from the light source L hits object O, the object reflects light according to its surface reflectance properties. Based on our definitions, surface reflectance is a transformation which we will denote as T_O.

Taking into account (1), the light leaving the object and hitting the retina of the perceiver is the following:

(2) $T_O(T_L(L_S))$

The formula in (2) defines a light source. It is the light source that L_S is transformed into after the two transformations T_L and T_O have been applied.

Next, in the visual system, the light is transformed again. Let's denote the way the visual system transforms or distorts light as T_V. The resulting (virtual) light source that is encoded in the brain is:

(3) $T_V(T_O(T_L(L_S)))$

Transformance for this simple system is defined as the complex transformation:

(4) $T_S=T_V(T_O(T_L()))$

Transformance is a physical property that describes how a system transforms a standard light source. It is the property I propose that should be equated with color. I will argue that surface reflectance by itself is not enough to describe color. Two other transformations need also be taken into account: the transformation of the actual light source relative to a standard light source and the distortion of light by the encoding process of our visual system. Color is then the abstract physical property that explains how a system transforms a standard light source.

The advantage of this proposal is that there is a very high correlation between our color sensations and T_S and the problem of perceptual variation is solved. This is because T_S takes into account all the elements that influence our sensation of color while not being a disjunctive property. All the variations because of illumination, object characteristics and the properties of the visual system are taken into account. If the lighting changes this means that T_L has changed and therefore T_S will be influenced. If the object's surface reflectance changes T_O has changed and therefore again T_S will be influenced. If the visual system adapts or changes or is different between viewers, this means T_V has changed and T_S again will be influenced.

Our definition will not surprise most visual scientists since it satisfies Hurvich's observation: "It should be clear by now that object color is not physical light radiation itself, that it is not something that inheres in objects, having to do exclusively with the chemical makeup of the object, nor is it only the nervous excitation that occurs in the eye and brain of an observer." (Hurvich 1997) Color according to my definition is a property that inheres in the combination of the light source, the object and the brain. It cannot be isolated in any one of them and can only be found in their combination. Transformance can also be viewed as a relational property of objects. This will be discussed in detail in the next chapter.

The Relation between Transformance and Surface Reflectance

As we saw above, in the case of a simple system transformance is defined as the complex transformation $T_S=T_V(T_O(T_L()))$

Let's assume for a moment that a viewer is looking at an object in typical daylight or some other very similar light. Also, let's assume that the standard light source we use to map from light sources to transformations is also typical daylight. Let's designate this light source as L_S. Because the actual light is the same as the standard light, the actual light is mapped to the identity transformation (since by definition $T_I(L_S)=L_S$). When the actual light is similar to the standard light, T_L is very close to the identity transformation and we can write T_S as:

$T_S=T_V(T_O(T_I()))$ which can be simplified to:

(5) $T_S=T_V(T_O())$

Let us also assume initially that the distortion of the visual system is small and therefore T_V can be also approximated by the identity transformation. Simplifying color again we get:

$T_S=T_O()$

Under the assumptions above, color is approximated well by T_O, which is the surface reflectance of the object.

Let us now make a slightly more realistic assumption about T_V, the distortion introduced by the visual system. Instead of assuming that it is negligible, let us make the more realistic assumption that it is substantial but almost constant across viewing conditions. The only things that change in this scenario are the objects viewed. Since T_V doesn't change significantly, the color sensation associated with each object is always the same. To humans it would seem in this case that color sensation only changes as a function of the object and not as a function of anything else and it would be natural to attribute color just to the object. This is analogous to the case of the person that weighs different objects only on earth. He would probably reach the conclusion that weight is only a function of the mass of the body because everywhere one weighs an object on earth the weight is pretty much the same. Weight though is also a function of the gravitational field the object is in but due to limited experience and lack of knowledge a person may not realize there is a link between weight and gravitational field. Understanding that there is a link between how the visual system encodes light and color is a conclusion that is just as difficult to reach.

Another factor contributing to our attribution mistake[3] is color constancy. As the lighting conditions, which are exemplified by T_L, change, T_V changes (the visual system adapts by changing its physical properties) so as to accommodate the change and keep T_S (the color of the system) constant as much as possible. Color constancy extends significantly the range in which the color of the system is a good approximation for the surface reflectance of objects. An example of such an adaptation process would be the sun taking into account the movements of the planets and changing its internal

---

[3] What the attribution mistakes entails about the theory and whether it renders it an "error theory" will be discussed in later chapters.

distribution of mass so as to keep the center of gravity of the solar system in the same place relative to the center of the galaxy.

Maloney makes the following comment about idealized conditions: “If there are idealized philosophers in this idealized environment, they would be wrong to consider color as anything other than a perceptual correlate of objective surface properties, the intrinsic colors of surfaces.” (Maloney 2003)

In many cases the assumptions above do not hold and therefore T_V and T_L cannot be neglected. The formulation of transformance explains why under “normal conditions” it makes sense to associate color with surface reflectance. In some specific range of conditions, transformance can be approximated well by surface reflectance.

As for the assumption that the actual standard light source the brain uses to represent transformance is very close to typical daylight, I believe it makes sense for two reasons. First, this would be the units in which computation is the simplest. It would allow the brain of the first organisms to evolve color vision to make the simplification formulated in (5) and at first basically “ignore” or discount the physical characteristics of the light source. Primitive color systems may have assumed that the actual light was typical daylight and could not handle variations in light at all. The later evolution of color constancy involved relaxing the assumption about the characteristics of the light source. The color constancy mechanism assumes that the light fluctuates around a standard light source instead of just assuming that the light is fixed. To summarize the first point, with the right choice of measurement units, the initial evolution of color vision could have been relatively simple and less expensive from a computational point of view. Second, if these units are the ones used, it would help explain why we are naturally inclined to tie color with normal viewing conditions and why we attribute color to objects. Whether this conjecture is true, is of course, a matter for additional research.

Contrast Colors

The simple system used so far as an example is not sufficient to explain how contrast colors such as black and brown can arise within the proposed framework. Let's consider a system made out of one light source L, two objects O1 and O2 and a perceiver P. The idea that I will now formalize is that T_V is a more complex transformation than previously assumed and that it has two light sources as inputs instead of one.

In this more complex system, light coming from L interacts with both objects and then light from these objects hits the retina. Since light from two sources hits the retina, the characteristics of two light sources get transformed (distorted), but what they are transformed to is also a function of the other light source. The light coming from O1 is transformed by T_V into the (virtual) light source:

(6) $T_V(T_{O1}(T_L(L_S)), T_{O2}(T_L(L_S)))$

Notice that T_V in this case is more complex and has two parameters. In order to take into account contrast effects the light coming from O2 needs to influence the transformation (distortion) of the light coming from O1. In the same way the light coming from O2 is transformed into the (virtual) light source:

(7) $T_V(T_{O2}(T_L(L_S)), T_{O1}(T_L(L_S)))$

Notice that the roles of the parameters in (6) and (7) are switched.

The transformance of O1 is the following transformation:

(8) $T_V(T_{O1}(T_L()), T_{O2}(T_L()))$

O1's transformance is 100% correlated with the color sensation we attribute to O1.

The transformance of O2 is the following transformation:

(9) $T_V(T_{O2}(T_L()), T_{O1}(T_L()))$

Again, the transformance of O2 is 100% correlated with the color sensation we attribute to O2.

It would be a mistake to say that the system has two different transformances simultaneously. The accurate way to speak would be to say that two different physical properties of the system are each attributed to the two objects in the system. Property (8) is attributed to O1 and property (9) is attributed to O2. Property (8) of the system is what we are referring to when we talk about the color of O1 and property (9) of the system is what we are referring to when we talk about the color of O2. The sensation of color that we attribute to O1 is 100% correlated with (8) and the sensation of color we attribute to O2 is 100% correlated with (9).

The way to generalize to a more complex system with more objects is to add more parameters to T_V, one parameter for every object.

Taking into Account Distance

So far we have ignored distances between components in our system because the experimental evidence suggests that T_V is sensitive to relative energy values between wave lengths and not to absolute values. Distance does not affect the relative values of energy between wavelengths and therefore doesn't influence color sensation much. For the sake of completeness we show in this section how distances can also be modeled as transformations and thus incorporated into our framework.

As the distance between a light source and an object gets larger, the amount of energy hitting the object generally grows smaller. If the light source emits energy in equal amounts in all directions, the energy hitting an object will be reduced by a factor of $1/r^2$ relative to the original energy of the light source where r is the distance between the light source and the object.

The distance can be modeled as a transformation, T_D, which reduces the energy in each wavelength by $1/r^2$:

(10) $(T_D)^w=(1/r^2)L^w$

Transformance (and therefore color) in a system with a light source L, an object O and a perceiver P and in which the distance between L and O is D1 and the distance between O and P's retina is D2 is the transformation T_S defined below:

(11) $T_S= T_V(T_{D2}(T_O(T_{D1}(T_L()))))$

The light emitted from the light source is transformed because of the distance to the object to a light source with less energy in each wavelength. This new light source is then transformed by the object into another light source which is transformed by the distance to the retina to another light source which is finally transformed by the visual mechanism of the perceiver to yet another light source. The distance between objects in the system is also a physical property of the system so color remains a physical property of the system.

Taking into Account Angles

To this point we have ignored the relative angle of the light source to the object and that of the object to the retina. If these two angles change the characteristics of the light, they can also be modeled as transformations:

$(T_A)^w = a^w L^w$

where a^w is an infinite vector of real numbers.

Transformance (and therefore color) in a system with a light source L, an object O and a perceiver P and in which the angle between L and O is A1 and the angle between O and P's retina is A2 is the transformation T_S defined below:

(12) $T_S = T_V(T_{A2}(T_O(T_{A1}(T_L()))))$

Systems that Change Over Time

Systems in which the light source fluctuates over time or surface reflectance changes over time can also be modeled using transformations. For the sake of simplicity, let's assume that just surface reflectance changes over time. To denote this we will add an index to the surface reflectance transformation: T^t_O. This notation should be read as the surface reflectance of the object O at time t. In order to accommodate averaging effects, T_V will have several parameters that include the surface reflectance at different times:

$T_S = T_V(T^t_O(T_L()), T^{t-1}_O(T_L()), T^{t-2}_O(T_L()) \ldots T^{t-n}_O(T_L()))$

What is the spacing between the time points and how many time points are averaged is information that needs to be researched. My aim here is just to show that effects due to fluctuations over time and averaging can also be modeled as transformance. For example, the Benham Disk effect can be modeled in this way.

Other Systems of Interest

Systems with Multiple Light Sources

Let's examine how one could analyze a system with multiple light sources. Consider a system with two light sources L1 and L2, an object O and a perceiver P. Under the assumption that L1 and L2 are not coherent we obtain that a standard light source would be transformed by this system into the following light source:

(13) $T_V(T_O(T_{L1}(L_S)+T_{L2}(L_S)))$

where the + operation between light sources implies adding the energies in each wavelength respectively. In the case that the phase information needs to be taken into account the + operation is of course more complex.

In this system the following transformation, which again is a physical property of the system, is highly correlated with the sensation of color that we attribute to O:

(14) $T_V(T_O(T_{L1}()+T_{L2}()))$

Systems with No Objects

Let's move on to a system in which there is no object but in which the light from the light source hits the retina directly. Our system therefore consists of a light source L and a perceiver P. The light from the light source is transformed by the brain to yield the following light source:

(15) $T_V(T_L(L_S))$

The physical property of the system that will be correlated with the sensation of color is therefore:

(16) $T_V(T_L())$

In this system the property formulated in (16) will never be a good approximation for the surface reflectance of an object because there is no object in the system. Systems with no objects were interacted with rarely during the period in which the human brain evolved and we normally discount the possibility of such systems completely. When we see a red car on TV we believe that there is a car in front of us and that it is red because that is how the car's body reflects sunlight even though in front of us is really an array of light sources created by electrons hitting a phosphor coated screen.

The color of the real car is the same as the color of the car on TV because the following holds:

$$T_{V1}(T_{car}())=T_{V2}(T_L())$$

where T_{car} is the surface reflectance of the car and T_L is the transformation that characterizes the light emitted from the television and we assume that viewing conditions for the real car are normal and therefore the left hand side of the equation has been simplified. T_{V1} and T_{V2} represent two different states of the visual system. One while watching a car outside in normal light and the other while looking at a television in artificial lighting.

Systems with No Light Source

From the point of view of physics a system with no light source or a system with a light source having zero energy in every wavelength are the same. Consider therefore a system with a light source with zero energy in every wavelength, L_0, an object O and a perceiver P. By definition $L_0=T_Z(L)$ for every L and therefore also for L_S. The system transforms a standard light source to the following light source:

(17) $T_V(T_O(T_Z(L_S)))$

Under the assumptions that O does not emit light independently and that the brain does not distort the zero energy light source significantly then (17) can be simplified to:

(18) $T_Z(L_S)$

and the color of the system is $T_Z()$. The color sensation we have when no light (zero energy light) interacts with our eyes could therefore be associated with the zero transformation.

Systems with No Perceivers

Systems without perceivers also have transformance (and therefore color). In the case of a system with a light source L and an object O that property is:

(19) $T_O(T_L())$

Of course the color of such a system would have to be estimated or measured by other means than human color sensation.

Systems with No Perceivers and No Light Sources

In the case that a system is made of just a simple object O and nothing else (no light source and no perceiver) its transformance would be:

$T_O(T_Z())=T_Z()$ (there is no energy to reflect)

Based on our analysis one could have been tempted to say that objects don't have colors, only systems do. This is incorrect because there are systems that consist of only an object. Objects that are systems in themselves do have color, but it is a very uninteresting color. There is some way such systems transform a standard light source, namely transforming it to the zero energy light source.

Systems with Filters

In a system in which there is a light source L, an object O, a perceiver P and a filter F between the object and the perceiver, the transformance of the system will be:

(20) $T_V(T_F(T_O(T_L())))$

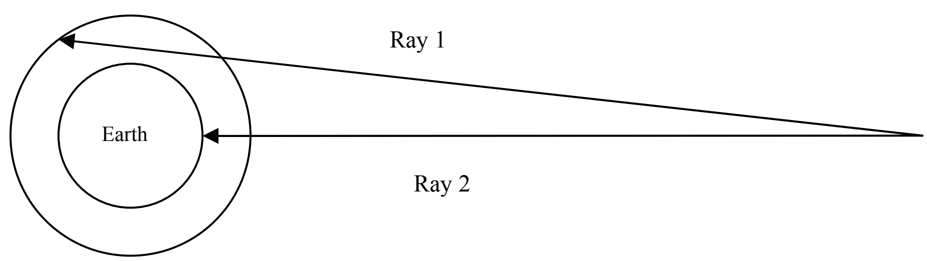
where T_F is the physical property of the filter that describes how it modifies light. The light source is modified by the object, then by the filter and finally by the visual system of the perceiver. Viewing a filter as a transformation is intuitive because a filter is defined by how it transforms the light that passes through it.

General Systems

Using the building blocks above one can provide a mathematical formulation for the transformance of any system based on the physical properties of the components of the system and the spatial relations between them. As an example I will analyze two kinds of complex systems that actually occur in real life and have philosophical interest.

Of Humans and Pigeons

The first kind of system is based on the example by Mohan Matthen involving the ability of the pigeon's visual system to sense ultraviolet light (Matthen 1999). The atmosphere scatters ultraviolet light more than it scatters light in the human visible wavelengths. Therefore, the more the light from the sun travels through the atmosphere, the more ultraviolet light is scattered relative to other light. As can be seen from the figure below, Ray 1 travels a larger distance than Ray 2 through the atmosphere.



It could very well be the case that for pigeons the sky does not look just one color. For the pigeon, the color of each piece of sky is a function of the angle of that piece to the sun. Let O1 and O2 be two pieces of the sky with angles A1 and A2 to the sun respectively. Let T_{Vh} be the transformation representing the human visual system and let T_{Vp} be the transformation representing the pigeon visual system.

Let's look at four systems:

S1: Light from the sun, piece of sky with angle A1, pigeon visual system.

S2: Light from the sun, piece of sky with angle A2, pigeon visual system.

S3: Light from the sun, piece of sky with angle A1, human visual system.

S4: Light from the sun, piece of sky with angle A2, human visual system.

The transformance of each system is as follows:

$T_{S1}=T_{Vp}(T_{O1}(T_{A1}(T_L())))$

$T_{S2}=T_{Vp}(T_{O2}(T_{A2}(T_L())))$

$T_{S3}=T_{Vh}(T_{O1}(T_{A1}(T_L())))$

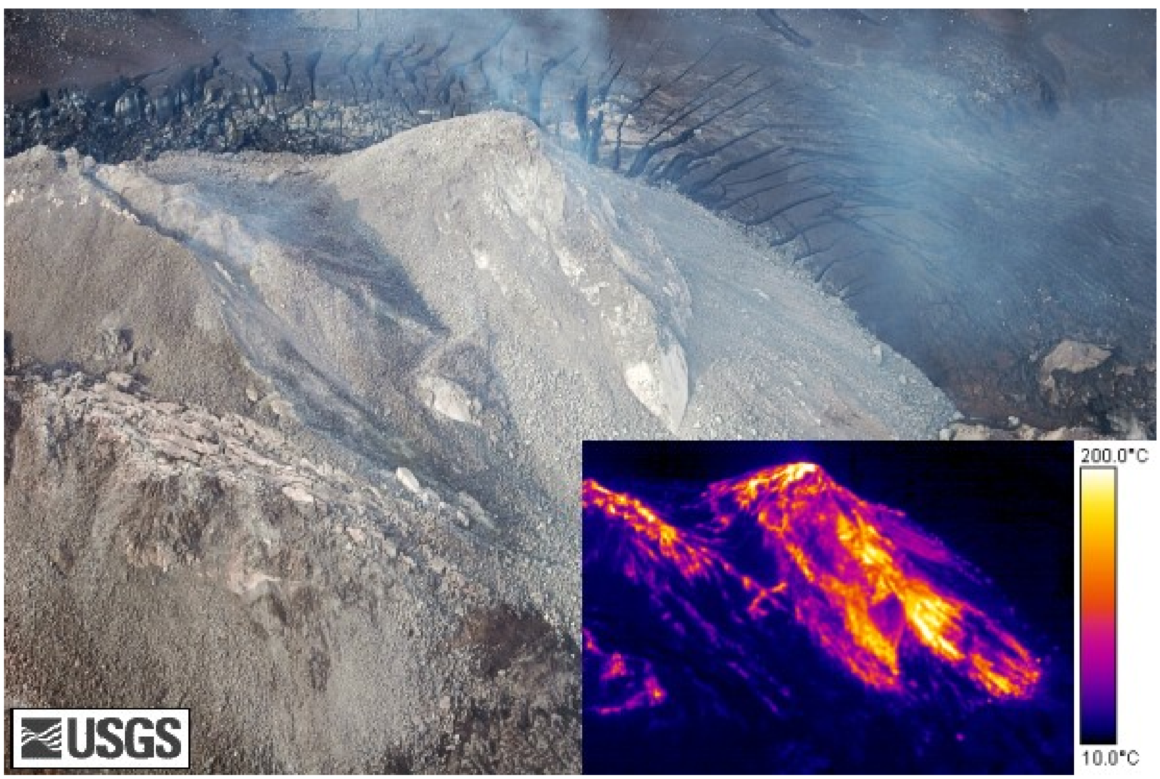
$T_{S4}=T_{Vh}(T_{O2}(T_{A2}(T_L())))$

Because $T_{S3}=T_{S4}$ the sensation of color from both pieces of the sky is the same for humans. The characteristics of T_{Vh} are such that the fact that T_{A1} and T_{A2} are different does not matter (an example of such functions will be given in the next chapter). On the other hand, since $T_{S1}\neq T_{S2}$, because of the characteristics of T_{Vp} it could be the case that the sensation of color or some other form of representation caused by the two pieces of sky may be different for the pigeon (whatever the pigeon sensations are exactly). Because the pigeon's visual system encodes and distorts light differently from the human visual system, the transformance of a system will change when a pigeon visual system is substituted for the human visual system.

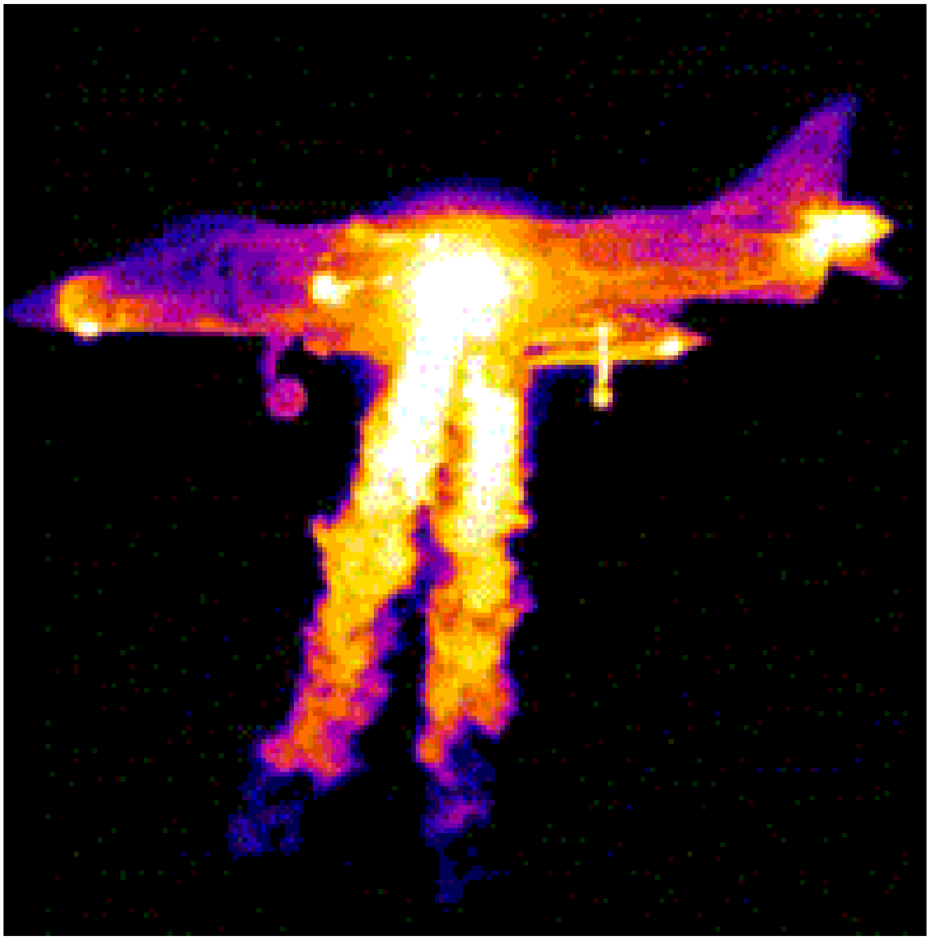
How should we speak about this, do the different pieces of sky have a different color or not? One way to speak of this is the following. We could say that relative to the human visual system, the two pieces of sky have the same color. Relative to the pigeon visual system, the two pieces of sky have a different color. I will discuss this issue in detail in the next chapter.

