

In Defense of Incompatibility, Objectivism, and Veridicality about Color.

Pendaran Roberts & Kelly Schmidtke¹

Published in Review of Philosophy and Psychology (2012).

The final publication is available at

<http://www.springerlink.com/content/g587787114328685/>.

Whether the following propositions about the colors are true is the subject of heated debate (for examples see Hacker, 1991; Byrne & Hilbert, 2003; Cohen, 2004): No object can be more than one determinable or determinate color all over at the same time (Incompatibility); the colors of objects are mind-independent (Objectivism); and most human observers usually perceive the colors of objects veridically in typical conditions (Veridicality). Recently, the truth of these propositions has been called into question by empirically based arguments of the following form: There is *mass* perceptual disagreement about the colors of objects amongst human observers in typical conditions (P-Disagreement); therefore, at least one of the three propositions Incompatibility, Objectivism, or Veridicality is false (Hardin, 1988; 2003; 2005; Cohen, 2004; 2006; 2009; Cohen et al, 2007; Kalderon, 2007; Mizrahi, 2007; see also McLaughlin, 2003; Chalmers, 2006). In this article, we defend Incompatibility, Objectivism, and Veridicality by calling into question whether the empirical literature really supports P-Disagreement.

Allen (2010) has questioned whether P-Disagreement is really supported by the empirical literature. Allen points out that the amount of disagreement measured by experiments depends on many methodological factors, including but not limited to the participants' native language (Berlin & Kay, 1969), age (Scheffrin & Werner, 1990), and retinal illuminance (Ayama et. al.,

¹ One can contact Pendaran Roberts at apxpr@nottingham.ac.uk, and Kelly Schmidtke at schmike@tigermail.auburn.edu.

1987). Moreover, Allen argues that the amount of disagreement is highly affected by the particular color(s) studied, with unique green causing the most disagreement. Thus, when analyzing multiple studies it is important to remember that disagreement will arise not only due to participants' color perception but also due to many methodological differences between the studies (Ayama et. al., 1987; Kuehni, 2004).

We admire Allen's attempt to bring some needed scrutiny to whether the empirical literature really supports P-Disagreement. However, one plausibly confounding factor that most philosophers including Allen have overlooked is task type.² In the present article, we argue that the type of task employed in most empirical studies that appear to support P-Disagreement calls this support into question. Broadly speaking, there are two types of tasks used in the empirical literature, matching and naming tasks.

Matching Tasks =df An experimental procedure in which participants are presented with at least two colored examples and asked whether they look the same or different.

Naming Tasks =df An experimental procedure in which participants are presented with colored example(s) and asked to name them.

Naming tasks depend on participants' having concepts of the particular colors (color concepts) associated with color words, while matching tasks do not so depend. Naming tasks require participants to say, for instance, which example is best described by a certain color term (e.g. 'unique red'). So, in order for participants to complete a naming task, they must have a color concept associated with the given color word. In contrast to naming tasks, matching tasks only require participants to be able to say whether colored examples look the same or different.

² Byrne and Hilbert (2007, footnote 5) mention that task type is important but other than this example we are unaware of any philosophers who have talked about task type in relation to P-Disagreement.

Hence, in order to complete a matching task, participants are not required to have, for instance, a color concept associated with the word ‘unique red.’ Of course, we admit that matching tasks require concepts. Matching and naming tasks require, amongst others, both the concept of something being colored and the concept of something being the same or different. Our claim is merely that matching tasks do not depend on participants having color concepts associated with color words, while naming tasks do so depend (Jordan & Mollon, 1994).

Philosophers use empirical studies to support P-Disagreement. An overview of the literature reveals that the majority of the studies that appear to support P-Disagreement use naming tasks not matching ones (see Allen 2010 for a review). Let us look at the tasks used by a few of them. Perhaps the most widely recognized large-scale color research is the World Color Survey (Cook, Kay & Reigier, 2011).³ In this research, participants are first presented with various differently colored chips one at a time and for each chip asked what basic color term best describes it. In a later task, participants are shown all the differently colored chips at once and asked to say which chip is best described by a given color term. Although not part of the World Color Survey, Wuerger, Atkinson, and Cropper (2005) use a similar task. In their task, participants are given a color term and then view 12 differently colored circles on a computer. After selecting the colored circle that is best described by the given color term, the computer presents 12 colored circles that give off a narrower range of wavelengths from which the participant again selects the best example for the given color term.

In other researchers’ naming tasks, the wavelengths that cause experiences as of the unique colors are determined by presenting participants with colored examples and asking

³ One may argue that the world color study deals with focal colors, not unique colors. This discrepancy is not problematic to our claims as Miyahara (2003) found that participants’ mean focal colors and unique hues were strikingly similar for red, green, blue and yellow.

whether they contain too much of a neighboring color to be a unique color. For example, when measuring unique green, participants see a green-ish color and say whether it contains too much blue or yellow to be unique green. If the participant says, for instance, that the example contains too much blue, then the experimenter adjusts the color by reducing the example's blueness. This process repeats for a pre-determined number of trials or until the participant indicates that the example is unique green (Scheffrin & Werner, 1990; Malkoc et al., 2005, Unique Hue Settings; Webster et al., 2000). It is important to note that many experiments that do not consist entirely of naming tasks have a naming task component. For example, Ayama et al. (1987) asked participants to *name* unique color examples at different illuminances. Then participants were presented with a new color example and asked to *match* this new color with their previously chosen unique colors at different illuminances.

We suspect that the additional conceptual factor in naming tasks (the color concepts participants' associate with color words) is responsible for a lot of the empirically measured disagreement that appears to support P-Disagreement. Thus, we expect results based on naming tasks to vary considerably more across participants than the results of matching ones. Our hypothesis implies that conceptual factors not perceptual ones explain a notable amount of the empirically measured variation. P-Disagreement is the proposition that there is mass *perceptual* disagreement about the colors of objects amongst human observers in typical conditions. Thus, if our hypothesis is correct, then P-Disagreement is in a precarious position. Without strong empirical support, we ought to seriously question whether P-Disagreement is true, as the premise is in conflict with common sense. According to common sense there is little perceptual disagreement amongst most human observers in typical conditions.

Our aim in calling P-Disagreement into question is to provide a defense of Incompatibility, Objectivism, and Veridicality from the empirically based form of argument under consideration. A defense of these propositions is an attack against selectionism, relationism, and eliminativism. This is because selectionism (the view that our visual systems select different properties to be the colors) implies that Incompatibility is false (Allen, 2010; Kalderon, 2007; Mizrahi, 2007); relationism (the view that the colors are relational properties that combine objects and perceivers) implies that Objectivism is false (Cohen, 2004; 2006; 2007; 2009), and eliminativism (the view that external objects are not colored) implies that Veridicality is false (Boghossian and Velleman, 1989; Chalmers, 2006; Hardin, 1988; Maund, 1995; and Pautz, 2006).⁴ Moreover, a defense of the relevant propositions is also a defense of the views we favor, primitivism (the view that the colors are non-reducible properties) and physicalism (the view that the colors are properties like those mentioned in modern physics). The reason for this is that both primitivism (Campbell, 1993; Westphal, 1987; 2005; Watkins, 2002; 2005) and physicalism (Byrne & Hilbert, 2003a) (as often conceived) are species of realism and objectivism about color.

The experiment

We ran an experiment to test our hypothesis that the relevant conceptual factor is responsible for a lot of the empirically measured variation. In a classroom lit with Philips Master 26W/840/P4 bulbs, 24 philosophy students (19 male) at the University of Nottingham voluntarily participated in our study to understand visual color disagreement, none of who acknowledged having any

⁴ Some believe that dispositionalism is a form of relationalism (for example see Cohen, 2009). For the purposes of this article, we understand relationalism to be incompatible with dispositionalism, because relationalism but not dispositionalism is the denial of Objectivism.

color deficiency.⁵ Each participant received two worksheets: half the participants received a green matching worksheet and a red naming worksheet, while the other half were given a red matching worksheet and a green naming worksheet. Participants could modify the angle they viewed the sheets from as they desired. The order of the worksheets and the items within the worksheets were counterbalanced across participants, and no order effects emerged.