Surface Reflectance or Temperature?

The second example involves systems that include a forward looking infrared (FLIR) device. A FLIR is a camera that takes pictures using light from the infrared area of the spectrum, an area that the human visual system is not sensitive to. Since bodies emit infrared radiation proportionally to their temperature, FLIRs can be used to detect objects warmer than their surround even in total darkness. Below are two examples of FLIR images. One is of the volcano Mount St. Helens (regular and FLIR image) and the other is of a jet during vertical takeoff. The different colors represent areas of the picture with different temperatures.



Mount St. Helens spine as seen from the northeast. USGS Photograph taken on September 2, 2005 by Matt Logan and Jim Vallance



A FLIR device may look very much like a set of binoculars and can be used in the same way.

Let's analyze four systems:

S1: A human looking at a "blue" ball in typical daylight.

S2: A human looking at a "blue" ball in total darkness.

S3: A human using a FLIR device to look at a “blue” ball in typical daylight.

S4: A human using a FLIR device to look at a “blue” ball in total darkness.

In addition the temperature of the ball is such that it causes a red sensation when viewed through the FLIR. The FLIR device is a kind of filter. The FLIR transforms the radiation coming from the ball and can be modeled as a transformation. Unlike the human visual system that initially transforms radiation by zeroing all the energies not in the visible range (400 to 700 nm), the FLIR zeroes all energies not in a narrow band of infrared radiation (either 8-12 micrometers or 3-5 micrometers). After analyzing the energy distribution of the radiation from the infrared region, the FLIR shifts this information to the visible spectrum by emitting light in the visible region. The FLIR can thus be modeled as a transformation that initially multiplies the radiation source by a window in the infrared region of the spectrum and then shifts the energies to the visible portion of the spectrum. Let’s designate this transformation as T_F. The surface reflectance of the ball will be designated by T_O and the distortion of the human visual system by T_{Vh}. T_Z is the zero transformation.

The transformances of the four systems are:

$T_{S1}=T_{Vh}(T_O()))$ (simplified because the illuminating light is the standard one)

$T_{S2}=T_{Vh}(T_O(T_Z()))$

$T_{S3}=T_{Vh}(T_F(T_O()))$ (simplified because the illuminating light is the standard one)

$T_{S4}=T_{Vh}(T_F(T_O(T_Z()))$

$T_{S3}=T_{S4}$ because T_F takes into account only the infrared light emitted from the ball and the fact that light in the visible range is reflected from it or not doesn't matter. I'll explain this in a more rigorous manner. T_F is a combination of two transformations. The first is a window transformation of the form:

$(T_{F1})^w=e^w$ if w is in the infra red range, 0 otherwise.

e^w is the energy emitted by the light source in wavelength w.

The second is a transformation that shifts energy from longer wavelengths to shorter wavelengths:

$(T_{F2})^w=e^{(w+c)}$ where c is some constant number.

The above should be understood as T_{F2} changing the energy at every wavelength w to the one at wavelength w+c. This would be similar to shifting the graph of the energy as a function of wavelength to the left. This models the fact that the FLIR takes energies in the infra red part of the spectrum and transforms them to energies in the visible range of the spectrum. When we combine T_{F1} and T_{F2} we get:

$T_F=T_{F2}(T_{F1}())$

The transformation describing the object is of the form:

$(T_O(L))^w=c^wL^w+e^w$

where c^w is the percentage of the energy reflected in wavelength w by object O and e^w is the absolute energy emitted independently by the object O in wavelength w. Since the object emits energy only in the infra red and reflects energy only in the visible range we can rewrite T_O as:

$(T_O(L))^w = c^w L^w$ in the visible range

$(T_O(L))^w = e^w$ in the infra red range

Once we apply T_F to the above we see that the only part of T_O that is relevant is the infra red one. It doesn't matter if the object reflects any light in the visible range or not because this wavelength range is not in the window of T_{F1}. In the dark, no light is reflected from the object. But since T_{F1} ignores the reflected light anyway, the end result is that $T_{S3}=T_{S4}$.

After Images and Bumps on the Head

If a person looks at a red square for one minute and then at a uniform gray area, he would have the sensation associated with an image of a blue-green square even though no blue-green square exists in front of his eyes. He would have blue-green sensations that are not associated with any real object but with some imaginary square. These kind of effects are quite common and are called successive color interactions or after images[4].

The exact causes of after images are still being researched but there is no argument that the causes are to be identified with physical changes in the visual system such as photopigment adaptation and neural activity. The same after images happen consistently in the same viewing conditions.

I think it is likely that after images are caused by faulty encoding of light by the visual system because the encoding is not optimized for certain relatively rare conditions. For example, consider the need to encode an electrical signal that is most of the time in

---

[4] See Valois and Valois in Color readings volume 1 for a discussion of after images and their physiological bases.

the range of 0V to 5V but rarely may be out of this range and even reach 50V. The signal is encoded by an 8 bit analog to digital converter. Every few milliseconds the changing voltage is converted to a binary number between 0 and 255 and the numbers are stored on a computer. One way to encode would be to play it safe and allocate the numbers to the whole range of possible voltages. Thus, 0V is converted to the number 0 and 50V to the number 255. In general under this scheme v Volts is converted to the number:

n=int((v/50)*255)

The function int(x) returns the biggest integer lower than x. The problem with this scheme is that in most cases sensitivity is compromised. Since the signal is mostly in the 0-5V range, the signal is almost never converted to a number larger than 25. The number 25 requires only 5 bits in binary format. Most of the time, 3 bits out of the 8 bits of the analog to digital converter are underutilized.

A better conversion method could be the following:

n=int((v/5)*255)   for v<5

n=255 for v≥5

This coding mechanism is much more sensitive in the 0-5V range at the price that it is completely non sensitive above 5V. All the signals above 5V are converted to the same number and cannot be distinguished. But, in the 0-5V ranges many more voltage levels can be distinguished. For any specific application the designing engineer would have to decide what tradeoffs to make and which option to choose.

Similar tradeoffs have emerged through the evolutionary process that culminated with the current capability of organisms to encode light. In order to maintain a higher sensitivity in the most likely scenarios, sensitivity is compromised in the unlikely scenarios. In other words, in order to encode with less distortion in likely scenarios, a larger distortion in unlikely scenarios is tolerated.

I view after images as the result of such tradeoffs. Organisms have tolerated an encoding mechanism that breaks down so to speak under certain conditions because this allowed more sensitivity in the more prevalent conditions. If this is the case, even the color sensations arising from after images, can be modeled by transformance. The color of a simple system is still defined by:

$$T_S=T_V(T_O(T_L()))$$

but T_V now models a physical state of the visual system that corresponds with encoding light very poorly. Most people view after images as illusions. I think the term “cases of really poor encoding” is more accurate to describe them. An advantage of transformance as color is that it can model even these difficult cases.

There is of course no way of reconciling our common sense view of after images as illusions with the scientific view that they are not illusions but cases where encoding performs poorly. The color sensations of after images are well correlated with the physical state of the brain and therefore with transformance. After images represent the physical state of the visual system and in this sense are not illusions. I will address the question of the colors of common sense in the next chapter.

This leaves us with strong bumps on the head. In many such cases people report “seeing stars” and having color sensations if they remain conscious. I think it is obvious that these are the result of the physical changes the brain undergoes because of the trauma.

These sensations therefore also represent the physical state of the brain and can be modeled by transformance.

What Color is Red?

When a digital thermometer with an accuracy of a tenth of a degree reports that the temperature is 32.4°C, the actual temperature can be any one of the temperatures in the range of 32.35°C to 32.45°C. In this range there are an infinite number of temperature values. In the same way, a specific color sensation, say of the red kind, is correlated with and represents an infinite number of transformance values. The color red is the set of all transformance values that are correlated with the sensation of red. Unlike in the case of surface reflectance, the set correlated with red does not overlap the set correlated with green or any other color. Each transformance value is found only in one set. This issue is revisited in the next chapter.

Just as two thermometers may be manufactured and calibrated differently, a particular hue of red may represent different transformance values in different individuals. But just as there is no fundamental difference between a thermometer that represents temperature in Celsius and one that represents temperature in Fahrenheit, the same applies for individuals and color sensations. This issue will be discussed further in chapter 4.

What is Transformance Then?

Intuitively, transformance is a very general abstract physical property that describes how a system transforms or modifies light or information that characterizes light. It is quite difficult to define in words complex physical properties without leaving out some aspects of the definition. That is why understanding in many sciences is achieved by mathematical formalization. Though the complete characteristics of transformance cannot be captured using words alone, given any system, I have shown how to represent its transformance using transformations of light sources. Hopefully, the many examples above have provided the reader with a good enough sense of transformance and how it relates to surface reflectance and color.

Chapter 3

Transformance is Color

In this chapter a philosophical defense of transformance as color will be given assuming that some form of color physicalism is true. As part of the defense, surface reflectance will be extensively evaluated against transformance. Objections of a more general nature against all forms of physicalist accounts of color and especially the issue of color sensations and how they relate to color will be discussed in the next chapter.

Transformance as defined in the previous chapter is a very general abstract physical property that describes how a system transforms or modifies light or information that characterizes light. If viewed as a relational property, it can be attributed to an object. As an intrinsic property, it is a property of a system. One can speak of an object having a certain transformance relative to the light and the visual system or alternatively one can say that the system that includes the light, object and visual system has a certain transformance. From a scientific point of view, both ways of speaking are identical, just as viewing weight as a relative property of an object or as an intrinsic property of a system makes no scientific difference. Both are different ways of speaking about the same things and the mathematical equations that model the relations between them. From a philosophical point of view, viewing transformance as a relative property of objects is more promising since it is more in tune with our common sense view of colors. This is the approach I will take.

My view then is a physicalist relationist one. It is different from the physicalist view described in the literature in substance but is different from relativist approaches described in the literature only in methodology. It is different from surface reflectance physicalism because it advocates transformance as color and not surface reflectance (or productance). My approach turns out to be relationist because transformance is a relational property of objects. Having undertaken the project of identifying the physical property that is best correlated with color sensation without any preconceived notion about its form, I subsequently discovered that it can be viewed as a relational property of objects. This is a welcome "side effect" but not something that was initially a goal of the project. In contrast, the advocates of color relativism initially define color as a relational property of objects and either ignore the issue of the physical basis of this property or make their definition conditional on such a property being found. For example Brian McLaughlin defines color as:

Relativized Colors: Redness for a visual perceiver of type P in circumstances of visual observation C is that property which disposes its bearers to look red to P in C, and which is had by everything so disposed (McLaughlin 2002).

McLaughlin identifies color with the basis for the disposition to "look red" and therefore his analysis is a functional one. Color is according to McLaughlin the physical property (if it exists) that "endows" objects with the disposition: "red things will be disposed to look red because they have the property of redness". Redness is not a disposition but the underlying (physical) property that is the cause or basis of the disposition. In the next chapter I will discuss the issue of how McLaughlin views the relation between color and color sensation or in his terminology "what it is like to see color".

In retrospect, the fact that transformance can be viewed as a relational property should not be a big surprise. Both color relationists and myself are primarily motivated by the issue of perceptual variation and a relativist account does solve the problem. McLaughlin is explicit about this concern: “Radical relativization seems to me the right response to the problem of standard variation. Rather than embracing colour irrealism, we should handle the problem by radically relativizing the colours; that is, we should relativize them to kinds of perceivers and circumstances so specific as to leave no room for variation in colour appearance.”

Transformance is fully compatible with McLaughlin’s definition above if viewed as a relational property. It is I believe “that property which disposes its bearers to look red to P in C, and which is had by everything so disposed.” (McLaughlin 2002) My theory implements the “radical relativization” McLaughlin has in mind by taking into account in the definition of color the physical properties of the perceiver’s visual system and the viewing circumstances.

Transformance as a property of objects is also in agreement with Jonathan Cohen’s definition of a relational property: “Suppose x is something red; then, as we modify things other than x, and thereby modify the relations x bears to other things, will x (necessarily) continue to be red? If so being red is non-relational; if not it is relational.” (Cohen 2005)

Whatever the initial motivation, the end result is the same in that I share the relationist views about color and therefore need to defend my view against objections to color relativism/relationism.

As for the physicalist side of the project, Byrne & Hilbert (B&H) have deftly weathered a twenty year barrage of criticisms and have immensely helped clarify the challenges facing any physicalist theory of color. Their work made me realize that even if surface reflectance is not color, surface reflectance must play an important part in any physicalist theory of color. Surface reflectance is an important part of transformance, and in certain conditions, surface reflectance is a good approximation for transformance[5]. I view my theory as a refinement or extension of their theory. Thus, most of my argument with B&H is about which property, surface reflectance or transformance, provides a better fit with the scientific data and is more explanatory from a scientific perspective.

I will begin by addressing the issue of perceptual variation.

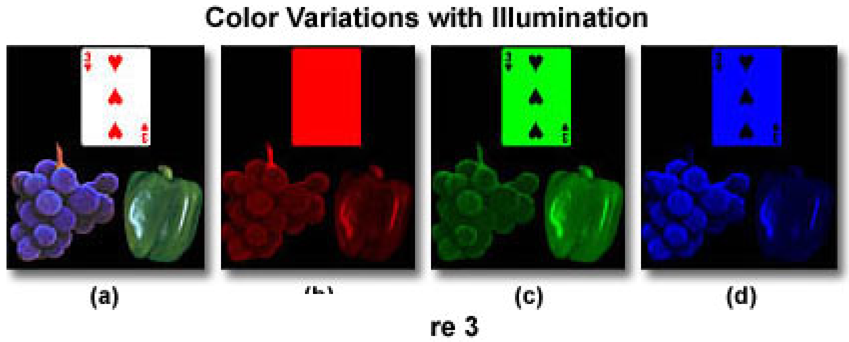
Perceptual Variation

There are several kinds of variations that a theory of color needs to explain. The different variations have been extensively dealt with numerous times in the literature. I will therefore just provide a short summary and an example of each type of variation.

Variations Related to Illumination

In different lighting conditions, the same objects cause different color sensations. Objects in moonlight cause different color sensations than objects in sunlight. The dress that caused a certain color sensation in the store causes a different color sensation at home. The color sensations caused by the grass in the shaded part of the lawn are different from the ones caused by the grass in the sunny part.

[5] See section "The Relation between Transformance and Surface Reflectance" in Chapter 2



The same scene illuminated by (a) white (b) red (c) green and (d) blue light

The theory that color is transformance explains well why color changes with illumination. Illumination is one of the components that make up transformance. In a simple system color is modeled by the complex transformation: $T_S=T_V(T_O(T_L()))$

T_L is the transformation that characterizes (models) the illumination. If T_L changes that could also change T_S and explain the variation in color due to illumination.

Variations Related to Human Visual Systems

Even among humans there is substantial variation in the perception of color. Under the same viewing conditions, people report different color sensations associated with different objects. Among "normal" viewers for example, unique green is distributed in the 490-520 –nm range. In addition, there are various anomalous variations that are classified colloquially as "color blindness". For example, certain kinds of color blind people have a problem identifying the number below.

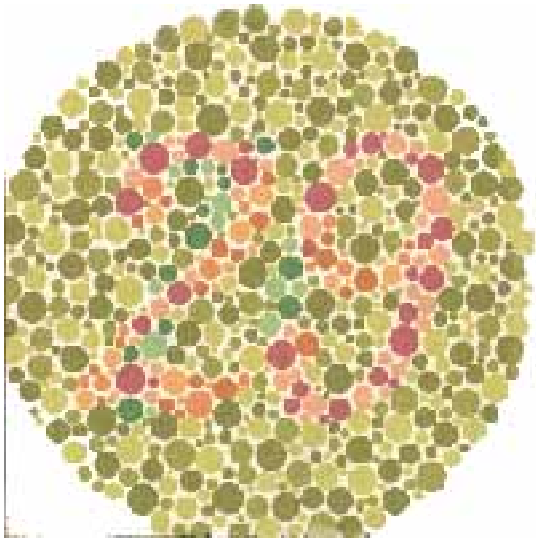


Figure 1 – The number 29, red on green background

Visual systems are modeled by the T_V component of transformance and therefore transformance physicalism explains variations between perceivers with different visual systems.

Variations Related to Surround

The same object may cause different color sensations as the surround of the object changes. For example, brown and black sensations can only be generated if an object has a certain background. An object generating a brown color sensation will generate an orange one once it is viewed through a reduction tube. The example below demonstrates one surround effect (Brown and Donald 1997).

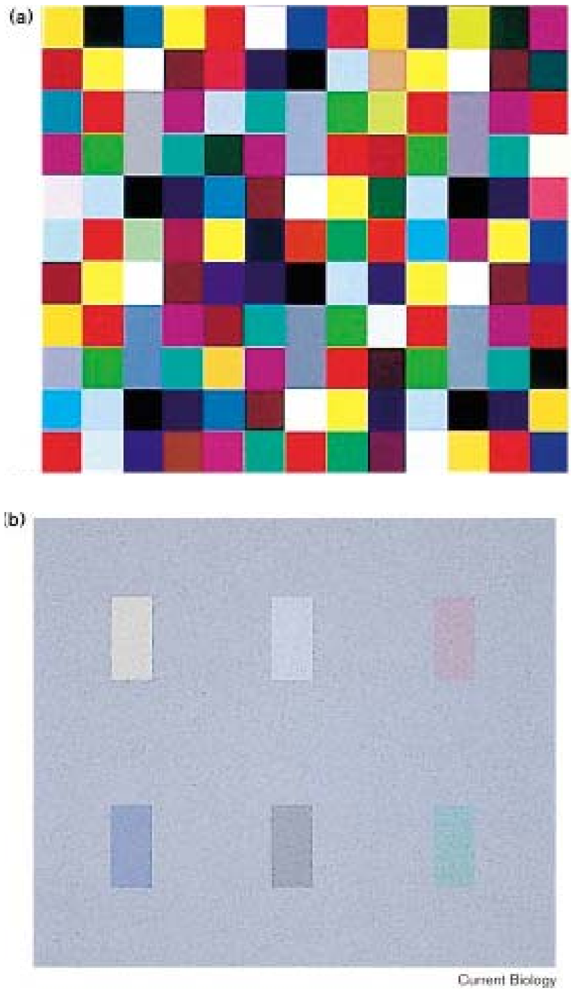


Illustration of stimuli. Two surrounds, with the identical space-averaged luminance and chromaticity, but different variances around that average, produce very different effects on the perceived colors of six embedded test rectangles. The six rectangles were predominantly gray tinged with yellow, white, red, green, black and blue (clockwise from upper left rectangle). (a) In a high-contrast richly colored surround, the six rectangles all appeared grayish and were difficult to discriminate from each other. (b) In the uniform gray surround, the six rectangles, which were each physically identical to the corresponding six rectangles in (a), appeared to be much more richly colored. (Brown and Donald 1997)

The theory that color is transformance can easily accommodate surround colors as was shown in Chapter 2. This is done by allowing T_V to be a transformation that has more than one parameter.

Metamers

Two objects are metameric under a given illuminant if they cause the same color sensation ("look the same color") even though their surface reflectance is completely different. Under a different illuminant the objects may not be metameric anymore and may cause different color sensations ("look different"). The problem of metamers arises because in cases where we expect to find perceptual variation, it is lacking.

Since transformance physicalism does not identify color with surface reflectance, it doesn't have any problem dealing with metamers. In fact, the theory that color is transformance explains metamers. From the experimental data it is well known that T_V is not a one to one transformation and that it transforms many values to the same value. Therefore for many combinations of the same T_O (object reflectance) and many different T_L (lighting conditions), T_V (the distortion during the encoding of light) will output the same result thus explaining the existence of metamers.

As a concrete example take the infinite family of functions defined below:

$f^r(x)=rx$ for $0<x<1$ where r is a real number between 0 and 1

$f^r(x)=0$ everywhere else

If r is 0.6 this designates the function $f(x)=0.6x$ for $0<x<1$ and $f(x)=0$ elsewhere. If r is 0.893 this designates the function $f(x)=0.893x$ for $0<x<1$ and $f(x)=0$ elsewhere.

Now, consider the function $g(x)=int(x)+1$. The function int(x) returns the first integer smaller than x. The last step in the example is to look at the composite function $g(f^r(x))$.

By definition:

$g(f^r(x))=g(rx)=int(rx)+1=1$ for $0<x<1$

$g(f^r(x))=g(0)=int(0)+1=1$ elsewhere

For all members of the family $f^r(x)$, the composite function $g(f^r(x))$ is the same function, namely, $f(x)=1$. Because of the specific characteristics of $g(x)$, the composite function $g(f^r(x))$ is the same one no matter what r is. Loosely speaking, because of the characteristics of $g(x)$, all the functions in $f^r(x)$ are "metameric". The transformation that describes how the visual system distorts light while encoding it (T_V) has similar characteristics and that is one reason why there are metamers.