The colored items were constructed using Microsoft Publisher's CMYK color index.⁶ Unique green and unique red were defined so that CMYK for green = 100.0.100.36 and CMYK for red = 0.100.100.20.⁷ The worksheets were printed using a professional grade color printer, run by a professional print shop, called 'The Xerox 700 Digital Color Press,' which has received the FOGRA Validation Print Certification measuring color accuracy and consistency.

Both the matching and naming worksheets contained 20 pairs of items (see figure 1, but keep in mind that the color accuracy will depend on the monitor or printer used). Each pair contained a standard item positioned left of a comparison item. The standard items on the matching worksheets were rectangles colored to exemplify unique green or unique red. In contrast, the standard items on the naming worksheets were the words "True Green" or "True Red."⁸ The comparison items for both worksheets were colored rectangles. Only one of the comparisons matched the standard. The remaining 19 comparisons differed from the standard in

⁵ A likely concern is to worry about the fluorescent lighting used in the room in which we conducted our experiment. Since we were concerned with comparing disagreement in the naming task with disagreement in the matching one, the only reason to be worried about the lighting would be if there were good reasons to suspect that it differentially affected our tasks. However, not only are our results for each task independently predicted by the psychophysical data (see p. 8-11), but also our pilot experiment conducted in natural daylight found comparable results (see p. 12).

⁶ The letters in the initialism 'CMYK' stand for Cyan, Magenta, Yellow, and Black respectively.

⁷ The black ink was added so as to decrease the lightness and prevent the items from appearing washed out. The green items needed more black ink than the red ones to obtain this goal.

⁸ We used the terms "True Green" and "True Red," because it is our understanding that these terms in the vernacular mean what "unique green" and "unique red" mean respectively to color scientists. We defined "true green" and "true red" for our participants in the same way that "unique green" and "unique red" are defined in color science.

that they contained different amounts of cyan, magenta, or yellow by 5 unit steps. Instructions on the top of both worksheets read, “Circle the box that contains the same items.” The instructions were also verbally explained. We explicitly told participants that they were only to circle one pair for the matching task and one pair for the naming task.

After checking that our participants followed the directions correctly, their responses were entered into analyses as the number of units that their chosen comparison differed from the standard. The standard deviation was highest for the green naming worksheet ($SD = 9.41$), followed by the red naming worksheet ($SD = 8.38$), and lastly the green and red matching worksheets, which both had the same standard deviation ($SD = 4.52$) (see figure 2). The Brown-Forsythe test for equality of variances was selected to compare these groups, as this test has greater power than other tests designed to compare variance with non-normal distributions (Conover, Johnson & Johnson, 1981; Algina, Olejnik, & Ocanto, 1989). These tests revealed that the naming worksheets produced more disagreement than the matching worksheets ($F(1, 46) = 7.93, p < 0.01$). This difference was significant for the green worksheets ($F(1, 22) = 6.06, p < 0.02$) but not the red worksheets ($F(1, 22) = 2.37, p > 0.05$). There was no significant difference between the green and red matching worksheets, and no significant difference between the green and red naming worksheets.

Discussion

Our results show that significantly more interpersonal disagreement emerges in naming than matching tasks. So, our results should cast doubt on whether P-Disagreement is true by supporting our hypothesis that the additional conceptual factor in naming tasks (the color

concepts participants' associate with color words) account for a lot of the empirically measured variation. P-Disagreement implies that at least one of the propositions Incompatibility, Objectivism, or Veridicality is false. Thus, by showing that P-Disagreement is in trouble, we have provided a defense of these three propositions about the colors. A defense of the relevant propositions is an attack against selectionism, relationalism, and eliminativism as well as support for physicalism and primitivism.

A competing hypothesis that would support P-Disagreement is that our results are explained by widespread color transformations rather than the relevant conceptual factors. There are three (approximate) reflectional symmetries in color space that would allow for three transformations all of which would be largely if not completely undetectable: red-green inversion, blue-yellow and black-white inversion, and complete inversion (Palmer, 1999). If these three inversions were distributed amongst the population, then there would be mass perceptual disagreement about the colors of objects but no reason to suspect increased disagreement about whether they look the same or different. Thus, if these three inversions were distributed amongst us, it may seem that there would be more disagreement in naming tasks than matching ones, and this is exactly what we find. Our opponent may have similar expectations for at least some of the behaviorally detectable transformations.