Another reason that metamers occur is the following. Let's assume (just for a moment) that the visual system does not distort light at all during the encoding process. This means that T_V is the identity function and the transformance of a simple system made of a light source, object and visual system can be simplified to:

$T_S=T_O(T_L())$

From a mathematical stand point, for a given T_S there is an infinite number of different pairs of T_O (surface reflectances) and T_L (illuminants) such that $T_S=T_O(T_L())$. All these pairs are thus potentially metameric. If we relax the constraint that there is no distortion but only assume that the distortion is constant across viewing conditions, the same result is obtained.

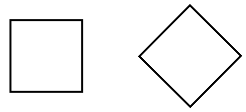
Surface Reflectance Physicalism, Transformance Physicalism and Perceptual Variation

Byrne & Hilbert (B&H) are the leading proponents of the theory that color is surface reflectance. In this section I will contrast how they address the issues associated with perceptual variation with how transformance physicalism deals with these issues.

B&H employ two strategies to solve the problems associated with perceptual variation. The first strategy is to make a distinction between the conditions necessary for perception and what is perceived. They employ this strategy in the case of contrast colors:

"However, if we avoid the confusion mentioned in section 1.3.3 above, between the conditions necessary for perception and what is perceived, there should be no temptation to think of brown as being a relational property different in kind from other colors. The conditions necessary to see an object as brown involve a relation between the object and its surround, but this is perfectly compatible with our claim that brown is a type of reflectance." (Byrne and Hilbert 2003)

The second strategy B&H use is to define color as a set of surface reflectances. All surface reflectances that are members of the set are of the same color. Furthermore, according to B&H the sets may overlap[6]. (Byrne and Hilbert 1997) Just as it is possible that something is both square and diamond, they claim it is also possible for it to be blue and bluish green at the same time.



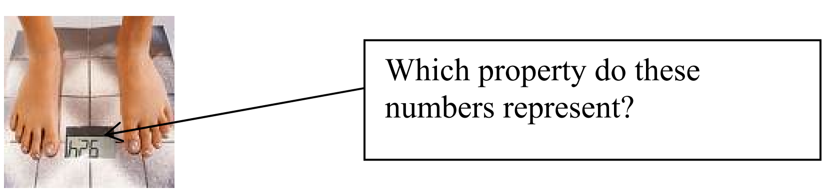
---

[6] The position that both sets could be true is not endorsed anymore by B&H.

B&H admit that they cannot identify which reflectance types are represented by which color sensations but reject the notion that this is a problem for their theory. Furthermore, they point out that color relativists cannot accomplish this task either (Byrne and Hilbert 2003). B&H use the second strategy to suggest a solution to the problem of determinate colors and metamers:

"Notice that even if we ignore metamerism, there is already a problem with determinable colors – red, green, purple, and so forth. Typically two purple objects will have different reflectances. The solution to this problem is clear: we can identify the determinable colors with reflectance types (or sets of reflectances) rather than with the specific reflectances themselves. For example, the property purple, on this modified account, is a type of reflectance rather than a specific reflectance. As a bonus, this proposal also solves the problem of metamers (and so it is not really an additional problem): both determinable and determinate colors are reflectance-types. Metameric surfaces are, according to the revised theory, the same in determinate color in spite of their physical differences." (Byrne and Hilbert 2003)

A better understanding of the differences between B&H's position that color is surface reflectance and my position that color is transformance can be found in asking questions about physical properties simpler than color for which we have a better intuitive feel. The first question is: What physical property do the numbers on a bathroom scale represent?



Let us consider two relevant answers. The first is that the number represents the mass of the object on the scale and the second is that the number represents the weight of the object on the scale. The mass of the object is the amount of matter and energy in the object. The weight of the object is the force applied on the object in a gravitational field. The mass of an object is the same on the earth and on the moon, but its weight on the moon is 1/6 of its weight on earth because the gravitational field of the earth is 6 times stronger than that of the moon. Mass is an intrinsic physical property[7]. The mass of an object does not depend on any other entity outside the object. Weight on the other hand is a relational property. The weight of an object on earth depends on the mass of the earth or in other words, is relative to the mass of the earth. The weight of an object on the moon is relative to the mass of the moon.

Returning to our question, let's say Mary is an advocate of the theory that the number on the bathroom scales represents mass. Will confronts Mary with new scientific evidence that on the moon, when the very same object was put on the very same scale the result was a number 1/6 as large being displayed. Since the mass of an object is the same on the earth and the moon, Will argues that the bathroom scale cannot be representing mass.

Mary ponders this question and responds to Will with two points. The first point is that Will is confusing the property represented with the conditions necessary for representing it correctly. The bathroom scale number does represent mass but the necessary condition is that it must be located on earth. My view is that Mary's first response is correct but not complete. The more complete answer in my opinion is that

[7] In Newtonian physics

bathroom scales represent weight, but in a constant gravitational field weight is only a function of the mass of the body on the scale and therefore it seems that the numbers represent mass. It is also a scientifically more satisfying answer because it is more general and explanatory. For example, it provides a sound scientific explanation why there are variations in the number the scale displays for the same object in different gravitational fields and also predicts what the variations would be. Mary's theory is neither predictive nor explanatory. It cannot predict what the variations will be and also cannot explain why they occur.

Mary's second response is that actually each number on the bathroom scales represents a set of masses. The number 10 perhaps represents the set of masses {10, 60} because on earth it would represent 10 mass units and on the moon 60 mass units. What about the number 60 asks Will? It would represent the set {60, 360} answers Mary. Will persists, when we see the number 60 on the scale, which of the two sets does it represent? Mary answers that it represents both sets. After all, something can belong to two categories at the same time, just as something could be both a square and a diamond.

Will persists and asks Mary to identify which bathroom scale numbers represent which masses. Mary admits that she can't do that but thinks that it is not a fair criticism. First of all, because Mary can't identify the corresponding sets, that doesn't mean they don't exist. If Professor Plum is found dead in the conservatory with a knife in his heart, we may not know who killed him but we know he was murdered by someone. In the same way, the correct matching between the sets exists, we just cannot identify it. In addition, Mary argues that Will's theory is also not up to the challenge and therefore it is not a fair criticism against hers.

Will answers that he too can't assign the sets to masses because it is impossible to do in any way that is not ad hoc. The Professor Plum analogy is not apt because for any mass, and any number on the scale, one can find a gravitational field such that if an object with that mass is put on the scale in that gravitational field, the scale would show that number. For example, say you have an object with the mass of 10 and you want the scale to show the number 10, put the object on the scale on earth. You want the number to show 5,000? Put it on a scale on Jupiter. You want some other number to be displayed? There will always be some place in the universe where one can weigh the object and get the desired number displayed. By varying the gravitational field, any mass can correspond to any number. Therefore, all numbers represent all masses. Mary's proposed sets are not just overlapping in some parts. They are identical and include all possible masses. In the case of Professor Plum we assume that one person murdered him. Therefore, someone murdered him. It is not true however that all the people in the world murdered him. If that were the case, there would just not be any way to identify THE person that murdered him and we could not reach the conclusion that one person (someone) murdered Plum[8]. It wasn't someone, it was everybody. Since every number represents every mass there is no non-arbitrary mapping between numbers displayed by the scales and sets of masses. Will's theory also cannot identify which sets of masses correspond with which numbers on the scale because it is just not possible.

On the other hand, the correspondence between weight and the number displayed by the scale is very simple. In theory, each number corresponds to just one weight. In practice, because the precision of the bathroom scale is finite, each number corresponds

---

[8] Is there a good answer to who was THE murderer in Agatha Christie's "Murder on the Orient Express"?

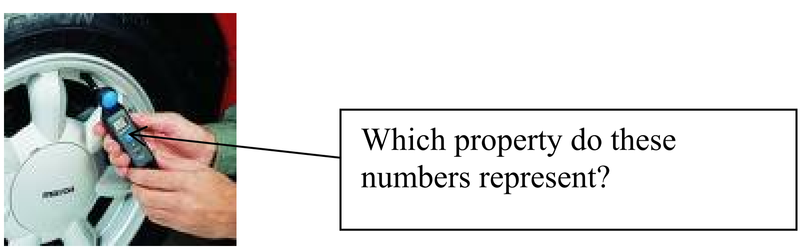
with a set of weights. For example, if the bathroom scale displays 83.1, this corresponds to the infinite set of weights {w: 83.05<w<83.15}. Note though, that unlike Mary's case, the sets are not overlapping.

In this case, weight, a relational property, rather than mass, an intrinsic property, is a much more probable candidate for what the numbers on the bathroom scales represent. The same I argue holds for color. All the arguments above in support of mass that Mary provides are just paraphrasing B&H's arguments in support of color as surface reflectance. For the very same reasons that weight is more appropriate than mass in the case of the scales, transformance, a relational property, is more appropriate than surface reflectance, an intrinsic property, in the case of color.

Another example of a relational property is velocity. Each object has an infinite number of velocity values associated with it, depending on what frame of reference the velocity is measured in. The velocity of the passenger relative to the train is different than his velocity relative to the platform that is different from his velocity relative to the sun and that is different relative to his velocity relative to the center of the galaxy etc. If one attempts to assign a set of velocity values to each object, the sets will be identical and infinite. Would B&H say that all objects have the same velocity? This does not make sense. The right way to speak is to say that the velocity of an object is relative to a frame of reference. Relative to the same frame of reference, two objects may or may not have the same velocity. One can also say that one object has the same velocity relative to one frame of reference as another object has relative to another frame of reference. In the same way objects have colors relative to the system that they are in and should be considered relative properties just as weight and velocity are. Just as velocities in two different frames of references can be compared, colors of objects in two different systems may be compared. This is issue is addressed in more detail in the section “Apples and Oranges”.

Again, we can see why the Professor Plum analogy is not apt. For simplicity’s sake let’s consider the speed property of objects which is just the magnitude of the velocity and not its direction. B&H can argue that they know that the speed of an object is somewhere in the range 0 to the speed of light and therefore an object must have some speed[9]. This is the analogy implied by the Professor Plum argument. The problem with this argument is that for it to be correct there must be some privileged frame of reference. Otherwise, the object doesn’t have some speed. It has all speeds depending on which frame of reference one chooses. There is no privileged frame of reference in the case of speed and I think the same holds for color. That is why it is more likely that color is a relational property like transformance and not an intrinsic property like surface reflectance.

Let’s look at another example which B&H raise. What physical property does the number on the so called “pressure gauge” represent (Byrne and Hilbert 2004)?



In this case there are three competing theories. Patrick argues that the number represents the amount of air in the tire. He points to a set of experiments that show that as

[9] This is the reply B&H give Cohen (Byrne and Hilbert 2003).

you add or remove air from the tire the numbers displayed change accordingly. Teresa on the other hand says the number represent the temperature of the air in the tire. She points to a different set of experiments that show that as you vary the temperature of the tire, the numbers displayed vary accordingly. Sam on the other hand formulates a theory that takes into account the two sets of experimental data and suggests that the numbers represent a third physical property which is obtained by multiplying the amount of air by its temperature[10].

Patrick argues that Teresa's data are not important because the condition for representing the amount of air is that the temperature is constant. Teresa argues that Patrick's data is not relevant because the condition for representing temperature correctly is that the amount of air is constant. I prefer Sam's argument that the number represents the multiplication of the amount of air and the temperature and that when the temperature is constant it seems as if the numbers represent the amount of air and when the amount of air is constant the numbers seems to represent the temperature. Sam's theory is more predictive and explanatory and therefore more scientifically sound.

The argument that color is surface reflectance is logically sound. If we assume that color is an intrinsic property like surface reflectance, then B&H's arguments are correct. If speed were an intrinsic property of objects then there would have to be some value in the range 0 to the speed of light that would be THE speed of the object. However, I believe the question is not whether if we assume that color is surface reflectance we can provide a consistent argument that it is so. B&H have answered that question in the

[10] The temperature units need to be Kelvin for this.

affirmative. I think the right question to ask is which of the two following theories better explains the scientific data:

1) Color is surface reflectance, an intrinsic physical property of objects

2) Color is transformance, a relational physical property of objects

The question is first a scientific one and only then a philosophical one. Since theory 2 is much more explanatory and predictive, I think it is preferable. The first theory cannot explain and predict the perceptual variations and lack thereof. On the other hand, if color is transformance, this explains the variation in color sensations in a scientifically pleasing way.

One additional clarification is required regarding "necessary conditions for detection". In discussing conditions necessary for the detection of color B&H also provide the following example:

"In order for a household thermostat to detect that the temperature is below 65°F, the thermostat dial must be set correctly. It does not follow that the property of being below 65°F is in any interesting sense dependent on, or relative to, thermostats or their settings. No one is likely to make this mistake of confusing temperature with conditions necessary for the detection of temperature. But an analogous mistake is for some reason often made in the case of color." (Byrne and Hilbert 2003)

I agree with all B&H say in the paragraph above except for the last sentence. B&H are conflating two different types of conditions required to measure physical properties. The first type of condition is that the measuring equipment functions correctly (Type I). The second type of condition is that other physical properties are kept constant while the property of interest is being measured (Type II). For example, when measuring the temperature of a gas it is not enough that the thermometer is reliable, it is also important to keep the pressure of the gas times its volume constant during the measurement. Otherwise, the measurement is meaningless since the temperature of a gas changes when its pressure times its volume changes as can be seen from the ideal gas equation:

$$pV = nRT$$

where p is the pressure, V is the volume, n is the number of moles of gas, R is the gas constant, and T is the temperature in Kelvin.

In the passage quoted above B&H describe a condition of Type I for which there is no argument that it is a genuine condition necessary for the correct measurement of temperature. If the thermometer doesn't work, the temperature of a gas cannot be determined. B&H use Type I condition to reach a similar conclusion regarding cases that involve a condition of Type II, but in those cases the conclusion does not follow.

To see why, let's apply the argumentation above to the problem of contrast colors. Regarding contrast colors such as brown B&H explain: "there should be no temptation to think of brown as being a relational property different in kind from other colors. The conditions necessary to see an object as brown involve a relation between the object and its surround, but this is perfectly compatible with our claim that brown is a type of reflectance." (Byrne and Hilbert 2003)

Let's paraphrase this statement to deal with weight instead of surface reflectance: "There should be no temptation to think of a weight of 40 Kg as being a relational property. The conditions necessary to measure the weight of an object as being 40 Kg involve a relation between the object and the gravitational field surrounding it, but this is perfectly compatible with our claim that weight is an intrinsic property."

Now let's paraphrase this again using speed: "There should be no temptation to think a speed of 50 miles/hour is a relational property. The conditions necessary to measure the speed of the object as being 50 miles/hour involve a relation between the object and a frame of reference but that is perfectly compatible with our claim that speed is an intrinsic property."

When in order to measure some property accurately we need to hold other physical properties constant, this is strong evidence that the property depends on those other properties and is relative to them. It is a very different case from that of a non-functioning measuring device and does not warrant reaching the same conclusions. I conclude that based on the scientific evidence and the analogy with other relational properties that it is more likely that contrast colors are relational physical properties and are dependent on their surround.

This discussion also highlights one point of potential disagreement that my theory may have with Brian McLaughlin's theory of color. McLaughlin's view is that color is a relational property but believes that it may still turn out to be an intrinsic property of objects: "It remains open, however, that rather than fundamental, emergent properties of certain physical whole, colors are in some sense derivative from microphysical properties of certain physical wholes. Indeed on the evidence, if there are colors, they are so derivative, for the properties that affect the behavior of light are microphysical properties or properties derivative from them. Let us call properties that are derivative from microphysical properties, 'physical properties'. Given that redness, if it exists, is such a derivative property, it is a job for vision science to identify the physical property in question and, thereby, to locate the place of redness in nature." (McLaughlin 2002)

I think the view that color is a relational physical property is not compatible with the view that it can be derived from the microphysical properties of objects (it can be derived though from the microphysical properties of systems). The scientific evidence in my opinion shows that color is more like weight or speed rather than mass. Just as weight and speed cannot be derived from the microphysical properties of objects, neither can color.

Lights, Filters and Volumes

What is the color of the sun? That of a flame in the fireplace? What is the color of the car on the TV? The lights on the Christmas tree? The one thing we can be sure of is that these colors are not related to surface reflectance. The color of these objects is associated with how they emit light rather than how they reflect it.

What is the color of the sky, a glass of wine, maple syrup and the water in the swimming pool? Again, these colors cannot be associated with surfaces and therefore do not involve surface reflectance.

What is the color of sunglasses? Light is not just reflected from them, it passes through them. It also seems that the color of objects changes when we put on sunglasses. How does the color of the sunglasses influence the color of objects?

As was shown in the previous chapter, transformance is a general enough property so that it is applicable to all the cases above. Every physical system transforms light in some way, and that is the reason transformance can be defined for any physical system.

Surface reflectance on the other hand cannot be defined for the cases described above. In order to expand their theory to objects emitting light, volumes, filters and what I call general systems, B&H define a new property which they call productance. The general idea is to characterize the object in terms of the light leaving it instead of just the light it reflects. In the case of an object emitting light just in one wavelength w, and reflecting light also just in that wavelength, if it is illuminated by light whose intensity at w is i_1 then the productance of the object is $(ri_1+e)/i_1$ where r is the percentage of light reflected at wavelength w. As can be seen, when e is 0 (the object doesn't emit light), productance reverts back to reflectance and that is probably why B&H define it in this way. Productance though is relative to an illuminant as can be seen from its definition. When the light illuminating the object changes, the productance of the object changes (assuming e is not zero).

B&H explain why they believe productances are illumination independent: "Although productances are relative to illuminants, it is important to stress that the productance of a surface is illumination independent– that is, independent of the actual illuminant. The surface of a stoplight or tomato has a certain productance relative to an illuminant I, and it has this productance independently of the light that is in fact illuminating it. Hence, it has a certain type of productance independent of the actual illumination."

I will paraphrase what B&H say using another relational property, weight: "Although weight is relative to a gravitational field, it is important to stress that weight of an object is gravitational field independent – that is, independent of the actual gravitational field the object is in. A book or a tomato has a certain weight relative to gravitational field G, and it has this weight independently of the gravitational field that is in fact acting on it. Hence, it has a certain weight independent of the actual gravitational field."

Just as it would be inaccurate to say that the weight of an object is independent of the gravitational field that it is in, it is also inaccurate to say that productance is independent of illumination. The mistake is the result of B&H moving from an intrinsic property, surface reflectance, to productance which is not an intrinsic property. The analogy is between surface reflectance and mass on one hand and productance and weight on the other. Mass is not dependent on the gravitational field the object is in. The mass of an object is the same on the moon as it is on earth. Weight is dependent on the gravitational field the object is in. The weight of an object on the moon is one sixth its weight on earth. The surface reflectance of an object is not dependent on the illumination. The productance of an object is.

Regarding weight, one can say that because of the object's mass, it will have a specific weight at any given gravitational field. This does not make weight an independent property of the same status as mass. The weight of an object undoubtedly depends on the gravitational field it is actually in. In the same way, productance is a relational property of objects and depends on the actual illumination even though surface reflectance is independent of the illumination.

Another problem with productance is identified in several comments in reply to B&H (Byrne and Hilbert 2003). Recall that in the simple case of an object emitting monochromatic light of wavelength w with intensity i_1 and reflecting r percent of the energy in wavelength w (0 in all others), the productance is defined as $(ri_1+e)/i_1$. The problem with this definition is that as i_1 approaches zero, the productance approaches infinity and thus is ill defined. B&H don't view this as a problem since they maintain that this problem also holds when reflectance is defined. Reflectance in wavelength w is the ratio between the energy reflected in that wavelength and the energy of the illuminant in that wavelength: $r=i_{refelected}/i_1$. B&H claim that in both cases as i_1 approaches zero, there is a problem and therefore productance should not be singled out.

B&H's mistake is to assume that there is a problem in the case of reflectance. There is no problem because both the numerator and the denominator in the definition of reflectance go to zero as i_1 approaches zero and therefore the result converges to the actual finite reflectance and not infinity. In mathematical notation:

$$\lim_{i_1 \to 0}\left(\left(\frac{i_{reflected}}{i_1}\right)\right) = r$$

While:

$$\lim_{i_1 \to 0}\left(\left(\frac{ri_1 + e}{i_1}\right)\right) = \infty$$

Because e is not zero, the numerator does not also go to zero in the case of productance.

A famous example of limit where both the numerator and denominator go to zero but the limit converges is the following:

$$\lim_{x \to 0}\left(\left(\frac{\sin(x)}{x}\right)\right) = 1$$

Is Transformance Always a Mind Independent Property?

As was shown in the previous chapter, transformance can be defined for any physical system including systems with no perceivers. However, the question remains whether in those systems that do include perceivers, transformance is also a mind independent property. This question can be rephrased as whether the visual system is an essential part of the mind or not. My view is that it is not. Just as a person remains himself when his arm is amputated even though his arm is a part of the person, one can extract a person's visual system to a great extent without that person losing his mind or becoming someone else (there is no need to go into the gory details). The visual system interacts with the mind but in my view it is either external to it or is a non-essential part of the mind. A rough and ready test would be to ask if a person is deemed to have "lost his mind" if his visual system is severely impaired. I think the answer is no. The question though is not a trivial one because many other kinds of injuries to the brain or components of the mind may lead us to the conclusion that a person has indeed lost his mind. The literature is rife with examples of people who have changed significantly and/or "have lost their minds" due to head injuries or dementia.

Also, the scientific evidence is clear that visual systems evolved before the mind did. Flies and bees have complex visual systems but are not likely to have minds. Flies and bees do have cognitive systems, but it is unlikely that they are conscious. An organism may have a visual system but not a mind, or a mind and not a visual system. It therefore seems quite possible that the mind and the visual system are two distinct things[11]. If that is indeed the case, then transformance is always a mind independent property and therefore so is color. Even if minds did not exist, colors would still be around[12].

Is Color Just One Property?

Both Mohan Matthen and Brian McLaughlin address the possibility that color may be more than one property (Matthen 2009, McLaughlin 2002). Matthen calls this possibility the "disunity of color" while McLauglin talks about the "problem of common ground".