Whether color transformations (behaviorally detectable or not) are widespread is an important question, but we do not think that they can explain the results of our experiment. Our argument is as follows: One would associate different color concepts with color words dependent on how one's color space was transformed. For example, someone who was red-green inverted would associate the concept of being green with the word 'red' and the concept of being red with the word 'green,' and so, despite the inversion in how things phenomenally look, he would

verbally agree with the non-inverted that, for instance, the forest is green and that fire trucks are red. In naming tasks, participants are presented with colored sample(s) and asked to name them. One can only name the relevant examples using one's color words. Thus, inversions in color space cannot explain why we found more disagreement in our naming task than our matching task. Of course, if one believes that such inversions obtain based on other grounds, one is going to be unconvinced by our experiment that P-Disagreement is in trouble. Regardless, our results support our conceptual hypothesis and not the color transformation one.

Another worry is that we only found that naming tasks result in significantly more disagreement than matching ones with respect to unique green. It is our opinion that this finding is sufficient to call P-Disagreement into question. A review of the empirical literature (Allen, 2010) reveals that disagreement primarily occurs with respect to unique green, while significantly less disagreement emerges with other unique colors. For example, Kuehni (2004) reviewed 10 color experiments, and found that the variation with unique green (*Mdn Range* = 62 nm) was larger than both unique blue (*Mdn Range* = 21.5 nm) and unique yellow (*Mdn Range* = 9 nm).⁹ Using Munsell chips, The World Color Survey supports the same pattern, with the most variation arising with green (*VAR* = 3.01 chips) then blue (*VAR* = 2.45 chips), red (*VAR* = 0.46 chips) and finally yellow (*VAR* = 0.31 chips) (Webster & Kay, 2005). Therefore, supporting our hypothesis with respect to unique green is sufficient to call P-Disagreement into question.

A third concern is that matching tasks also suggest disagreement, and so is the disagreement measured using these tasks not enough to support P-Disagreement? While matching tasks do produce disagreement, the disagreement is dramatically less than that in

⁹ The largest exception to this general pattern is Ayama et al. (1987). The small number of participants ($N = 2$) in Ayama et al.'s experiment can explain the observed deviation, because such small numbers of participants likely make variation larger. Also, it is interesting to note the absence of data available for unique red in Ayama et al.'s experiment. Unique red is a non-spectral color and the tasks compared by Ayama et al. all use spectral stimuli.

naming tasks, as evidenced by the present experiment as well as many others. One frequently used matching task is the Rayleigh match. In this task, participants view a split stimulus where the first half emits a combination of wavelengths (e.g. green + red) and the second half emits a single wavelength (e.g. yellow). The participant then adjusts the light waves emitted by the first half so that it matches the second half. The researcher then records the percentage of red light in the first stimulus so that the scores range from 0 to 1. It has been found that participants' mean Rayleigh matches vary little across (from $M = 0.547$ to 0.555) and within experiments (from $SD = 0.021$ to 0.037) (Lutze et al., 1990, Table 5).

In other research using matching tasks, it was found that people with clinically defined normal color vision can distinguish between wavelengths of 1 nm at the middle of the spectrum where green is located (500 and 600 nm). People's ability to distinguish wavelengths does degrade to about 6 nm at the ends of the spectrum where violet and red are located (Wright & Pitt, 1934), but recall that an nm is only one billionth of a meter. Of course, people with color deficient vision vary more in color perception tasks than those with normal color vision (Barbur, 2008), but only a small minority of the population (about 4%) have color deficiencies. Thus, while matching tasks do suggest disagreement, the disagreement is not nearly as much as naming tasks suggest and certainly not sufficient to support P-Disagreement. On the contrary, the results of matching tasks are what we would expect if P-Disagreement were false.

We pointed out in the introduction that Ayama et al's (1987) experiment includes a matching component, and so this study's results are important to addressing the present concern about matching tasks. In support of Incompatibility, Objectivism, and Veridicality, the results were that participants demonstrated a spectacular ability to match new colors with previously chosen ones. The location of unique colors did change with retinal illuminance for both tasks but

never so much as to cause disagreement about unique colors (e.g. examples identified as unique blue were never identified as unique green). The disagreement that was measured by Ayama et al. must be cautiously considered. First, the experiment included a naming task component and so is susceptible to our general worry about using such tasks to support P-Disagreement. Second, like many color studies this experiment included a small number of participants ($N = 2$), and studies that use such small numbers of participants are insufficient to provide more than a modicum of support for P-Disagreement.