Their worries are not identical. Matthen is more bothered with cases such as the pigeon's in which color may represent direction instead of surface reflectance. As was shown in the previous chapter, transformance is general enough to address such issues. The way I see it, both the human and pigeon color systems represent transformance but the pigeon's representation is fine enough to allow it to infer directions while our representation is rougher making it impossible for us to do so.

---

[11] For a more nuanced discussion and overview of prevailing philosophical thought on this issue see (Allen 1997)

[12] It could very well be though that color sensations are mind dependent. This issue will be discussed in more detail in the next chapter.

The case of the pigeon is analogous to that of a person that has a very accurate weighing device and measures with it some known mass in different locations on earth. With such a device differences in weight for the same object will be measured even on earth. The measurement still represents the weight of the object but from the differences in weight in different locations the person can infer which regions of the earth are more or less dense than others while another person with a less accurate weighing device would not have the information to make the inference. Different organisms infer different things about their environment using color. That much is undisputable. However, the fact that color can be used as input to different inference processes does not warrant the conclusion that color is not one property. I suggest that at the input to all these processes is a representation of one property, transformance, and that is what color is.

The less an organism distorts the light in the process of encoding it, the more inferences the organism can make because it loses less information. People with normal vision can infer that the number 29 is depicted in Figure 1 while color blind people cannot make the inference because color blind people lose more information while encoding light. The hypothetical case in which a visual system encodes light without loss is modeled by setting T_V to the identity transformation. Since there is no distortion, the light is not changed in the process of encoding and that is why T_V is the identity transformation. In the general case of a simple system the transformance is: $T_S=T_V(T_O(T_L()))$. In our hypothetical case since T_V is the identity transformation, this can be simplified to:

$$T_S=T_O(T_L())$$

which is still a value of the transformance property.

McLaughlin's worry is different than that of Matthen: "The really serious problem for colour-realism is, I believe, the problem of common ground. If anything is disposed to look red to P in C, there will be many heterogeneous chemical properties that dispose their bearers to look red to P in C; many heterogeneous chemical properties will be metameric with respect to a given type of perceiver P in a given circumstance C."

McLaughlin's concludes that if the problem is to be solved it will be solved by an abstract property. I propose that this property is transformance. What is common to all objects that look red to perceiver P in circumstances C is that their transformance as modeled by a complex transformation is a member of a certain set of transformations.

Arguments Against Color as a Relational Property

The difference between transformance physicalism and the other relativized theories of color is the physicalist underpinning of the theory. As I will show, the fact that color is associated with a well defined physical property helps to overcome objections to relational theories of color with relative ease. This is the case because many well understood physical properties are relational and one can use analogies from their (relatively) uncontroversial metaphysics and their use in language to understand the case of color.

Jonathan Cohen has addressed most of the objections to viewing color as a relational property in "Color Properties and Color Ascriptions: A Relationalist Manifesto." In what follows I rely on the list of objections he has compiled and where I agree with him, also on his answers. Since my theory is a physical one, when we are in agreement, I reproduce his arguments using examples that are familiar physical properties.

Too Much Color

If colors are relativized to systems, doesn’t an object have an infinite number of color property values? Is this an ontological perversion that should count against relativized colors? The cases of weight and speed show that this not an issue. An object has an infinite number of speed values associated with it depending on the frame of reference that the speed is relative to. An infinite number of weight values can be associated with each object, depending on the gravitational field we put the object in. If this ontological explosion is not a problem for weight and speed, why should it count against transformance?

There is another way to diffuse the issue. The weight of every object can be derived from the masses and locations of other objects. Therefore, weight can be viewed as reducible to mass and location. There are strict bridge laws from which we can derive the weight of every object in the universe given the masses and locations of all the objects. We therefore do not have to include weight in our fundamental ontology. The same goes for speed. We can derive all the speeds of all the objects relative to any frame of reference from their locations at different points in time. In the same way, the transformance of an object is not fundamental and can be derived from the microphysical properties of the pertinent system. Each object has an infinite number of transformance values associated with it because it is part of an infinite number of systems, but these values are not ontologically fundamental and can be derived from a finite number of micro-physical properties.

Color Language and the Colors of Common Sense

There are two problems for color as a relational property that are related to color language. The first is that all the evidence of color language usage suggests that we don't view color as a relational property. We say "the shirt is red" instead of the "the shirt is red for viewer x in viewing conditions y". I very much agree with the factual basis of this objection. The folk ontology of color is not a relational one. In our pre-theoretic mode colors are intrinsic properties of objects that do not depend on the viewer or viewing conditions. Tomatoes are red in the dark and the shaded part of the lawn is the same color as the part in the sun. Though I have not done a scientific study, I have yet to encounter a person that disagrees with the previous sentence (in an unreflective mode of course).

Cohen describes the second related objection:

"A fourth and final objection notes that ordinary talk about color presupposes the possibility of agreement, disagreement, and errors of color attribution that seems, prima facie, difficult to reconcile with relationalism. If S1 utters (P1) because the raspberry looks gray when viewed in a circumstance C1 where there is extremely low illumination, S2 says something pertinent, and indeed disagrees with S1 when she utters (P2) because the raspberry looks red when viewed in a circumstance C2 where there is strong illumination by direct sunlight.

(P1) This ripe raspberry is gray in C1.

(P2) This ripe raspberry is not gray in C2.

If relationalism were true, then it seems that the two color attributions just considered could not conflict, just as the following two sentences cannot conflict:

(R1) It is raining in Vancouver.

(R2) It is not raining in San Diego." (Cohen 2004)

I view the two problems as related because the latter problem also arises precisely because color is considered by the folk as an intrinsic property. If color is intrinsic then P1 and P2 are contradictory, just as the folk take them to be. If color is intrinsic then if at one time a person thought an object was one color and later under different viewing conditions he changes his mind, then in at least one of the cases he must have been mistaken. If the shirt I bought was burgundy in the store but when I look at it at home I come to believe it is red, in my pre-theoretic mode I would conclude that I made a mistake in my initial color ascriptions, not that the color of the shirt has changed because of the viewing conditions and that both of my observations were true.

Cohen believes that "ordinary thought and talk about colors presupposes particular ways of filling in the parameters to which color properties are relativized, but that these presupposed parameters are tacit in our ordinary thought and talk." (Cohen 2004) I do not find this plausible. A simpler explanation is that the folk don't think of color as a relational property that depends on viewers and viewing conditions. The folk believe that each object has one color and that is it.

The arguments from color language bring to the fore the question of how to reconcile the common sense view of colors with color relativism in general and transformance physicalism in particular. I believe that the property of speed and how we talk about it is relevant in this discussion also. When we say that the horse is fast, are we tacitly assuming a frame of reference? I think this is very unlikely. Very few people understand that speed is a relational property. Most people never realize that speed is relational unless they attend science classes. Yet, everybody understands the sentence "the horse is fast" or the sentence "the car is speeding". People also have no problem understanding the sentence "the Mars Rover is slow" even though, because it is on Mars, it is hurtling away from us or toward us at a tremendous speed and may be considered fast. This rules out the possibility that we tacitly assume that the surface of the earth is our frame of reference. The Mars Rover is moving quite quickly relative to it. I think the best explanation is that we assume an intrinsic speed property. The Mars Rover is slow both on Earth and Mars just as its mass is the same on Earth as on Mars.

The other possibility is that we tacitly assume a frame of reference relative to the surface an object is on or near. I find this improbable because of the lesson from the following chestnut: How fast can a VW Beetle go? 200 Km/h if you throw it out an airplane. It is clear from this old pun that there is something not right about how we are discussing the speed of the car. Beetles are slow even when you throw them from planes because we believe that they are intrinsically slow. We would not agree that Beetles are fast even though they can reach 200 km/h when thrown from planes.

Both in the case of speed and in the case of color I propose that we just assume erroneously that these are intrinsic properties that do not depend on anything outside the object. I think this also holds in the case of weight. Few people realize that it is a relational property. The situation is analogous to the case of a person measuring the pressure of a tire once at noon and then a second time late at night. Not knowing that temperature influences the reading of the pressure gauge he may say, "I thought earlier that the pressure was x but it is really y". The pressure gauge isn't mistaken or working improperly, it is the interpretation of the result that is mistaken. Our color sensations are almost always correlated with transformance but we have no knowledge of this and assume that when we perceive correctly, what we perceive identifies an intrinsic property of objects.

I therefore think a more nuanced understanding of the declaration that “color relativism makes color illusions very rare” is required. Our color sensations are almost always correlated with transformance and in that sense are near-infallible. However, since humans have a wrong theory of what kind of property color is, we interpret our sensations in an erroneous way and this is reflected in our language of color. Our color mistakes or so called illusions are not at the level of the raw color sensations which are almost always veridical. Rather, the mistakes are caused at a later stage in which we use an incorrect theory of color to infer things about objects based on the color sensations that they are associated with.

Are we making a big mistake? I think not. When we deal with physical properties and their representations, it is not absolute accuracy we are after. What we need is a good approximation. When we compute the circumference of a circle, we often approximate π as 3.14. This is sufficient for most cases even though we never will compute the true circumference using this approximation. For many problems in physics we would use Newton’s simpler equations instead of Einstein’s more complex ones even though the former are not accurate and are just approximations (that are appropriate in most everyday situations). The extra accuracy is just not worth the effort in obtaining it. The person in the tire pressure example above is not living a “grand illusion” about tire pressure. His theory of tire pressures is not scientifically accurate but the approximate theory that he does have is quite sufficient for owning and operating a vehicle. (This issue is expanded in chapter 5).

As was discussed in chapter 2, by viewing color as an intrinsic property, humans are making a simplification that in a range of conditions ("normal conditions") is quite accurate and it is certainly accurate enough for our survival as a species. There just doesn't seem to be any good reason why a more complex concept of color was necessary or is necessary. Human interaction is with objects, not light or visual systems, so as long as color is viewed as a property of objects, the more complete picture of viewing it as a relational property of objects and understanding its dependency on viewing conditions and visual systems is just not necessary or advantageous for any everyday interaction.

Comparing Apples to Oranges

B&H raise a very interesting objection to relativized colors: "Imagine that you have just eaten a tasty crimson fruit, and that you are now looking at another fruit of the same kind. (To avoid irrelevant distractions about color language, imagine you are an Old World monkey.) You recognize the fruit as having the same distinctive shade of red as the first, and that's why you reach for it. Rather surprisingly, this simple explanation of your behavior is not available to the color relativist. Call the first "type of circumstance of visual observation" (Jakab & McLaughlin's phrase) C_{F1}, and call the second C_{F2}. Unless the relativization to types of circumstances is to be pointless, the relativist must concede that the details of the example could be filled out so that C_{F1} and C_{F2} are different. We may assume, then, that $C_{F1} \neq C_{F2}$. According to the relativist, the color the first fruit appeared to have was what we can call "crimson for you in C_{F1}," and the color the second fruit appeared to have was "crimson for you in C_{F2}." Never mind how we should understand these unfamiliar expressions – the important point is that, because $C_{F1} \neq C_{F2}$, the expressions are supposed to pick out different properties (just as being soluble in water is a different property from being soluble in alcohol). According to the relativist, the first fruit seemed to you to have a different color than the second, and hence the relativist cannot endorse the simple and obvious explanation of your fruit-eating behavior." (Byrne and Hilbert 2003)

Transformance physicalism can address B&H's worry. Transformance is a relational property of objects but an intrinsic property of systems that include the objects. The old world monkey can make the comparison between two viewing conditions because it is comparing the transformance of the two systems. Of course the monkey does not know that this is what he is actually doing. At most the monkey like humans, erroneously believes that he is comparing an intrinsic property of objects. Every system has the transformance property and therefore any two systems may be compared. In normal conditions, comparing the transformance of systems is a good approximation for comparing the surface reflectance of objects in those systems and therefore the effort is useful for the monkey[13]. Under this interpretation when we say that our speed relative to the platform is the same as our speed relative to the train we mean that the same property, call it relative velocity between the two objects in the system, is the same for the system [me, platform] and the system [me, train].

The reason two different relational properties can be compared is that one can derive the relational properties from intrinsic properties of systems. The weight of object O_1 is a relational property of O_1 but if O_1 is on earth it is also an intrinsic property of the

[13] See chapter 1, section "The Relation between Transformance and Surface Reflectance"

system [O_1, earth]. The weight of object O_2 is a relational property of O_2 but if O_2 is on the moon it is also an intrinsic property of the system [O_2, moon]. Let's call this intrinsic property of systems W^*. The description of this property is the force exerted on the first object in the system by the gravitational field of the second object. The value of this property for the system [O_1, earth] can be compared to the value of this property for the system [O_2, moon] because these are values of the same kind of property.

Since all the different weight properties (weight relative to earth, weight relative to moon, weight relative to Jupiter etc.) and their values can be derived from the W^* property and its values, they are fundamentally the same property and can be compared. More explicitly, all the values for the property weight relative to earth come from the W^* property of all systems in the set:

{all systems of the form [x, earth] where x can be any object except the earth}

All the values for the property weight relative to the moon come from the W^* property of all systems in the set:

{all systems of the form [x, moon] where x can be any object except the moon}

Since all the values of both the property weight relative to earth and the property weight relative to the moon, are also values of the property W^*, the values of the property weight relative to earth can be compared to the values of the property weight relative to the moon.

Going back to the case of the old world monkey, the two seemingly different relational property values, "crimson for you in C_{F1}" and "crimson for you in C_{F2}" are also values of transformance, and therefore can be compared.

Chapter 4

Color and Color Sensation

In the previous chapter I addressed possible objections to my theory assuming that some form of color physicalism was true. In this chapter I will defend my theory against the more general criticism that applies to any physicalist theory of color. The main argument against all forms of color physicalism is that there cannot possibly be any physical property that does justice to the structure of our color sensations. The idea is that somehow the relations between color sensations should constrain the relations between colors and that there could not be a physical property that is compatible with the constraints. One constraint for example is that since the sensation of orange is more similar to the sensation of red than to the sensation of blue, orange must be more similar to red than to blue. However, there are no relevant physical properties that display such relations. All physical properties that are viable candidates for being red, blue or orange, show no such relations. Interestingly, there is no clear argument in the literature as to why color sensations should constrain colors. It seems to be a strong intuition that some philosophers have that apparently requires no justification.

If the length of a rectangle is 3 inches is it prime? Anybody tempted to answer yes should consider that in centimeters the length is 7.62 and in feet 0.25. The property of being prime is a property of the specific representation and not of the length. The fact that 3 is prime does not reflect back on or constrain in any way whatsoever the length of the rectangle. Zenon Pylyshyn writes: “The temptation to make the mistake of attributing to a mental representation the properties of what it represents is difficult to avoid.” (Pylyshyn

2003) The reverse is also true: “The temptation to make the mistake of attributing to a physical property the properties of its mental representation is also difficult to avoid.” But in order to understand color, it must be avoided.

Color sensations are representations of color, they are not color. Some representations do constrain the objects they represent while most just differentiate between objects in a certain domain but do not pose constraints on the relations between the objects or their properties. For example, in the US, the Office of Homeland Security uses colors to represent security threat levels. The relations between blue and green obviously do not constrain in any way the relations between the security level of “guarded” and that of “low” even though blue represents “guarded” and green represents “low”. If the Office of Homeland Security would have decided that blue represents “low” and green “guarded”, this would have been fine also. There is no necessary connection between the representing color and the represented security levels. The fact that the representation is of a specific form does not constrain the underlying property in any way.

Security threat levels color representations

In this chapter I will show that for both philosophical and scientific reasons it is unlikely that color sensations do constrain colors and that the properties that are found in color sensations are just properties of the representations themselves but not of color. If that is indeed the case, the general argument against physicalist theories does not stand. To set the stage, I will begin by discussing the naïve argument against physicalism.

The Naïve Argument against Color Physicalism

Both the microphysical or abstract physical properties of surfaces are invisible to us. Nevertheless, we do see colors. If this is the case, how can color be equated with microphysical or abstract physical properties? This naïve argument is usually accompanied by the acceptance of the doctrine that Mark Johnston (1992) has labeled "Revelation": All there is to know about color is revealed to us by our visual experiences. The naïve objection is the starting point for attacking color physicalism. It is however easily diffused once the clear distinction between color and color sensation is made: "In considering Revelation, it is important to note that colors are one thing, the phenomenal characters of color experiences, another. Red, for instance, is not what it's like to see red. Redness is a property of surfaces and volumes and, thereby, of the objects and materials of which they are surfaces and volumes. What it's like to see red is an aspect of visual experiences of red; it is a property of such experiences. The doctrine of Revelation for colors should be distinguished form the doctrine of Revelation for what it's like to see colors – the doctrine that the nature of the phenomenal character of a color experience is revealed to us when we have the experience." (McLaughlin 2003)

In a physicalist context, to "see colors" means to have color sensations that represent colors by being correlated with them through a complex causal process.

Imagine a standard mercury thermometer that is put into water. The water is very slowly heated from room temperature to boiling point. The height of the mercury in the thermometer represents the accurate temperature of the water at any given moment in time. The temperature of the water though is fundamentally the mean kinetic energy of the water molecules. How, then, can one know or see what the temperature of the water is just by looking at the height attained by liquid mercury contained in a glass container? After all, temperature is an invisible microphysical property…

Color sensations are not color just as the height of the mercury is not temperature. If one incorrectly defines the temperature of the water as the height the mercury attains when the thermometer is inserted into the water then of course temperature is the height of the mercury. But since temperature is the mean kinetic energy of the molecules, the height of the mercury is not the temperature, but just one way of representing temperature (other representations include volumes of gasses, height of any number of liquids and numbers on a LCD screen). In the same way, the sensation of color is not color, but a representation of color. Both representations are a result of a complex causal process that is mostly “invisible” to us. We are aware only of its end result. Science is required to elucidate the whole causal process.

When temperature first began to be used around the year 1600 it was just defined operationally, for example as the height of the liquid in the thermometer. All that was known was that temperature changes cause changes in the height of the liquid in the thermometer. But the actual changes that occurred in the object being measured were not known. All that scientists could do was measure the effects of these changes. It took another 250 years of experimentation and theory before it was understood that temperature fundamentally is the mean kinetic energy of molecules. What is happening when we measure temperature is that different values of the mean kinetic energy of the molecules cause the liquid in the thermometer to reach different heights. In our brain, different values of color cause different sensations of color. As we shall see, color sensation is just one way of representing color in our brain.

Color sensations are cognitively impenetrable. Looking at a ball we know is "red" and is lit by a blue light, we will have blue color sensations even though we believe we should have red sensations. Our knowledge that the ball is red cannot influence the "erroneous" sensations we have. When we look through a reduction tube at a "brown" dot that we know is brown, we cannot help but have orange color sensations. Our color sensations are not influenced by what we know or believe. They are a result of a complex causal chain over which we have no control[14]. My theory is that at the beginning of this complex chain are transformances and at the end of the process are color sensations which represent non-overlapping sets of transformances. In the case of the thermometer, at the beginning of the causal chain there is temperature and at the end of the chain there is the height of the mercury that represents temperature.

The more serious objections to color physicalism are not naïve. They are not based on the conflation of color and color sensation. Rather, they acknowledge that color sensation is not color but only a representation of color, yet argue that the nature of the representations constrains the nature of color. We will look now at these arguments.

---

[14] Of course, we can look away but the point is that once our look is fixed, we have no control over what color sensations we will have.

From Properties of Color Sensations to Properties of Color

There are things we presumably know about colors either from common sense or from color science. For example, most people would agree to the following statements:

1) The sensation of orange is more similar to the sensation of red than to the sensation of green.

2) The sensation of purple is more similar to the sensation of blue than to the sensation of green

3) The sensation of pink is more similar to the sensation of red than to the sensation of blue

If color sensations do indeed constrain color then the following holds: Since the sensation of orange is more similar to the sensation of red than to the sensation of green, orange is more similar to red than to green.

According to my theory orange, red, blue, green, purple and pink are non-overlapping sets of transformances. It is extremely unlikely that these sets exhibit the above relations even though the sensations associated with them do. It is not even clear how to define similarity between such sets and why there would not be many ways of defining similarity that would give different results. If indeed the similarities between the representations of color constrain color, my definition is not adequate since it does not respect these constraints.

In addition to a similarity structure, color sensations also have an opponent structure and a unitary-binary structure (Hurvich 1997). Red/green and blue/yellow sensations cannot be caused at the same time by one object (system) and are called opponent colors (they should really be called opponent color sensations). There are red, yellow, green and blue hues that are unique in the sense that a normal observer for example can mix monochromatic light in such a way so as to obtain a yellow sensation that does not have any component of green or red sensations by that observer. In contrast, binary hues like orange and purple are always perceived as a mixture and there is no mixture of light sources that can induce a sensation that is deemed unique.

Given these facts, Hardin (1993) formulates the problem in the following way: “The unitary-binary structure of the colors as we experience them corresponds to no known physical structure lying outside of nervous systems that is causally involved in the perception of color. This makes it very difficult to subscribe to a color realism that is supposed to be about red, yellow, green, blue, black, and white – that is, the colors with which we are perceptually acquainted.”

Again, there is no denying that sets of transformances do not display the same relations or structure as the color sensations that represent them. But the question remains: Since the experience of color is different from color, why do relations between color experiences constrain colors? Hardin just assumes that they do since probably it is intuitively clear to him that they must. To understand whether this assumption or intuition is warranted we need to take a deeper look at representations and later on evaluate whether the philosophical arguments and scientific data support Hardin’s intuition.

Representations

A representation in its most general form is a relation between two domains, a represented domain, and a representing domain (Gallistel 2008, 2009). It is important to note at this stage that we are not discussing just mental representations but representations in general.[15]

Fred Dretske gives the following definition of a representation (Dretske 95): "The fundamental idea is that a system, S, represents a property, F, if and only if S has the function of indicating (providing information about) the F of a certain domain of objects. The way S performs its function (when it performs it) is by occupying different states s_1, s_2,….s_n corresponding to the different determinate values f_1,f_2,….f_n, of F."

Thus for Dretske, the speedometer of the car represents the speed of the car because that is its function. That is what it was designed to do. On the other hand, the angle in which the smoke of the fire rises to the sky does not represent the speed of the wind, because it was not designed to do so. As for mental representations, Dretske accepts the notion that evolution can provide design without a designer.

Dretske's attempt at defining representations based on natural (scientific) concepts shows how difficult a project this is. First of all, Dretske has to rely on two concepts, "information" and "design" (or function) to define representations. But both "information" and "design" have no good definitions themselves so it is not clear how Dretske's definition helps. In addition, Dretske's definition is susceptible to the following problem. Imagine that by a chance of nature, a column of mercury gets caught in a glass container. A human finds it and uses it to measure temperature. It is difficult to explain how before the human found the artifact the height of the mercury did not represent

---

[15] Also, we are not concerned with issues of meaning, semantics, understanding and ways of naturalizing these. In addition, we are not just limiting ourselves to reliable causal covariance since according to our initial definition, mathematical domains can represent physical domains and vice versa. The definition is too lenient but since we will be mostly discussing concrete uncontroversial cases of representation, this will not matter.

temperature and once the human found it, the representation properties suddenly emerged in the glass container. Or take Dretske's example with the column of smoke. Does the fact that someone measures the angle of the smoke and infers from it the wind speed, suddenly confer representational properties to the smoke?

I will not make an attempt at providing a general definition of what a representation is. Gallistel remarks about representations: "As with most complex concepts, it can be truly understood only through the consideration of several examples." In what follows I will limit myself to concrete examples which I hope there will be no argument as to whether they are representations or not.

Representations come in many shapes and forms. Gallistel provides an example of two rich featured representations, the representation of geometry by algebra and the representation of proteins and organism structures by DNA. He also provides an example of a very simple representation which involves raising the flag on the mail box when there is mail for the mailman to pick up. In this case the raised flag represents the fact that there is mail to pick up.

In the simple mail example the represented domain consists of two states:

{there is mail to pick up, there is no mail to pick up}

The representing domain consists also of two states:

{the flag is up, the flag is down}

Based on his knowledge of the representation (the relationship between the two domains), the mailman can infer from knowledge of the prevailing state in the representing domain the prevailing state in the represented domain. In other words, because the mailman knows the flag is up he can infer that there is mail to pick up.

It is quite common as well as convenient that the representing domain consists of mathematical entities such as numbers, vectors, matrices or functions but this isn't necessary as the example of DNA and the mailbox flag show. It is important to note though, that any representation involving two physical domains is equivalent to a representation between a mathematical domain and the first physical domain followed by a representation between the same mathematical domain and the second physical domain.

For example, take the case of the mail flag. Let's define the following representation from the physical domain {the flag is up, the flag is down} to the mathematical domain {1, 0}. In this case the flag being up represents the number 1 and the flag being down represents the number 0. In the same way lets define another representation from the mathematical domain {1, 0} to the physical domain {there is mail to pick up, there is no mail to pick up}. In this case the number 1 represents that there is mail to pick up and the number 0 that there is no mail to pick up.

We now have a chain of representations: The flag is up represents the number 1 which in turn represents that there is mail to pick up. The flag is down represents the number 0 which in turn represents that there is no mail to pick up. The whole chain defines a representation from a physical domain to another physical domain but in the middle of the chain there is a mathematical domain. The consequence of this is that we can simplify the discussion about representations between physical domains by "inserting" mathematical domains in causal processes that are strictly physical. I will do this implicitly and often in this chapter. The mathematical domains should be understood simply as a mechanism to aid explanation and understanding.

Representations Do Not Generally Constrain What They Represent

In the most general case, a representation does not constrain at all the underlying domain that it represents. Going back to the simple mail example, it is clear that it does not matter if it is the flag up or flag down state that represents the fact that there is mail to pick up. Either representation is fine as long as the relation is clear to the subjects who use the representation. It is just not necessary that the fact that there is mail be represented by the flag up. The flag down would do nicely also just as a chalk mark on the mailbox would be fine. It is just a matter of defining the representation as to differentiate between the mail and no mail state and letting the users of the representation "know" exactly what the relations between the representing and represented domains are.

In the same way, in a computing device, the fact that there is mail could be represented by any number as long as the fact that there is no mail is represented by a different number[16]. So {1, 0}, {-100, 100}, {1, 2} are all good representations because they differentiate between the two states. In fact, the representing domain can consist of functions (or any other mathematical entity) apart from numbers. For example the representing domain may be $\{y=x, y=x^2\}$. The function $y=x$ represents the fact that there is mail while the function $y=x^2$ represents the fact that there is no mail.

[16] "Represented by a number" actually means represented by a specific physical state of transistors in modern computers.

Another way to look at this is to ask the following question: If we know that there are a 1 and a 0 in the representing domain (it is either {1, 0} or {0, 1}), is this enough in order to infer which state of the mail each number represents? It is not because 1 could perfectly well represent both the fact that there is mail and the fact that there is no mail. The same goes for 0. The properties of 0 and 1 and the relations between these properties do not constrain in any way the properties of the represented states or the relations between the properties of the represented states.

In the example above the properties of 0 and 1 play no role whatsoever in helping us understand any of the properties of the represented domain. In order to make this even clearer let's examine a slightly more complex example. Let's assume there is a simple controller that monitors a machine engine in a factory for temperature irregularities. The controller sends a signal to a display and also to another computer that integrates data from the entire factory. The represented domain is therefore {temperature of engine too low, temperature of engine normal, temperature of engine too high}.

The representing domain on the display could be {blue, green, red} (the actual "colors" are displayed). It could be the sentences {"Warning! Temperature Low", "Temperature Normal", "Warning! Temperature High"}. The sentences themselves could be displayed or an automated voice system may repeat them. In the computer to which the controller sends information, the representing domain can be any 3 mathematical objects that are different. For example, the representing set may be {1, 2, 3} or it may be {0, 1, 2} or {2, 4, 8}.

Since in this case there are an infinite number of possible representing domains, each with different relations between the properties of the objects in the domains, it is not plausible that all these properties constrain temperature. For each specific property that is claimed to constrain temperature, an argument must be provided in order to explain why that specific property of the representation does constrain temperature.

In specific cases, these arguments can be easily evaluated based on our knowledge of the actual causal chain. For example, let’s assume that the representing domain in the computer is {2, 4, 8}. One could attempt to argue that this cannot be a representation of the temperature since the temperatures do not exhibit the property of each being twice the temperature preceding it. However, since by examining the actual causal chain from the temperature to the representation or by talking to the engineer that designed the system, we can readily verify that {2, 4, 8} is in fact a valid representation of the temperature, this argument should not impress us. And when we reflect on the fact that it is quite simple to put in place a causal chain so that the representing domain consists of any three different numbers, it becomes clear that there is just no necessary constraint that a representation using numbers (or transistor states) can put on temperature.

In the most general case then, representations just differentiate between represented entities and do not pose any additional constraints. If the representation of A is different from the representation of B, all we can infer in the general case is that A is different from B and nothing else. In the general case, we cannot infer any additional relation between A and B based on their representations.

A Puzzle Concerning Constraints

There are two different ways to understand what a constraint is. The first is the following: Do properties of representations and the relations between them *necessarily* constrain the properties of the underlying physical properties and the relations between them? The answer to this question is clearly no as the many examples above show.

However, there is another way to look at this issue. Take for example a thermometer. Given the whole causal chain, the height of the mercury must represent temperature because of the laws of nature. Doesn't the fact that two measurements are closer mean that the underlying physical properties are more similar? If the height of the mercury in one measurement is closer to the height of the mercury in another measurement than to the height of the mercury in a third measurement, does that no mean the temperature of the first object is closer to the temperature of the second than to the temperature of the third object? In fact, in this specific case, the height of the mercury does constrain temperature.

In each specific causal chain some properties of the representation *may* constrain the underlying physical property. But this is not the question of interest. The question is rather the following: Not knowing the exact causal chain, can we assume that *all* properties of the representation constrain the underlying relations between the physical properties? Or, can we know a priori which specific properties if at all, of the representation will in fact constrain the represented physical properties? The answer to these questions is of course no, as the examples above demonstrate.

In the general case, there are no properties of the representation that constrain the represented physical property. Given a specific causal chain, we may be able to point to some properties of the representation that do constrain the underlying physical property. In the case of the thermometer only the height of the mercury constrains the temperature but the width of the mercury or the color of the mercury do not constrain the temperatures that are measured.

In the case of color, the eliminativists point to specific properties of the color sensations and insist that theses properties do in fact constrain color. They do not actually point to any specific property and identify it as such, but rely on our intuitions that there are such properties and we have some "immediate" knowledge of them. But lacking full understanding of the actual causal chain from color to color sensations, how can the eliminativists actually know that they have identified the right properties? Couldn't it just be the case that no property of color sensations constrains color, and that color sensations just differentiate between colors but do not pose any additional constraints on them? I will argue extensively in what follows that most likely this is indeed the case.

An Analogy Based On A Digital Thermometer

Before providing the detailed philosophical and scientific argumentation for my position, I will analyze a digital thermometer and point out where I think the analogy with color vision holds. These and other points will be elaborated upon in the remainder of the chapter.

Digital Thermometer

Let's analyze the digital thermometer. First of all it is a measurement instrument. This means for example that when the temperature of the water in a fish bowl in which the thermometer is inserted is x degrees, the number displayed on the thermometer's screen is x (or very close to it). Regressing for a moment back to the naïve objection, let's ask if the number displayed on the screen is the temperature of the water in the metaphysical sense? We may be tempted to answer in the affirmative but numbers are not temperature and neither are LCD screens. The temperature of the water is a physical property of the water. Temperature is the average speed of the water molecules and when we say that the water is at temperature x we mean that the average speed of the water molecules is a certain specific value.

How then is the number displayed by the thermometer related to temperature? It is related through a complex causal process. The metal probe of the thermometer is a thermistor, a semiconductor whose resistance changes with temperature. At low temperatures a thermistor is basically an insulator and as its temperature goes up, its conductance gets better and better. This happens because more and more mobile electric charges are released inside the thermistor. Inside the body of the thermometer is a simple computer (microprocessor) that monitors the thermistor's electrical resistance and displays the appropriate temperature either in Celsius or Fahrenheit by sending appropriate electric signals to the liquid crystal display (LCD) screen. Whether the temperature reading is displayed in Celsius or Fahrenheit is determined by the position of a simple two position switch on the thermometer.

The complete causal process is as follows: The water molecules hit the probe, thus transferring energy to its molecules and raising the temperature of the probe. The temperature of the thermistor in the probe rises also and the thermistor's resistance goes down (its conductance goes up). A microcontroller monitors the resistance of the thermistor through an analog to digital converter. Using the number at the output of the converter the microcontroller runs a simple program based on a lookup table. The program determines in which units, Celsius or Fahrenheit, the temperature needs to be displayed based on the position of the switch, does a simple calculation and sends the appropriate commands to the small LCD screen.

I apologize for the seemingly irrelevant minutiae. My aim is simply to show that there is nothing mysterious about digital thermometers and how they work. The numbers on the thermometer screen are related to temperature via a causal process that always for the same temperature results with the same numbers being displayed. The numbers are correlated with the temperature because of the laws of nature.

The correlation is not one to one though. The numbers displayed are accurate only to a tenth of a degree (in our specific case) and therefore it is not the case that every number is correlated with a single temperature. A whole range, basically an infinite number of temperature values, is related to each number. Because of its finite precision, the thermometer distorts the temperature as it measures it. It rarely if at all displays the number that represents the exact temperature. Instead, it displays a number that represents a whole range of temperature values and we know that the actual temperature is represented by a number that falls somewhere in this range.

It is not only in the display that the thermometer distorts the temperature measured. At several points the thermometer encodes temperature while distorting it. The first point is the analog to digital converter which the microprocessor monitors. Since it has a finite number of levels, any temperature values that fall between two levels are encoded as the same temperature. In the microprocessor, finite accuracy arithmetic is done with the temperature and this may cause further distortion. Finally, the LCD display is limited to a tenth of degree accuracy and this again will cause an infinite number of temperature values to be encoded as the same value.

There is one obvious representation that is immediately noticeable in the thermometer; the numbers on the LCD represent the temperature of, say, the water the thermometer is in. However, there are many other representations that are not as obvious. The temperature of the probe represents the temperature of the water also. The conductance of the thermistor represents both the temperature of the probe and that of the water. The output of the analog to digital converter represents the temperature of the water by representing the conductance of the thermistor. The state of the registers of the microprocessor represents the temperature of the water as well as many other physical properties in the causal chain. The electrical signal the microprocessor sends to the LCD represents the temperature and finally also the number displayed on the LCD represents the temperature. In general, in one causal chain, one can identify many representations.

In the microprocessor itself there could be several representations of the temperature, for example, one in Celsius and one in Fahrenheit. Also, the microprocessor does not have to be connected to an LCD. It could be connected to a speaker that says the temperature or to a printer that prints the numbers that represent the temperature. The LCD is not an essential part of the thermometer.

Now, consider the same thermometer but with the following change in the microprocessor software that makes it “crazy”. The “crazy” digital thermometer, instead of displaying the temperature in a straight forward way, transforms the temperature according to an internal lookup table. For example 70^o may be transformed to 32^o and 69^o to 95^o and so on with no apparent logic. Every possible temperature is transformed into another value but no two temperatures are transformed into the same value. The lookup table is finite because the resolution of the thermometer is finite (in our specific case one tenth of a degree).

Both the regular and the “crazy” thermometer measure temperature accurately. In both cases, the numbers displayed on the LCD represent temperature. The regular thermometer’s representation is better though because it allows us to infer more things. For example, if we use both thermometers to measure the temperature of two objects A and B, we can infer from the measurement with both thermometers whether the temperature of A and B is the same or different. However, only with the normal thermometer can we infer, in the case where the temperature of the two objects is different, whether the temperature of A is higher or lower than that of B.

In the case of the normal thermometer, the relation between representations constrains the relations of the actual physical properties that it represents: If the number displayed for A is bigger than the number displayed for B, then the temperature of A is higher than the temperature of B. In the case of the "crazy" thermometer, the relation between the representations does not constrain the actual physical property in the same way. If the number displayed for A is bigger than the number displayed for B, we cannot infer anything about the actual relation between the temperature of A and the temperature of B except that they are different.

An intermediate case between the normal thermometer and the "crazy" thermometer is the following. Imagine that the engineer designing the thermometer had a very limited LCD which could only display one digit and that he designed the thermometer in such a way that when the temperature was 0^o, 10^o, 20^o, and so on the number displayed would be 0. When the temperature was 1^o, 11^o, 21^o, and so on the number displayed would be 1. In general the number displayed would be the temperature value modulo 10 instead of just the temperature value. I'll refer to this thermometer as the "modulo" thermometer.

What is interesting about the modulo thermometer is that in a limited number of cases, its representation is just as good as that of a normal thermometer. In the case of a macroscopic object whose temperature changes slowly, if the representation of the temperature in the modulo thermometer changes from 4 to 5, we can infer that the temperature of the object being measured is rising. If we knew that previously the temperature was 24, we can infer that the temperature is currently 25. But if we leave the object and don’t track the temperature changes, we lose our ability to make inferences about the relations between temperatures measured at different times.

When measuring the temperature of two different objects with the modulo thermometer, we cannot infer from the numbers representing them which temperature is higher than the other. The relation between the representations does not constrain the relation between the actual physical properties when the modulo thermometer is used, just as in the case of the “crazy” thermometer. Although in some specific cases the representation of the modulo thermometer does constrain the relations between the actual temperature, in most cases it doesn’t. A measurement of 5 is not necessarily higher than a measurement of 4 because the underlying temperatures may be 35° and 44° respectively.

I believe that the modulo thermometer provides a framework to understand the relations between color and color sensation:

1) Color sensations come about at the end of a causal process. Since they are highly correlated with color, just as the numbers on the LCD are correlated with temperature, we (as physicalists) can assume as much.

2) Color sensations represent a set of transformances and not one transformance value just as the finite precision temperature reading represents a set of temperatures.

3) Color sensations do not constrain the actual physical properties they represent just as the display in the modulo thermometer does not generally constrain the relations between the underlying temperatures. In the general case, they just differentiate between colors but do not pose any additional constraints.

4) In the brain, color may have several representations, just as the temperature is represented in several forms in the thermometer in general and in the microprocessor in particular.

5) The same color sensations may represent different sets of transformances for different people because their brain is wired differently, just as one thermometer may represent the temperature in Celsius and the other in Fahrenheit. In the two thermometers, the same number represents a different set of temperature values just as in two different individuals the same color sensation may represent a different set of transformances.

I proceed now to the details of the argument.

A Computing Device Can Impose Any Structure On Representations

To begin with, it is important to point out that a computing device such as the brain can create good representations that have *any* internal structure. This internal structure is specific to the representation and does not constrain the physical property that is represented in any way. Therefore, the similarity, binary and opponent structure of color could easily be a side effect of the brain computing the representation and not a constraint on color.

Let's start with a simple example that shows how the microprocessor in the digital thermometer could create representations of temperature with an internal structure very similar to color sensations. For simplicity's sake let's assume that there are only seven levels of temperature represented initially by $\{1, 2, 3, 4, 5, 6, 7\}$. By writing a simple program that takes the number in the previous representation and multiplying it by itself and adding one, the following representation is generated: {2, 5, 10, 17, 26, 37, 50}. These new numbers are also a representation of the temperature. There is a clear causal path from the temperature level to one of the numbers (a computation is also a causal process) and if we adopt Dretske's definition, the new presentation was designed to indicate different temperature levels and is therefore a representation. It is simple to see that any computation that maintains the differentiation between the temperature levels creates an acceptable representation.

Now, consider the following domain: {(6, 3) (3, 1) (2, 0) (0, 0) (2, 2) (0, 3) (3, 6)}. Each of the seven pairs in the domain represents one of the seven temperature levels. It is trivial to write a program that creates this representation from the initial representation (using a look up table for example or a series of "`if x then y else Z`" commands). By running this program on a computing device, a causal chain is created that connects the temperature to the new representation that is based on pairs of numbers.

What is interesting about this new representation is that it has an "opponent" structure and a "unitary-binary" structure. The numbers 6 and 1 are "opponent" numbers and never show up together in one pair. Neither do the numbers 2 and 3. There are 5 pairs that are "binary", they have two different numbers and there are two pairs that are "unitary" and have the same number twice in the pair.

Furthermore, the similarity structure of this representation is quite interesting. Is (6, 3) more similar to (3, 6) or to (3, 1)? Most people would say that indeed (6, 3) is more similar to (3, 6) than to (3, 1) because they are comparing properties of the representation and not properties of the represented domain. Why people would focus on certain properties and decide that they are the salient ones to determine similarity is not clear but what is clear is that this is completely unrelated to what the different pairs in fact represent. In this case, it is obvious that the human imposed similarity criterion does not constrain the underlying physical property.

I would like to present an example that demonstrates that sometimes a paired representation makes sense even though a representation by a single number is also possible. Suppose that a person has the need to measure the relative speed between trains. He cannot get on the trains though and needs to perform the measurements while standing near the tracks. He has in his possession a speed gun. The obvious solution is to measure the speed of one train, then the other, and then subtract one speed from the other. This would give the relative speed between two trains. It is possible therefore to represent the relative speed between two trains as a pair of numbers, the first is the speed of the first train relative to the tracks and the other is the speed of the second train relative to the tracks. Together though, the two numbers represent the relative speed between the trains. For example, if the speed of train A is 100 Km/h and the speed of train B is 70 Km/h then the pair (100, 70) represents the fact that the relative speed between the trains is 30 Km/h.

The paired representation is more complex structurally than the one numbered presentation but it is much simpler in one crucial aspect. The paired representation can be created and used without the ability to subtract two numbers. In some cases it makes sense to retain a representation that may initially seem more complex because it simplifies the whole system by avoiding additional complex computations (such as subtracting one number from another). The tradeoff makes sense, because even though the specific representation is more complex, the system, of which the representation is just one part, is simpler.

Since the brain is a computational device, the examples above show that it is quite possible that the process of creating representations for colors in the brain could be responsible for creating the unitary-binary, opponent and similarity structure of color sensations (the representations of colors), even though this structure is not found in color itself. It could very well be a by product, a side effect, of the computation and should not necessarily be viewed as a constraint on color.

Color Is Represented In Several Different Ways In The Brain

There is strong scientific evidence that color is represented not just as color sensations in the brain. Before reviewing this evidence, I would like to discuss music and its related sensations and introduce a simple (thought) experiment.

It is quite obvious that there are several different representations of music in the brain. First and most prominent are the sensations we undergo when we listen to music. We also know that air vibrations are first converted to electrical and neurological signals that are highly correlated with sensations of music that we have. Arguably, these are representations of music also. In addition, humans have also the ability to generate music. Having heard a song at one time, we can sing it later. The song then must have been encoded and put into memory to be retrieved later. The representation of the music in memory is not the same representation as when we listen to music. We are not conscious of the music encoded in our memory and it does not have any of the "what it is like" features of the music sensations we undergo while listening to music. In our memory, music must be encoded in principal like it is encoded on CD's, tapes or computers.

Nobody would argue that these encodings have quale like features. They are good representations of music but quite different from the representation we are conscious of when we listen to music.

The same I believe holds for colors. Though we don't have the ability to generate colors directly like we have with music, we are able to do this indirectly. We can remember a scene including its color details and at a later time reproduce the color details quite accurately using actual paints or computer software. It is not plausible that in our memory colors are represented as color sensations since they do not have the "what it is like" features of color sensations. The color representations in memory persist even when we are not conscious of them.

Suppose that you are looking at a blank white wall. Then, someone inserts an object which is blue, green or red into your view and asks you which color it is. You of course would answer correctly. But since your color sensations where only white ones, what did you compare the new sensations to in order to identify it as say a green sensation? When you identified the object as green, your brain was not comparing a color sensation to another color sensation. It was comparing the green color sensation to some other representation of color that had no sensation like features. Not only does it seem that color can be represented in different ways in the brain, there must exist a way to convert between the different representations and compare them.