A final concern may result from wondering about the impact that the large-scale disagreement evidenced by naming tasks has on P-Disagreement. In reply, P-Disagreement is that there is mass *perceptual* disagreement about the colors of objects amongst human observers in typical conditions. Our experiment suggests that the additional factor in naming tasks (the color concepts participants' associate with colors words) accounts for a lot of the empirically measured variation. Hence, in order to support P-Disagreement using naming tasks, it must be that the color concepts we associate with color words affects the colors that our visual systems represent, but it is unclear whether the relevant factor can do what is required. In fact, whether this factor can influence what our visual systems represent touches on contentious issues relevant to the debate about perceptual content (McDowell, 1994; Tye, 2000) and cognitive penetration (Raftopoulos, 2005; Macpherson, 2012).¹⁰ Thus, although the disagreement evidenced by

¹⁰ One reason to be suspicious as to whether the relevant factor can influence what our perceptual systems represent is that it would seem that how the colors phenomenally look to people does not change based on their concepts of the particular colors. Here is an argument: Assume that how the colors phenomenally look changes based on peoples' concepts of them. Necessarily, if two experiences E1 and E2 differ in phenomenal character, then they differ in representational content (Representationalism). So, we get that when someone first forms a concept of a color like aquamarine, he comes to represent something new. However, the correct view of what is happening when someone first forms the concept of a color like aquamarine is that he has come to have the concept of the color property represented by his visual system when in his life he had phenomenally aquamarine experiences. Thus, we can conclude that how the colors phenomenally look does not change based on peoples' concepts of them.

naming tasks is relevant to the debate about P-Disagreement, the relevance has no impact on our argument to the effect that P-Disagreement is in an unstable position.

Why is there a significant difference between our green naming and matching worksheets but not one between our red naming and matching worksheets? We believe that this result is best explained by the participants associating either (1) a larger number of relatively narrow concepts with the word ‘unique green’ or (2) one or more broader concepts with ‘unique green.’¹¹ If (1) were true, this would mean that the question being considered is primarily explained by there being more disagreement between people about which concept is associated with the word ‘unique green’ than there is about which concept is associated with ‘unique red.’ If (2) were true, this would mean that the question is mostly explained by there being many more greens that satisfy the concept people associate with the word ‘unique green’ than reds that satisfy the concept people associate with ‘unique red.’

An alternative explanation is that the participants’ concepts were comparable, but that they had less plausible examples of unique red from which to choose than unique green. Some of the examples on our red worksheet look orange to us, while all the examples on our green worksheet look green. We do not think that this alternative explanation is correct. In a pilot experiment very similar to the main experiment of this paper, which was conducted in a room with natural daylight, the comparisons for red differed from the standard by 3 unit steps while the comparisons for green differed by 5 (as opposed to both differing by 5 in the present experiment), but this did not make a difference. The worksheets with green items revealed that the naming worksheet generated more variation than the matching worksheet; in contrast, the worksheets with red items generated similar variation.

¹¹ A concept is only broad or narrow relative to another concept. A concept C1 is broader than another C2 =df a greater number of differing entities can satisfy C1 than C2. A concept C1 is narrower than another C2 =df a fewer number of differing entities can satisfy C1 than C2.

With respect to our preferred explanation, the question remains whether our participants had a larger number of relatively narrow concepts or one or more broad concepts associated with the word ‘unique green.’ As Hardin (1988) reports, peoples’ unique color settings remain stable and reliable even for experimental sessions that are weeks apart (p. 39). In other words, there is a lot of intrapersonal consistency in the samples that people say fall under, for instance, the concept associated with ‘unique green,’ and so it would seem that the participants sharing one or more broad concepts cannot explain our results. Nevertheless, there is reason to be concerned about whether what Hardin reports can be used in this way. The reason is that plausibly the answer that a participant gives to the question, for example, “Which colored sample is unique green?” in the first session of an experiment designed to test his unique color settings has some influence on the color concept he associates with ‘unique green.’¹²