The considerations above provide strong prima facie evidence that color is represented in the brain in different ways and that the brain as a computing device can easily generate one representation from another. In addition to the conscious representation (color sensations) that we are aware of, there are also unconscious representations that we are unaware of. When combined with the scientific evidence which I will now present, the only reasonable conclusion is that color sensation is just one way of representing color in the brain.

In the early seventies, Larry Weiskrantz demonstrated the existence of unconscious vision in humans (Weiskrantz 86). He called this phenomenon blindsight. Weiskrantz showed that when coaxed, a patient with extensive damage to the visual cortex could guess shapes and forms at an above chance percentage. Yet, the patient reported not having any visual sensations and was just as surprised as Weiskrantz at his ability to “see”. Later research with this patient showed that he could also “see” color even though he reported no color sensations (Stoerig and Cowley 92).Since then, blindsight has been confirmed in dozens of additional patients. The evidence from blindsight leads Nicholas Humphrey to remark: “the existence of blindsight is as good a proof as we could ask for that visual perception does not necessarily *have to* involve sensation” (Humphrey 2006) In the specific case of color, blindsight provides evidence that “seeing” color does not necessarily involve having color sensations and therefore color sensation cannot be the only representation of color in the brain.

More recently, Boyer, Harrison and Ro demonstrated unconscious vision in people with normal brains (Boyer, Harrison and Ro 2005, see Lamme 2006 for an overview). They induced blindsight in people with normal vision by temporarily disabling their primary visual cortex using TMS (transcranial magnetic stimulation). Basically, a small area of the brain was zapped using a magnetic pulse in order to temporarily disrupt neural processing. Two experiments were conducted. In one, a horizontal or vertical line was presented to the subjects. In the other, a red or green disk was presented. In both cases blindsight was clearly demonstrated.

Specifically in the case of color: "When subjects reported being unaware of the target, they were nevertheless still able to accurately guess the color at significantly above chance levels (*M*=81% SD=0.13), *t*(4)=5.43, *P*=0.006. This result further demonstrates that, in the absence of normal V1 functioning and conscious visual perception, another structure(s) is still processing color information." (Boyer, Harrison and Ro 2005)

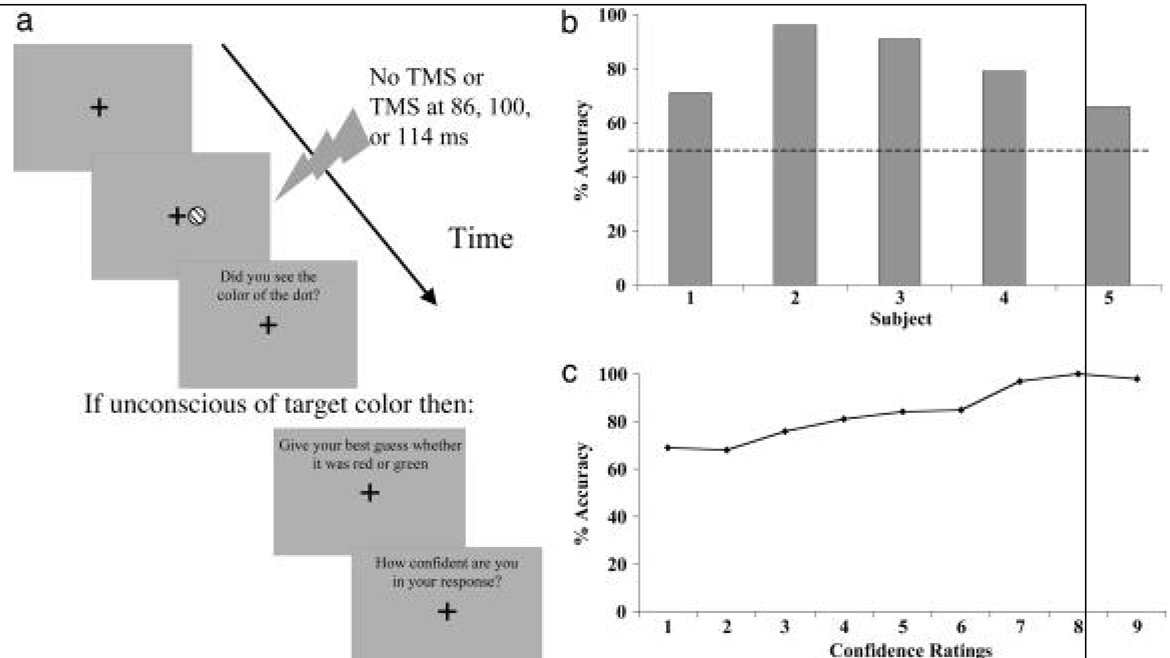


Fig. 2.

Unconscious processing of color. (*a*) A schematic illustration of the sequence of trial events in experiment 2. To examine color processing without primary visual cortex and in the absence of awareness, a red or green disk was presented while suppressing primary visual cortex processing using TMS. When subjects were unaware of the stimulus, they were asked to guess the color and to provide a confidence rating for their guess. (*b*) The mean accuracy in guessing the stimulus color on unaware trials. Despite the absence of awareness, guessing performance was significantly above chance. (*c*) The correlation between accuracy in guessing the stimulus color on unaware trials and the confidence ratings for those guesses.

Another interesting aspect of the experiment is the late timing of the TMS pulse relative to the initial stimulus that was required in order to suppress consciousness. The pulse delay was calibrated for each subject but was in the range of 80 to 140 ms after the stimulus onset. This means that for about 100 ms after encountering a visual stimulus, the brain processes color without there being any color sensation. It is not plausible that throughout this period color is not represented in some way in the brain.

The experiments done in Ro's lab make unconscious color vision a Moorean fact. In principle, any person skeptical of unconscious color vision can go to Ro's lab and experience it for himself. This elevates the evidence for unconscious color vision to the status of "common sense" and should be added to the things that we "know" about color (this will be elaborated in the next chapter). In any case, the scientific evidence is quite clear that color sensation is not necessary for seeing color.

If indeed color is represented in the brain in several different ways, and not only by color sensations, this reduces significantly the plausibility that any specific property of color sensation constrains color. If several representations exist, why is it the case that the properties of one specific representation constrain color? Isn't it more likely or just as likely that no property of color sensation constrains color?

Color Sensations Don't Represent Only Colors

Synesthesia is a phenomenon found in a small percentage of the population and is characterized by certain stimuli giving rise to perceptual experiences not usually associated with them. For example, a grapheme-color[17] synesthete will see a black "a" as red (a grapheme is a letter, number or word). Richard Feynman, the famous physicist, was a grapheme-color synesthete: "When I see equations, I see the letters in colours – I don't know why. As I'm talking, I see vague pictures of Bessel functions from Jahnke and Emde's book, with light-tan j's, slightly violet-bluish n's, and dark brown x's flying around. And I wonder what the hell it must look like to the students." (reported in Pearce 2006 from Feynman 1988).

Synesthesia is involuntary and automatic as well as consistent. Synesthetes cannot control their sensations nor can they change the sensation paired with a given stimuli. The prevailing theory is that synesthesia occurs when connections in the brain are not pruned well during brain development (Pearce 2006, Simner 2006).

While grapheme-color synesthetes are the majority (see Simner et al. 2006 for a differing view), many other kinds of synesthesia exist. The table below is a rough approximation of the prevalence of synesthesia that involves color sensations:

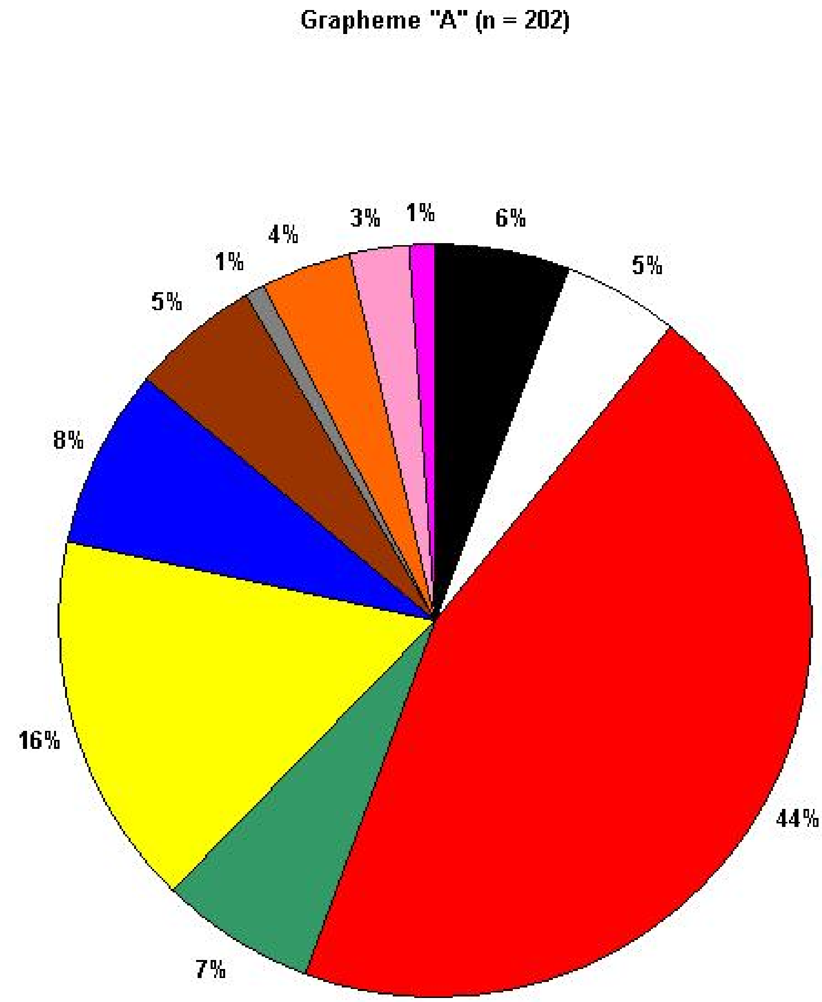
Graphemes -> color sensations	**64.9%**

[17] I am following here the accepted nomenclature. To be accurate, one should write grapheme-color sensation. The graphemes do not cause colors; they cause color sensations.

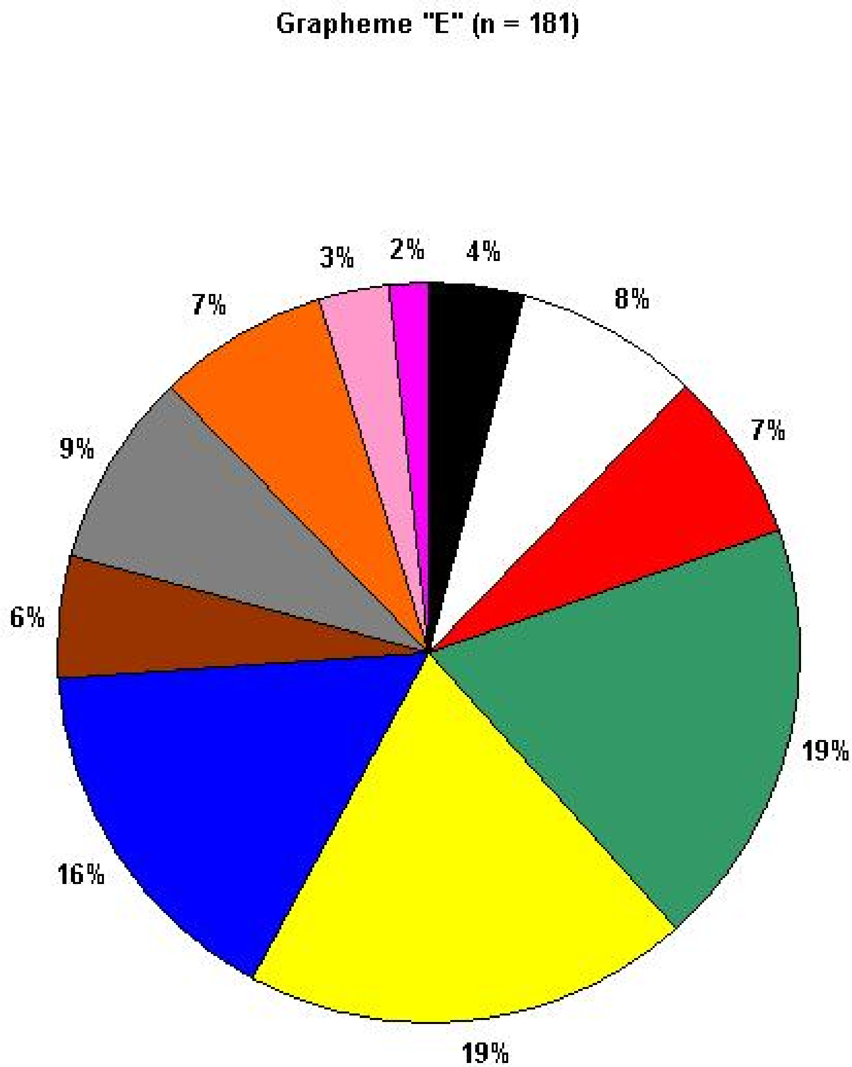
Time units -> color sensations	**23.1%**
Musical sounds -> color sensations	**19.5%**
General sounds -> color sensations	**14.9%**
Phonemes -> color sensations	**9.2%**
Musical notes -> color sensations	**9.0%**
Smells -> color sensations	**6.8%**
Tastes -> color sensations	**6.3%**
Pain -> color sensations	**5.5%**
Personalities -> color sensations ("auras")	**5.4%**
Touch -> color sensations	**4.0%**
Temperatures -> color sensations	**2.5%**
Orgasm -> color sensations	**2.1%**
Emotions -> color sensations	**1.6%**

From the Synesthesia List (http://home.comcast.net/%7Esean.day/)

What is equally intriguing is the variations in the color sensations associated with different graphemes among grapheme-color synesthetes. Sean Day's research on native speakers of Germanic and Romance language speaking synesthetes shows the following distribution for the color sensation associated with the letter "a":


Grapheme "A" (n = 202)
1%
6%
5%
44%
7%
16%
8%
5%
1%
4%
3%

The distribution for "e" is the following:



(Day's research can be found at http://home.comcast.net/%7Esean.day/Trends2004.htm) Notice how different the "a" and "e" distributions are.

Color sensations not only can represent letters and not color, the same color sensation can represent different letters. Therefore, color sensation does not necessarily represent color or anything else for that matter. With the appropriate cross wiring in the brain, color sensation can represent a myriad of physical properties as different synesthetes demonstrate.

Color Sensation Is Generated By An Independent Module Or Process In The Brain

The evidence from the previous section taken together with the evidence of color blindsight, gives plausibility to the assumption that color sensation is generated in the brain by a dedicated process or module that can receive inputs not just from the color vision parts of the brain. For example, in music notes-color synesthetes the part of the brain responsible for analyzing music connects to the color sensation generating module and not only to the sound sensation generating parts of the brain.

My view is that the color sensation module is very much like the LCD in the digital thermometer. The LCD is not essential for measuring temperature. The microprocessor in the thermometer can be connected to a printer, a speaker, a television set and many other devices that can represent temperature in a form suitable for humans and that can be controlled by the thermometer’s microprocessor. Also, the same kind of LCD can be part of many different measuring devices. In addition to being in a digital thermometer, it may be in a pressure gauge, an ohm meter, an ampere meter, a luminosity meter, a Geiger counter etc. In the same way, the color sensation process or module in our brain can represent any physical property (not just color) depending on the causal chain that it is part of. Also, just as the LCD is not necessary for measuring temperature, the color sensation module is not necessary for color vision.

The case of a color blind synesthese described by Richard Cytowic provides clear support for this model (Cytowic 2003 in the after word). This particular synesthese has an S-cone deficiency that makes it difficult for him to distinguish between blues and purples. He reports "seeing" numbers in colors that he cannot see in the real world and names these colors "Martian" colors. My model easily explains such occurrences. The synesthese's color vision is impaired but the process or module responsible for generating color sensations is not. Because color vision is impaired, the color sensation module is not used to its full potential. The analogy here is with an LCD that can display the number nine but is in a specific causal chain such that the appropriate electric signals that would result in the LCD displaying the number nine are never sent to the LCD. In the case of the synesthese, the signal to generate the (say) purple color sensation is never sent to his module from the impaired color vision system, but the appropriate signal to generate the purple sensation is instead sent from the part of the brain responsible for detecting numbers.

Confusing Properties Of Representations With The Properties Of The Represented Entities

Is 一 more similar to 三 or to 零 ? Most people in the West would answer that 一 is indeed more similar to 三 than to 零 based on graphical similarities. However in the Far East, many people would know the meanings behind these symbols and that they represent 1, 3 and 0 respectively in written Japanese. They may answer differently because 0 is closer to 1 than to 3.

When we are asked to compare say object A to object B, it is not always obvious whether we need to compare the features of the actual objects or the things both objects represent. I think that if we don't know what a symbol represents or its meaning, we automatically put quotation marks around it. So for example, we read the question above as:

Is "一" more similar to "三" or to "零"?

The Japanese read it though as:

Is 一 more similar to 三 or to 零? Or simply as: Is one more similar to three or to zero?

Since the difference between 1 and 0 is smaller than the difference between 1 and 3, some Japanese may conclude that 1 is more similar to 0 than to 3. Temperatures of 3°, 7° and 11° have nothing special in common even though the numbers representing them are prime. 3° is not more similar to 13° than to 12° even though 3 and 13 are prime and odd. Whether we compare properties of the representations or properties of what is represented, makes a major difference and depends on whether we realize that something is being represented and can differentiate between the underlying property and its representation.

Based on the scientific and philosophical debates, we can safely conclude that generally humans do not know what color is. This is an unavoidable conclusion for any physicalist. If color is something that we learn about through science, and not by immediate experience, at the minimum, any person not aware of the science associated with color does not know what color is. For example, if the physicalist story is true, a person cannot understand color without understanding at least surface reflectance (if not transformance). This means that most of humanity does not understand or know what color is.

Therefore, when humans are confronted with the question "is orange more similar to red than to blue", we unavoidably base our answer on what we are familiar with which are color sensations. Had we known, what stands behind the color sensations, what they actually represent; our answer would be quite different. My answer to the question above would now be that it is ill defined until the person posing the question decides on a method of determining similarities between sets of transformances and that my "feelings" or intuitions based on color sensations are not relevant to the answer.

It is easier for "normal" people to see this exact process happening in synesthetes. The following passage documents Franz Liszt, the famous composer, mistaking a property of the representation with the property of the object represented: 'When Kapellmeister in Weimar (1842), he astonished the orchestra saying: ''Oh please, gentlemen, a little bluer, if you please! This tone type requires it!'' Or: ''That is a deep violet, please, depend on it! Not so rose!'' First the orchestra believed Liszt just joked; later they got accustomed to the fact that the great musician seemed to see colours there, where there were only tones.' (Reported in Pearce 2006 based on an anonymous article in a German newspaper from 1895)

The musicians in Weimar were probably quite amused by what they correctly viewed as Liszt's obvious mistake of assigning color properties to tones. But their smugness was unwarranted. They were in fact making a very similar mistake (as all humans do) by assigning color sensation properties to colors. The only difference between the two mistakes is that the latter is committed by all people and the former is relatively rare. Had most of the population been synesthetes like Liszt, no one would have thought that Liszt was committing an error.

Our only connection to the world is through a complex computational process that creates representations of the world. We are only conscious of some of the representations and they capture just a miniscule amount of the information “out there”. On the other hand, the representations are good enough to allow our survival and do not overload us with information that is not essential. We should be humble about our abilities to understand complex physical properties based on their representations alone. We need science in order to complete the picture.

The Argument Revisited

The physicalist story that I am arguing for is as follows. The color sensation module/process (CSM) is an independent one and can receive inputs and send outputs to many other processes/modules in the brain and not only to the color vision related ones. The intuition that color and color sensation are necessarily connected is proven wrong by color blindsight and by synesthesia. Color sensations are a result of a complex causal process in which some input that is correlated with color is fed into the CSM. This complex causal process takes about 100 ms. During this process, several complex representations of color are computed in the brain. Color blindsight is evidence for this.

Since representations do not necessarily constrain what they represent, my theory is that color sensations merely differentiate between sets of transformances but do not put any additional constraints on them. Our intuitions regarding these constraints are wrong because we are mixing properties of the representations and properties of the underlying physical property. This is something we are apt to do when we don't understand what something represents.

Since the benefit of color vision is mainly in differentiating between objects, it seems highly unlikely that it also evolved to be able to estimate similarities between obscure and highly complex physical properties such as transformance and surface reflectance (even if we knew how to identify the salient similarity measure between sets of transformances). Any algorithm to compute a representation of color that was simple enough to differentiate between colors, even though similarity information could not be inferred from this representation, would have been sufficient. It is therefore much more likely that such a relatively simple algorithm evolved rather than a more complex one that created a "better" representation in the sense that similarity could be inferred from it (again, if one could even define what similarity means in the case of sets of transformances). Mechanisms to differentiate between colors evolved because they provided concrete advantages to organisms. On the other hand, the ability to deduce that two colors are more similar from a physical point of view to each other than to a third color does not seem to provide any advantage to the organism so there is no reason to believe that it was selected for.

In addition, color has several different representations in the brain. Why is color sensation more "important" than the other representations and why believe that it has some special properties that restrict color? The other representations in the brain are probably very similar to how different colors are represented in a computer and probably can be described in some numerical format. The other representations occur both before and after color sensations and just make it more likely that the particular properties that eliminativists point to are just properties of the specific representation and do not have anything to do with color (since the other presentations do not have this property).

The brain being a computational device can create any property for the representations it uses. It is therefore quite possible that the structure of color sensations is a side effect of the brain and is not a property that constrains color.

There is also another way to argue my point. Since transformance is 100% correlated with color (and it is not a disjunction), this is a strong indication that it is a viable candidate to be color. Since it is not necessary that representations constrain what they represent, why not accept the hypothesis that indeed color sensations do not constrain color?

The choice is between one of two hypotheses:

1) There is a physical property that is 100% correlated with color sensations but it is not color because it is not constrained by color sensations.

2) There is a physical property that is 100% correlated with color sensations and it is in fact color because like in the most general case with representations, color sensations as representations of color, do not constrain color.

Hypothesis 1 creates a significant challenge for the eliminativists as they need to explain how this correlation occurs. It is not likely that it occurs by coincidence. Based on the arguments put forward in this chapter, I find hypothesis 2 more plausible both from a scientific and a philosophical point of view.

Chapter 5

Knowing Color

What is it that we know about color? What are we justified in believing about color? I will attempt to answer these and similar questions in this chapter. I will also be exploring the boundaries between cognitive science and epistemology and the boundaries between our sensations and our epistemic faculties.

What do we know about color then? The question posed in this way is not specific enough because it can be understood in two completely different ways:

1) What does common sense (together with arm chair philosophizing) tell us about color?

2) What does science tell us about color?

By focusing on just one of the two questions above, a biased view of what we know about color may emerge. Just by picking one meaning of the question over the other, one may inadvertently decide which source of knowledge to privilege and miss a crucial aspect of the issue at hand. To a large extent, the philosophical problems associated with color involve deciding which source of knowledge should be given greater weight and how to reconcile the apparent contradictions between the two sources of knowledge. Put bluntly, what should we trust more in understanding color, our common sense or what science tells us?

Hardin's "Color for Philosophers" was a turning point as it succeeded in achieving one of Hardin's explicit aims: "Finally, I have wanted to encourage and provoke other philosophers to come to grips with the relevant scientific material, and to promulgate within the philosophical community the opinion that, henceforth, discussions about color proceeding in ignorance of visual science are intellectually irresponsible. If this book should help to effect such change in attitude, I shall think of it as a great success even if the other philosophical theses it contains should be consigned to the flames as sophistry and illusion." (Hardin 1988)

It seems then that Hardin's most important objective in writing his book was to set epistemological norms for thinking about color. Reading through the color literature published since, it is obvious that he has succeeded in his goal. While pre-"Color for Philosophers" common sense was the prime arbiter regarding color, few philosophers would now argue against the strong epistemological assertion that "discussions about color proceeding in ignorance of visual science are intellectually irresponsible". Before Hardin's book, common sense alone was accepted as adequate evidence for a philosophical investigation into color. Following Hardin's book, the epistemological norms changed and an investigation based solely on common sense has become "intellectually irresponsible".

The almost consensus in the philosophical community about the meaning of the question "what do we know about color" is that it means "based on science and common sense, what do we know about color". But what should be done in the cases that science and common sense do not agree? My view is that the scientific evidence should be our main guide to a true theory of color (I will explain why shortly) but that any theory of color is not complete unless it can explain two interesting facts.

Two Paradoxes[18]

The first fact is the apparent success of the common sense theory of color. It is not enough to assert that common sense is wrong in the case of color. The additional required step is to explain why it is wrong and how humans could have managed so well with an erroneous theory of color. I think it is unavoidable that once the results of visual science are given weight, any theory of color will be an error theory and that it will not be identical to our common sense view of color. However, it is also an indisputable fact that our common sense theory of color, while being plain wrong, has served humans well. This is a paradox that any theory of color needs to explain. Why is it the case that a wrong theory works so well?

The second fact that requires explaining is the relation between the common sense theory of color and science. Science is based on observation and experimentation which of course involve using color. Newton was able to observe the red apple falling towards earth because the apple had a color that was different from its background. The color of the apple was crucial to the ability to observe it. Our scientific findings are in many cases based on our erroneous theory of color. For example, following is a list of scientific observations that rely heavily on color:

1) The color of blood changes as a function of its oxygen content.

2) Once the hydrogen fuel at the core is exhausted, a star of at least 0.4 times the mass of the Sun expands to become a red giant.

---

[18] I use paradox here in the loose sense of the word.

3) The Andromeda Galaxy is moving towards our own Milky Way Galaxy within the Local Group; thus, when observed from earth, its light is undergoing a blue shift.

4) One can measure pH using indicator paper that turns different colors corresponding to different pH levels.

5) Brilliant Blue FCF is a colorant that may be added to foods to induce a color change. It is denoted by E Number E133. It has the appearance of a reddish-blue powder. It is soluble in water; solution has maximum absorption at about 630 nm.

6) Brain activity can be monitored based on color changes in fMRI and PET scans.

Any error theory of color needs to explain not only how humans could manage with an erroneous theory of color, but also how science itself, which was built using a wrong theory of color, could be a reliable source to provide information about color. This is not only a problem that my theory encounters. It is also a problem that all modern theories of color must address head on. For example, eliminativists such as Hardin use science to argue that common sense colors do not exist. But since in most scientific theories the colors of common sense are assumed, how could science, which starts by assuming the existence of common sense colors, end up saying they do not exist? I think this is a problem eliminativists have not seriously addressed.

Somehow, we use an incorrect theory of color to build scientific theories that we believe are true and then use these theories to discredit the common sense view of color. This does not seem kosher and is highly suspect from an epistemological point of view. It is an epistemological paradox that any modern philosophical theory of color must address.

Why Science Trumps Common Sense

Before discussing why my theory overcomes the two problems outlined above, I will explain why the scientific evidence about color should be given priority over our intuitions and common sense view of colors. In general, the language of science is mathematics but this is especially true of physics. We gain understanding of physical phenomena by being able to represent them mathematically. Every mathematical system though is based on axioms or postulates that are taken as granted. Changing the axioms changes the "truths" of the systems. If our senses and intuitions are to be a good guide to the laws of nature, they must be a good guide to determining the correct axioms of the mathematical systems that represent the laws of nature. However, the history of the last two centuries shows us that this is not the case.

The discovery of non-Euclidean geometries undermined philosophers' of mathematics confidence that that our senses and intuitions can reveal special mathematical axioms: "The alternatives to Euclidean geometry persuaded people that the senses and sensory imagination cannot certify controversial axioms and postulates." (Goldman 86) The discovery that space itself was not Euclidean (cannot be modeled mathematically by Euclidean geometry), drove home the point that when it comes to describing the world, our senses and intuitions should not be the sole arbiter or the preeminent one.

What our senses and intuitions do is give us an initial simple theory of the world. It is a theory that allows us to succeed in the biological sense (survive long enough to pass on our genes to future generations) but our senses and genes do not provide us with an accurate theory of the world. Our common sense theory of the world is an approximation that is good enough for all biological purposes. We did not evolve to represent the world accurately. We evolved to be successful in our particular biological niche, and for that a simple and approximate theory of our world suffices. Complex calculations require more energy and time than simpler ones, so an approximation that leads to the same results as a more complex accurate computation one would have been selected for.

There could have been no benefit whatsoever for our ancestors on the plains of modern day Kenya to believe that space was not Euclidean or even that the surface of the earth was not flat. Since they lived all their life in a limited area, the assumption that the earth was flat posed no problem for them whatsoever and had the advantage of leading to simpler cognitive processes that while being only approximations, were more than sufficient. Our ancestors would not have hunted, scavenged, found refuge or in general reproduced better had they believed the world was not flat. On the contrary, this would have been a hindrance to them since it would have made many calculations in the brain much more complex.

My assumptions are very similar to those of Herbert Simon (1982) and Gigerenzer (Gigerenzer 93, 96, 99) in their discussion of rationality. Their main assumption is that an organism's resources are limited. Our ancestors had limited time, limited knowledge and limited computational capacities and therefore rationality needs to be defined with these constraints in mind. Simon proposes a "satisficing" standard for rationality which he refers to as bounded rationality. Gigerenzer and Goldstein define a good inference mechanism as one which is "reliable, fast and inexpensive (frugal)". I am not interested here in the question of whether reasoning using our simplified model of the world is rational or not. My aim is to only to suggest that in an environment of limited resources, simple theories that are just approximations of reality and not strictly true, would likely have been selected for.

A Limited Definition of Approximation

To understand what I mean by “approximation” better, let’s examine the task of computing the circumference of a circle given the diameter. If we know the diameter, we obtain the circumference by multiplying it by π. The problem though is that π is an irrational number that requires an infinite number of bits to represent it in any form that is amenable to multiplication on a computing device. If we want to accurately obtain the circumference of the circle, we would need either infinite storage or infinite time. We can get around the problem of infinite storage by computing π “on the fly”, that is, using a finite program to compute another digit of π when the multiplication requires it. Since π needs an infinite number of digits to be represented accurately, this would literally take forever.

To avoid such problems, we usually represent π as either 3.14 or 22/7. Why? For most of us the reason is because that is what we were taught in school. Our teachers told us that these approximations were good enough, since if we used them we could still get full scores on our tests. Had we use 3.1 or 3 as approximations our test scores would be lower. Had we used 3.141, our scores would not be higher. Therefore, 3.14 turned out to be an adequate approximation since any greater accuracy would have not served our goals. In fact, it would have hindered them. It takes more time and it is more prone to error to multiply a number by 3.141 relative to multiplying it by 3.14.

Approximations are judged to be good or bad relative to our goals. For a student, an approximation of 3.14 for π is good because using it he attains the highest possible test scores. An engineer may need to use an approximation of 3.1415 in his calculations in order to makes sure that a bridge does not collapse and a NASA scientist may have to use an approximation of 3.14159265358979323846 in computing the trajectory of a mission to Mars in order to make sure it succeeds. In each case the approximation is deemed good or bad relative to the goal of the person making the computation.

An example of a different kind of approximation is Newton's theory of mechanics. We know it to be wrong and certainly less accurate than Einstein's theories. However, in everyday situations, in which the speed of objects is a small fraction of the speed of light, Newton's theory provides us with excellent results. It is good enough to satisfy our prediction and design goals. For example, if we need to predict where a missile will hit, but not err by more than 1 foot, then Newton's theory would be adequate. If we want to keep our prediction error below 1/8 of an inch, we may have to use Einstein's theory.

Approximations of the kind that I am interested in can occur at two different levels. The first level is the input level to the computation. This is the case of representing π as 3.14. An approximation can also occur at the level of the computation. Instead of a complex computation, a simplified one that is less accurate may be used. This is the case of Newton's theory being used instead of Einstein's.

The concepts of approximation, theory, model and representation and how they relate to each other have been notoriously difficult to define (Stanford 2006). I will not attempt to do this here[19]. The few paragraphs above were meant to give an intuitive feel for what I mean by approximation. Instead of giving a general theory, I will take advantage of the fact that we are dealing with a relatively narrow and concrete subject matter and explain how my theory overcomes the two paradoxes outlined above. When I talk about "approximation" below, I will always give a concrete example and my efforts should not be construed as giving a general theory of approximation.

Solving the Two Paradoxes

The first paradox I outlined is the apparent fact that humans get along very well with a wrong theory of color[20]. The solution for this paradox is that the common sense theory, while inaccurate, is a good approximation for the true theory of color in many cases. Just like Newton's theory is a good enough approximation for common situations, the common sense theory of color is also a good approximation for situations we are likely to find ourselves in. In some cases, one theory can be shown to be the limit of another theory. Einstein's theory of special relativity converges with Newton's theory when we approach the limit in which the speed of light is infinity. As I explained in chapter 2, my theory converges to the common sense theory of color in a wide range of viewing conditions. When we assume that (1) lighting conditions are close to typical daylight or within the range that color constancy is effective and (2) that the distortion of the visual system is the same or changes little over a wide range of viewing conditions, it

[19] For example, what is the difference between a representation and an approximate representation? Aren't all representations approximations? What is approximate truth? Aren't theories themselves approximations of reality? This is a quagmire I would not like to tread into.

[20] It is probably an overstatement to call the human conception of color a theory. It is more like a model since it is something used in doing computations and building representations and not something used to describe reality. Also, the human conception of color is silent on several critical features that would have to be an essential part of any theory of color. For example, the human conception of color does not specify what kind of physical property color is. The concept of surface reflectance is not part of our non-reflective thinking about color.

is easy to explain why humans believe that color is an intrinsic color of objects and why this is not such a bad theory. Under the two assumptions above, which actually hold in many cases the only apparent changes in color come from changes in the objects viewed and not from changes in the light or visual system. It is reasonable to conclude then that color is an intrinsic property of objects.

Eliminativists are not able to defuse this paradox because they deny that colors exist. Even if we interpret eliminativism simply as the theory that the colors of commons sense do not exist (and not that color in general does not exist), eliminativism still does not explain why humans are successful even though they use a wrong theory. Since they do not present an alternative to the common sense view of color, and perhaps even deny that there is an alternative, the eliminativists cannot show what the relation is between the alternative and the common sense view. My theory, on the other hand, can show that the common sense view is a good approximation to the scientific view and thus explain how a wrong view can be so useful.

The second paradox is the following. How is it possible that we use the incorrect common sense theory of color to build scientific theories that we believe are true and then can use these theories to discredit the common sense view of color? How does this epistemological magic happen? How can an incorrect theory be a basis for the science that then identifies that the theory is incorrect? If one adopts my theory, a simple solution to this paradox is possible. Scientific experiments are usually done in labs which have very similar lighting conditions. When a second lab repeats the experiments of the first lab, the lighting conditions in the second lab are very similar to those in the first lab. In addition, the visual systems of the scientists are usually very similar and those scientists that are “color blind” are usually aware that they are different and do not trust their own color observations. Therefore, the common sense theory of color is good enough not to hamper science. It is a good enough approximation in most scientific experimental environments.

For example, scientists would examine PH indicator paper in good lighting and would not rely on observations taken in a dark room. In addition, each indicator paper usually has two distinct hues such as yellow and red. These are easy to tell apart even when viewing conditions are not optimal. This is often the case when colors are used for observations. Scientists do not rely on subtle color differences based on observing different shades but usually use color in a coarse manner and rely on hues for differentiation.

So far we have seen that my theory of color is able to resolve two epistemological paradoxes that eliminativists and people that deny that colors are physical properties would have a hard time explaining. I will now move on to discussing what we actually know about color.

What Do We Know About Color?

What then do we know about color? Let’s start with what we presumably know “not as a student of physics or physiology, but simply in your capacity as a subject of visual experience” (B&V 1991). Let’s begin with a list put forward by Boghossian and Velleman (B&V 1991):

1) Red and orange are properties.

2) They are different properties of the same kind.

3) Red and orange are less different from each other than they are different from blue.

4) Red and orange cannot be simultaneously instantiated in the same place.

5) Red and orange are properties that things visually appear to have in certain conditions that are also known to us.

According to Boghossian and Velleman if the above were not true our knowledge could be hostage to "future empirical discoveries" and that one would have "to consider the possibility of obtaining evidence that red and orange are in fact the same property or, conversely, that they aren't similar at all". In general, I am disposed to believe (as Hardin believes), that our metaphysical theories should be hostage to empirical discoveries. If electrons figure prominently in our best scientific theories, my view is that electrons should be included in our ontology even though they are not part of our common sense ontology.

In the specific case of color, understanding the difference between color and color sensation allows us to explain how what we know about similarities between color sensations does not lead to knowledge about similarities between colors. In the specific case of color, the understanding that it is a relational property dramatically alters what we thought we knew about color in our non-reflective mode. In the case of color, confusing physical properties (colors) with their physical representations (color sensations) and misunderstanding the kind of property color is leads to incorrect knowledge statements.

In this light let's analyze Boghossian and Velleman's knowledge claims. The essence of their knowledge claims is composed of two statements:

1) It is not possible that red and orange are in fact the same property

2) It is not possible that red and orange are not similar

Let's defuse the second claim first. As we saw in previous chapters, it is quite possible that red and orange are not similar but their representations, the sensations of red and the sensations of orange, are. As for the first statement, it is not clear what is meant by it. Does it deny the possibility that the sensation of red and the sensation of orange represent the same color? Or is it merely a statement that the sensation of red is different from the sensation of orange? Given the structure of the original sentence, I think it is the first meaning that is pertinent. Between individuals it is uncontroversial that the same color can and is represented by different sensations. There is the extreme difference between "color blind" people and "normal" people, but as was discussed, there is a substantial variation within the "normal" group.

Perhaps Boghossian and Velleman are talking strictly about the same individual and asserting that for him, it is not possible that red and orange represent the same property? This does not help their case much since "color blindness" can easily be induced in an individual by injury. The best case for Boghossian and Velleman is one in which the same person is looking at the same time at two objects, one that causes a red sensation and another that causes an orange sensation. Would it be correct to say in this case that it is not possible that the sensation of red and the sensation of orange represent the same color? No, it would not. We must remember that Boghossian and Velleman are discussing color as an intrinsic property of objects (something akin to surface reflectance which is an intrinsic property) since this is the basis of our common sense "knowledge" of colors. It would be quite simple to set up a scene in which two physically identical balls are each predominantly lighted by different lights (say a red one and an orange one) and therefore the same “color” (according to Boghossian and Velleman) is represented by different color sensations. This example shows that that statement 1 is wrong if it is interpreted as talking about representations. The representations of red and orange (the red and orange sensations) can in fact represent the same common sense property.

My view is that we know very little about color in our non-reflective mode. We know that color is a property but we do not know what kind of property it is. We are not aware that color is transformance, a complex relational property. Furthermore, we are also confused about the relationship between color and color sensations. In what follows I will argue, that though our beliefs about color are mostly mistaken, they are justified. In regards to color we do not attain the epistemological goal of knowledge but we do reach the lesser level of epistemological attainment in that we have justified (but unfortunately false) beliefs. While we are wrong about color, we are at least justified in holding the beliefs we have about color (only in our non-reflective mode of course). I will first provide an intuitive argument in support of this view and then proceed to a more rigorous argument based on a process reliabilist theory of justification for color beliefs.

Our Common Sense View of Color is Justified

How do science and common sense interact in the case of justification? Alvin Goldman provides an example of a child looking for the first time at an oar in the water (Goldman 1986). The oar definitely looks bent to the child. Is the child justified in believing that the oar is bent? If in general the child is justified in believing that his perceptions portray the world as it is, why is he not justified, in this specific case, in believing that the oar is bent?

My intuitions about the case of the oar are not clear. I am not sure if the child is justified or not. Perhaps when it comes to cases when we can employ more than one of our senses, using just one of our senses is not enough to achieve justification (the child could have put his hand in the water to feel the oar but did not do so). However, in a similar case my intuitions are much clearer. Until the 20th century, were people justified in believing that space is Euclidean? The two cases are similar because in the case of the oar we are asking if it is bent or not and in the case of space we are basically asking the same question, is space bent (a two dimensional Euclidean space is like a plane while a non-Euclidean two dimensional space may be like the surface of a ball)?.

Until the 19th century the possibility of non-Euclidean geometries was not even contemplated. Euclid's five postulates were taken to be self-evident truths and not just axioms that can be changed to create new mathematical systems. It is therefore not surprising that it was obvious to people that space must be Euclidean. All the evidence, both common sense and scientific pointed in this direction. While the common sense evidence did not change, the scientific evidence accumulated in the 19th century led to Einstein's theory of relativity and the understanding that space is not Euclidean.

Was Newton, or any contemporary of his for that matter, justified in believing that space was Euclidean? My intuition is that given the evidence that they had that overwhelmingly portrayed space as Euclidean and the immense difficulty in imagining and verifying an alternative to this view, Newton and his contemporaries were justified in believing space was Euclidean. But as the scientific counterevidence mounted, scientists became less and less justified in their belief.

Before Newton's time, were people justified in believing that two bodies cannot exert force on each other from a distance? Based on human experience, most probably the answer is yes. Newton had to posit action at a distance and was unable to justify his seemingly irrational position except by pointing to the elegance of his theory of gravitation and its ability to account for astronomical observations. We are justified in our common sense beliefs until scientific evidence to the contrary emerges. But when we become aware of the additional evidence provided by science, we are not justified anymore in maintaining our common sense beliefs. Our justification is defeated by the counterevidence.

Justification also depends on how easy it is to think of alternatives to our common sense view and how easy it is to verify these alternatives. In the case of the nature of space as well as in the case of color, this is very difficult as can be discerned from the actual historical thought process of the community contemplating these issues. These difficult cases can be better understood using the framework put forward by Goldman in "The Sciences and Epistemology" (2004). According to Goldman, epistemology is not a scientific undertaking but it is constrained by science. If science tells us that our human reasoning facilities are limited or evolved to tackle some problems better than others, this constrains epistemology since it does not make sense to judge humans by norms that are not commensurate with their cognitive abilities. When we set epistemological norms, we should take into account the limited reasoning abilities of humans and not be overly demanding. Given human cognitive limitations, it is not reasonable to expect Newton and his contemporaries to have imagined and tested an alternative to the possibility that space is Euclidean. It is also not reasonable to expect humans to arrive at a correct theory of color just by introspection and arm-chair philosophizing. Hundreds of years of communal thought and research were required in both cases.

The Interface between the Visual System and the Reasoning System

Before presenting my process reliabilist theory of justification regarding beliefs about color, I would like to address the issue of where, when and how epistemology meets color. This is important because it is a required step in identifying the process to be analyzed. As previously discussed, our epistemic faculties cannot influence what color sensations we have. Our color sensations are cognitively impenetrable in general and specifically, our epistemic faculties cannot influence our color sensations. Our color sensations are not influenced by our beliefs and epistemic norms. When we look at a "blue" ball that is illuminated by a red light we cannot help but see the ball as red even though we know and believe that the ball is blue. Since color sensations are highly correlated with a physical property (if my theory is right), we can assume as physicalists that color sensations are a result of some causal process over which we have no control.

How then does color vision interact with our epistemic faculties? How do we come to have beliefs about color? Put differently, the question is how the raw sensation of blue causes the belief in the proposition that the sky is blue. There must be some minimal information that is passed between the two systems (by some causal interaction presumably). The visual system must make some propositions or their representations accessible to the reasoning faculties (these are often called "perceptual beliefs" in the literature). Otherwise, we would just have sensations without the ability to infer anything from them. The sensations would be a "dead end" so to speak from an epistemological standpoint.

Let's assume Tom looks for the first time at an orange. Furthermore, let's assume that he does not have the concept of an orange. He has not heard of oranges and discovers that they exist for the first time by looking at one. Tom would come to believe that there is an unknown object in front of him with an unknown color. This chain of events cannot happen unless the visual system creates some representations of propositions and makes them accessible to the epistemic faculties. The proposition represented is at the minimum the following: There is a round like object (known or unknown) with some (known or unknown) color in such and such location. Though our visual system is quite different from our reasoning system, there must be some module (cognitive process) in the mind that knows how to take presentations in the visual systems and convert them into a belief. Some philosophers describe belief creation as putting propositions in "belief boxes". I think that a more accurate way to view what is happening in the context of the computational theory of mind is to posit the existence of a computational process that converts representations of propositions from one format to another. The process converts propositions from the format of the visual system to the format of the reasoning system.

Format conversions are quite common in computational systems. For example, integers are often converted to floating point numbers and vice versa. A more complex case is the conversion of music from the format on a CD drive to the more space saving MP3 format. Format conversion in many cases is not an information conserving computation and is often irreversible. For example if we convert the number 5.34 from a floating point representation to an integer representation we obtain the number 5. This is a less accurate representation since we have lost all the digits after the decimal point. If in the future we want to convert 5 back to 5.34, we cannot do so since we lack the required information. The conversion of a floating point number into an integer format is almost always irreversible. The conversion of music to the MP3 format also has the same characteristics.

The format conversion module that is the interface between our visual system and our reasoning faculties is also one that results in loss of information. This module is the source of the common sense view of color because it implants the beliefs about color into our reasoning system. When we look at a blue ball we come to believe that in front of us is a ball with the intrinsic property of being blue because that is the belief the module creates or implants in our reasoning system. The detailed information in our visual system about its internal states and the lighting conditions is mostly discarded or converted to a simple belief only about an object.

Given the discussion above, we can now begin more concretely to discuss the cognitive process which is the source of our color beliefs. The relevant process is the one that starts at the eye and ends with beliefs created by the visual system and implanted in the reasoning system. The input to the process is light hitting the retina and the output is a set of coarse beliefs about objects and their colors (we don't usually have beliefs about subtle shade properties of objects unless we compare them side by side).

A Process Reliabilist Theory of Justification

Process reliabilism is an externalist theory of justification put forward by Alvin Goldman (1979). According to Goldman a belief is justified if it is produced by a reliable (truth conducive) process. How can the process described above be truth conducive given the fact that it produces mostly false beliefs? It can't. But building on the intuitions developed in the discussion of the geometry of space, I propose that the definition of a reliable process needs to be expanded. A reliable process would be one that is either truth conducive or one that consistently produces good approximations (in the limited sense discussed above) of physical properties. Intuitively, a process can be reliable because its results are true or it can be reliable because "it gets the work done" even though it does not produce true results.

For example, the process of multiplying the diameter of a circle by 3.14 is not conducive to finding its true circumference, but it is still reliable because it is conducive to producing good approximations to the circumference. A person who didn't know that π was different from 3.14 would be justified in believing that the circumference he computes is the right one. Newton's theory is not truth conducive but it reliably produces good approximations (in everyday situations).

Goldman defines a process as reliable if it provides true beliefs a certain percentage of the time. I suggest that a process that provides no true beliefs whatsoever can also be reliable. Again, consider the following process of determining the circumference of circles: measure the diameter using a ruler and multiply it by 3.14 and then come to the belief that the circumference is the result of the multiplication. My intuition is that this process is a reliable one even though it is highly unlikely to produce any true beliefs. I would venture to say that most people if asked if the process I just described is reliable would answer in the affirmative.

In fact, most non-scientific processes that entail measuring a physical property are not going to produce true beliefs. An example of such false belief is the belief that "the result of the measurement is 53.3° Kelvin". Scientists explicitly acknowledge this by providing error bars around measurements. Measurements are never taken as true but usually as an indication of the mean of the range in which the true measurement would likely be found. Thus the statement "the result of the measurement is 53.3° Kelvin" would really mean to scientists something like "there is a likelihood of 90% that the true measurement of the temperature falls in the range of 53.3° Kelvin plus or minus 0.4°" and the latter sentence is true.

Humans in their non-reflective mode do not make this distinction and therefore our beliefs about values of physical properties are by and large false. When we measure the height of our kids or their temperatures we form beliefs such as "Tom is 4 feet and 5 inches tall" or "Tom has a fever of 39.5° Celsius". It is highly unlikely that these beliefs are true. What we measure are good approximations of height and temperature but not the true height and temperature.

According to my theory color is a physical property which the visual system measures or estimates in order to represent in several different ways. These measurements are then converted into the belief format that our reasoning faculties use in their epistemological endeavors. The format conversion though is not lossless and in fact very simple beliefs are created relative to the amount of information contained in the visual system. Since the visual system adapts and also compensates for changes in lighting the creation of the following kinds of beliefs is possible since the information to create them is available:

1) The color of the object is blue but without the color constancy computations it would have been light blue.

2) The color of the object is blue but this may change since the visual system needs more time to adapt.

3) The lighting is low relative to typical daylight and the color constancy computations cannot compensate for that; however the best estimate for the color of the object is blue.

4) The lighting is very low and not good enough for the visual system to detect hues. Therefore the best it can do is assign the color light gray to the object. However, it is quite likely that in typical daylight, the color of the object would be different.

But none of the complex beliefs above is ever formed. The format conversion process creates only very simple beliefs and disregards most of the information in the visual system. In all the above cases the belief that would be formed regarding color is simply that it is blue or gray. As was discussed previously, there is just no need for a more complex analysis of color in everyday situations and in fact a more complex belief system about colors would hinder humans. That is what makes the beliefs that are actually formed good approximations since they are commensurate with our goals. And because of that our common sense color beliefs are justified (but false).

Addressing the Problem of Interpretation

My view then is that the content of our color beliefs are false but that we are justified in holding them. This view is quite different than that of David Lewis who in

"New Work for a Theory of Universals" (Lewis 1983) argues not only that our beliefs about color properties are true, but that they are about natural properties.

Lewis arrives at this conclusion from the premise that it is the only one that would allow solving the problem of interpretation. Lewis argues that the only way to solve the problem of interpretation is to accept the following constraint: "We have no notion how to solve the problem of interpretation while regarding all properties as equally eligible to feature in content. For that would be to solve it without enough constraints. Only if we have an independent, objective distinction among properties, and we impose the presumption in favor of eligible content a priori as a constitutive constraint, does the problem of interpretation have any solution at all."

Lewis argues that we need natural properties because they will "impute a bias toward believing that things are green rather than grue". Lewis believes that when someone believes that the lawn is green he has the belief that the lawn has some property. Lewis also believes that the reason the person has the belief that the lawn is green rather than the belief that it is grue is because green is a more natural property than grue. Lewis does not give an explicit definition of what natural properties are, but in general they are what physics is in the business of discovering. He also says that non-disjunctive properties are less natural than disjunctive ones (in his discussion of adding and quadding he says that "but quadding is worse by a disjunction").

Lewis' view is very different from my view because it not only posits that when we believe that the lawn is green we believe that the lawn has a specific property but also that this property has some specific features that make it prone to being natural (Lewis does not argue that we know about or are aware of these features). Lewis would argue that our belief that the lawn is green is true because the content of our beliefs are constrained to being more natural properties and therefore the "green" part of the belief that the lawn is green refers to some natural property (even if we don't know what this natural property is).

I believe that science has produced strong counterevidence to Lewis' view, at least in the case of colors. Let's define a property called broange. Something is broange if it is brown or it is orange when viewed through a reduction tube. When we see something brown we have a strong bias to believe that it is brown and not broange. I think this is beyond dispute. According to Lewis, the bias is because brown is a more natural property than broange. Based on what science tells us, this is not true. In fact, broange is a more natural property than brown. Brown is a contrast color and any object that looks brown when viewed through a reduction tube looks orange. Not only that, but Lewis would view broange as disjunctive and brown as non-disjunctive thus additionally favoring brown as the more natural property. It seems to me that broange is more of a property revealed by science than brown is. It more accurately reflects our scientific knowledge than brown does. In other words, calling something broange is more accurate than calling it brown, if we take into account what science tells us. Also, though both brown and broange are categories that depend on humans, broange does so to a lesser extent. This is the case because brown depends on how humans look at the object while broange depends less on how exactly the object is viewed by humans.

Nevertheless, even if the Lewis view is incorrect, an implicit argument against my view can be found in Lewis' argument: If indeed the content of our color beliefs are as I argue, and not as Lewis posits, the problem of interpretation remains unsolved. My response to this is as follows. As I previously argued, the content of our color beliefs is determined by the process that interfaces our visual system and our reasoning system. Let's call this process VtoR. This process formats the outputs of the visual system and creates our perceptual beliefs. As described above, much information is lost during this formatting process. Most of the information available to our visual system is not passed on to our reasoning system.

We have no control whatsoever over VtoR and over how it works as VtoR is cognitively impenetrable. VtoR evolved in parallel or as part of our visual or reasoning systems not to satisfy the constraint of being true, but in order to maximize its usefulness to the survival of our species. To this end, it has on the one hand to provide the reasoning system with useful information but on the other hand not overload it with information that will hamper decision making by either making it take more time or require more energy without significantly improving the usefulness of the decisions.

Our visual system does not convey any information on color constancy corrections because this information is not useful in almost all everyday situations. We care about differentiating between objects, but the exact algorithm used to do this, or the parameters the algorithm used, are just not interesting to us as organisms. Why create the perceptual belief "the color of the object is blue but without the color constancy computations it would have been light blue" if the belief "the color of the object is blue" would suffice in almost all situations? The extra energy and time required for more complex computations with the more informative belief would not have given our ancestors any clear advantage, and therefore were probably selected against. VtoR is in essence following the platitude "keep it simple, stupid" or KISS.

Even if VtoR is not a result of a selection process but came about by chance, it is extremely plausible that it functions very similarly between different individuals. We all form very similar if not identical perceptual beliefs with the same level of detail. In the end, VtoR determines the content of our beliefs about color and it is not surprising that we understand each other, since the content of our color beliefs is created by basically the same cognitive process. So why do we all believe that the lawn is green rather than that it is grue? Not because green is a more natural property than grue. Rather it is because VtoR, the process that creates these beliefs, works in such manner; either by chance, or more likely because it was selected to function in this way. In general our folk ontology is detailed enough to allow us to function well in our environment, but not overly detailed as to swamp us with information that is not useful and would only hinder us. I am hinting here, but am not elaborating enough about a teleological theory of content that would replace Lewis' philosophical principle that we are biased towards natural properties. I concede that this is required in order to provide adequate refutation of Lewis. However, I will leave this project for a future date.

Conclusion

If color eliminativists are right, then Lewis is obviously mistaken, since language in the case of color would refer to no property at all (and obviously not to a natural property). My view is much more charitable than that of the eliminativists in two different dimensions. First, though I agree with the eliminativists that the common sense view of color is wrong, I argue that the extent of the mistake is not big and that the common sense view is a good approximation for my view. Eliminativists though are more likely to argue that color is a “grand illusion”. It is not a small mistake humans make they would argue, but a very substantial one.

The second dimension in which my view is more charitable is the epistemological one. I have shown that humans could be justified in their color beliefs. Justification is not as valuable as knowledge, but it is an important epistemological goal. Eliminativists so far have not put forward any argument showing how eliminativism is compatible with humans being justified in our common sense beliefs about color.

It would have been a wonderful coincidence if the scientific view of color and the common sense view turned out to be identical. But with hindsight, it is safe to say that that would have been just as likely as Maxwell’s equations being compatible with the common sense theory of what light is or the common sense view of planetary motion being identical with Newton’s theory of gravity. Just as humans use light to see without understanding what light is, they use color without fully understanding what it is. Gravity is very useful to us, even to humans who have never heard of Newton. An untold number of children have enjoyed playing with toys that include magnets without having any clue what magnetism is. Taking advantage of some physical phenomena does not require understanding it.

Gaining understanding and knowledge of physical phenomena is what science is in the business of doing. Over the last 300 years or so science has provided humanity with knowledge we did not previously possess. Much of the knowledge science produced was based on counterintuitive principles and metaphysics: Forces at a distance, light being both a particle and a wave etc. The scientific process in physics has culminated with general relativity and quantum mechanics which are almost completely incompatible with human intuitions and folk metaphysics. One of the cornerstones of modern biology is Deoxyribonucleic acid (DNA). DNA, RNA and many other foundational molecules of modern biology have no compatibility whatsoever with folk intuitions about life or reproduction. Chemistry has discovered that water is not a fundamental element and that it can be created out of two gases using electricity.

Given the above, should we really be surprised then that color is different from its common sense view? I don't think so. What seemed initially surprising, at least to me, is how good an approximation the common sense view of color is to the scientific view. But on second thought, this should not be that surprising either. Our common sense theory of color could not have served us so well unless it was a good approximation of reality.

Color is the most important property for our visual system. Without being able to differentiate between colors, we would not have been able to differentiate any other property using our visual system. If we cannot first differentiate an object from its background using color, we cannot determine its shape, size, kind, orientation, distance or any other property of the object. Given color's importance to us, it may be somewhat disconcerting for some that we don't fully understand it and that our common sense view of it is wrong. But, we are not living a grand illusion. Color is a property of objects just as we commonly believe. It is just a more complicated property than we realize. Furthermore, we are justified in our beliefs about color and that is an impressive epistemological achievement. Taking into account the enormous complexity of the world and our very limited cognitive capabilities, the facts above suggest a cause for admiration and not for concern.

Bibliography

Armstrong, D. (1968). A Materialist Theory of the Mind. Routledge, London.

Armstrong, D. (1987), Smart and the secondary qualities, in [Byrne and Hilbert 1997b]

Boghossian, P. A. and J. D. Velleman (1989) Colour as a secondary quality. Mind 98:81-103.

Boghossian, P. A. and J. D. Velleman (1991) Physicalist theories of color. Philosophical Review 100: 67-106.

Boyer J. L., Harrison S and Ro T. (2005), "Unconscious processing of orientation and color without primary visual cortex", PNAS, vol. 142, no. 46

Boynton, R. M. (1979). Human Color Vision. Holt, Rinehart and Winston, New York.

Brown, R. O. and MacLeod, D. (1997) "Color appearance depends on the variance of surround colors", Current Biology, 7:844–849

Byrne, A. and Hilbert, D. R. (1997a). Colors and reflectances. In Byrne, A. and Hilbert, D. R., editors, Readings on Color, Volume 1: The Philosophy of Color, pages 263–288. MIT Press, Cambridge, Massachusetts.

Byrne, A. and Hilbert, D. R., editors (1997b). Readings on Color, Volume 1: The Philosophy of Color. MIT Press, Cambridge, Massachusetts.

Byrne, A. and Hilbert, D. R. (2003). Color realism and color science. Behavioral and Brain Sciences, 26(1):3–64.

Byrne, A. and Hilbert, D. R. (2004). Hardin, Tye, and color physicalism. The Journal of Philosophy, CI(1):37–43.

Byrne, A. and Hilbert, D. R. (2006). Color primitivism. In R. Schumacher, editor, Perception and Status of Secondary Qualities. Kluwer, Dordrecht.

Campbell, J. (1993) A simple view of colour. In: Reality, representation and projection, ed. J. Haldane & C. Wright. Oxford University Press.

Cohen, J. (2003a). Color: A functionalist proposal. Philosophical Studies, 112(3):1–42.

Cohen, J. (2003b). Perceptual variation, realism, and relativization, or: How I learned to stop worrying and love variations in color vision. Behavioral and Brain Sciences, 26(1):25–26.

Cohen, J. (2003c). On the structural properties of the colors. Australaian Journal of Philosophy, 81(1), 78–95.

Cohen, J. (2004). Color properties and color ascriptions: A relationalist manifesto. The Philosophical Review, 113(4), 451–506.

Cohen, J. (2006a). Color constancy as counterfactual. Australaian Journal of Philosophy. in press.

Cohen, J. (2006b). A relationalist's guide to error about color perception. Noˆus. In press.

Colin A. (1997), "Animal cognition and animal minds", in P. Machamer & M. Carrier (eds.) Philosophy and the Sciences of the Mind: Pittsburgh-Konstanz Series in the Philosophy and History of Science vol. 4. Pittsburgh University Press and the Universitätsverlag Konstanz: pp. 227-243.

Cytowic, R. E. (2003) The Man Who Tasted Shapes, Cambridge, MA, MIT Press.

Drake, S., ed. (1957) Discoveries and opinions of Galileo. Doubleday.

Dretske, F. I. (1995) *Naturalizing the mind*. Cambridge, MA, MIT Press.

Gallistel, C.R. (2008) Learning and representation. In R. Menzel (Ed) Learning theory and behavior. Vol 1 of Learning and Memory - A Comprehensive Reference. 4 vols (J. Byrne, Ed). Oxford: Elsevier. pp. 227-242

Gallistel, C. R., & King, A. P. (2009). Memory and the computational brain: Why cognitive science will transform neuroscience. New York: Wiley/Blackwell

Goldman, A. (1979) "What is Justified Belief?" in George Pappas, ed., Justification and Knowledge, D. Reidel, pp. 1-23

Goldman, A. (1986) "Epistemology and Cognition", Harvard University Press

Goldman, A. (2002) "The Sciences and Epistemology," in P. Moser, ed., The Oxford Handbook of Epistemology (pp. 144-176). Oxford University Press

Hardin, C. L. (1988). Color for Philosophers: Unweaving the Rainbow. Hackett, Indianapolis.

Hardin, C. L. (2003). A spectral reflectance doth not a color make. The Journal of Philosophy, 100(4):191–202.

Hardin, C. L. (2005). A green thought in a green shade. Harvard Review of Philosophy.

Hilbert, D. R. (1987). Color and Color Perception: A Study in Anthropocentric Realism. CSLI, Stanford.

Hubel, D. H. (1988). Eye, Brain, and Vision. Scientific American Library, New York.

Hurvich, L. M. (1981). Color Vision. Sinauer Associates, Sunderland, Massachusetts.

Hurvich, L. M. (1997), "Chromatic and Achromatic Response Functions", in *Readings on Color, Volume 2: The Science of Color*, ed. A. Byrne and D. R. Hilbert (Cambridge, MA: MIT Press, 1997)

Hurvich, L. M. and Jameson, D. (1957). An opponent-process theory of color vision. Psychological Review, 64:384–403.

Jackson, F. & Pargetter, R. (1987) An objectivist's guide to subjectivism about colour. Revue Internationale de Philosophie 41:127–41.

Jackson, F. & Pargetter, R. (1997) An objectivist's guide to subjectivism about color. In: Readings on color, vol. 1: The philosophy of color, ed. A. Byrne & D. R. Hilbert. MIT Press.

Jakab, Z. and McLaughlin, B. (2003). Why not color physicalism without color absolutism? Behavorial and Brain Sciences, 26(1):34–35.

Johnston, M. (1992). How to speak of the colors. Philosophical Studies, 68:221–263. Reprinted in [Byrne and Hilbert, 1997b], 137–176.

Lamme V. A. (2006), Zap! Magnetic tricks on conscious and unconscious vision, Trends in Cognitive Science, 2006 May;10(5):193-5

Lewis, D. (1997). Naming the colors. Australasian Journal of Philosophy, 75(3), 325–342.

Maloney, L. T. (2003). Surface color perception in constrained environments. Behavioral and Brain Sciences, 26(1):38-39.

Matthen, M. (1988) Biological functions and perceptual content. *Journal of Philosophy* 85:5-27.

Matthen, M. (1999) The disunity of color. *Philosophical Review* 108(1):47-84.

Matthen, M. (2005). Seeing, Doing, and Knowing: A Philosophical Theory of Sense Perception. Oxford University Press, Oxford.

McDowell, J. (1985). Values and secondary qualities. In Honderich, T., editor, Morality and Objectivity: A Tribute to J. L. Mackie, pages 110–129. Routledge and Kegan Paul, London.

McGinn, C. (1983). The Subjective View: Secondary Qualities and Indexical Thoughts. Oxford University Press, Oxford.

McGinn, C. (1996). Another look at color. The Journal of Philosophy, 93(11):537–553.

McLaughlin, B. (2002) “Colour, Consciousness, and Colour Consciousness”, in Quentin Smith ed. Consciousness: New Essays (Oxford: Oxford University Press).

McLaughlin, B. (2003). The place of color in nature. In Mausfeld, R. and Heyer, D., editors, Colour Perception: Mind and the Physical World. Oxford University Press, New York.

Peacocke, C. (1984). Colour concepts and colour experiences. Synthese, 58(3):365–81. Reprinted in [Rosenthal, 1991], 408–16.

Smart, J. J. C. (1975). On some criticisms of a physicalist theory of colors. In C. Cheng, editor, Philosophical Aspects of the Mind-Body Problem. University Press of Hawaii, Honolulu. Reprinted in Byrne and Hilbert (1997b), 1–10.

Stroud, B. (2000) The quest for reality: Subjectivism and the metaphysics of colour. Oxford University Press.

Teller, D. (2003). Color: A vision scientist’s perspective, Behavorial and Brain Sciences, 26(1):48–49.

Thompson, E., A. Palacios & F.J. Varela (1992) Ways of coloring: Comparative color vision as a case study for cognitive science. Behavioral and Brain Sciences 15:1-74.

Thompson, E. (1995a). Colour Vision: A Study in Cognitive Science and the Philosophy of Perception. Routledge, New York.

Thompson, E. (1995b) Colour vision, evolution, and perceptual content. Synthese 104:1-32.

Tye, M. (2000). Consciousness, Color, and Content. MIT Press, Cambridge, Massachusetts.

Curriculum Vita: Eliezer Mintz

Education:

Dates	*School*	*Degree*
1982-1985	Hebrew University, Jerusalem, Israel	B.Sc., Physics, Mathematics and Computer Science (1985)
1990-1991	INSEAD, Fontainebleau, France	MBA (1991)
2004-2009	Rutgers University, New Brunswick, NJ	Ph.D., Philosophy (2009)