Regardless of whether our participants had a larger number of relatively narrow concepts or one or more broad concepts associated with the word ‘unique green’, it is important to appreciate that participants recognize more different greens than different reds. On a 160 Munsell chip array about 30% of the chips are described as falling under the basic color term ‘green,’ while less than 10% are described as falling under the basic color term ‘red’ (Roberson et al., 2000). Our rationale for thinking that this is important is twofold: (a) If people can distinguish between more greens than reds, then plausibly during our lives we encounter more objects that look not to be red than not to be green. So, during our lives we are likely presented

¹² In order to test whether participants have a larger number of relatively narrow concepts or one or more broad concepts associated with ‘unique green,’ we propose a one trial test thus avoiding the above worry. In our proposed test, participants would be presented with either green or red worksheets like the naming sheets in figure 1. Participants who received the red worksheet would be asked to circle every box that exemplifies unique red and participants who received the green worksheet would be asked to circle every box that exemplifies unique green. It is important that participants be informed that a unique color is one that appears to have no neighboring hues in it, as plausibly they would not know what the word ‘unique’ means given the context. Our pilot experiment (which was very similar to the main experiment of this paper) suggests that if participants are not specifically instructed to only circle one box, they will circle multiple boxes to exemplify a unique color.

with more plausible samples for unique green than unique red. (b) There is empirical evidence that both the variation in the association of concepts with words across individuals and the narrowness of a concept within individuals depend on the number of different examples used during instruction (Posner & Keele, 1968; Heit & Feeney, 2005).

With regard to the association of concepts with words, Fried & Holyoak (1984) found that participants trained using myriad visually disparate examples (e.g. more dissimilar checkerboard patterns) exhibit more variable performance when asked to perform certain relevant tasks than people trained using less disparate examples (e.g. more similar checkerboard patterns).¹³ Regarding the narrowness of concepts, French et al. (2004) conducted a study in which infants were familiarized with either pictures of cats or dogs. The cats represented by the cat pictures were highly variable in their features (e.g. ear and hair length), while the dogs represented by the dog pictures were less variable. After repeated exposure, the infants were presented with both a novel cat and a novel dog picture. Those infants who were familiarized with cat pictures showed no preference for looking at either novel picture, suggesting that they had formed a broad concept satisfied by both dogs and cats. In contrast, those infants who had been familiarized with dog pictures preferred looking at the novel cat picture, suggesting that they had formed a narrower concept that was not satisfied by cats.

In addition to people recognizing a lot more greens than reds, they also recognize a lot more blues than reds (Roberson et al., 2000). Thus, just as we found significantly more variation with the green naming than matching task, we expect more variation in blue naming tasks than matching ones. This consequence not only shows a plausible way of testing our preferred

¹³ The results found with arbitrary stimuli, such as checkerboard patterns, extends to more natural stimuli such as speech sounds (Wade, Jongman & Sereno, 2007) and to detection of dangerous items in a briefcase via an X-ray image (Gonzalez & Madhavan, 2010).

explanation but also further weakens the support that the empirical evidence for mass interpersonal variation provides for P-Disagreement.

References

- Algina, J., Olejnik, S., & Ocanto, R. (1989). Type I error rates and power estimates for selected two-sample tests of scale. *Journal of Educational Statistics, 14*, 373-384.
- Allen, K. (2010). Locating the unique hues. *Rivista di Estetica, 43*, 13-28.
- Ayama, M., Nakatsue, T., & Kaiser, P. (1987). Constant hue loci of unique and binary hues at 10, 100, and 1000 Td. *Journal of the Optical Society of America A, 4*, 1136–1144.
- Barbur, J. L, Rodriguez-Carmona, M., Harlow, J. A., Mancuso, K., Neitz, J., & Neitz, M. (2008). A study of unusual Rayleigh matches in deutan deficiency. *Visual Neuroscience, 25*, 507–516.
- Berlin B., Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution*. Berkeley: University Press.
- Boghossian, P. A., & Velleman, J. D. (1989). Colour as a secondary quality. *Mind, 98*, 81-103.
- Byrne, A., & Hilbert, D. R. (2003). Color realism and color science. *Behavioral and BrainSciences, 26*, 3-21.
- Byrne, A., & Hilbert, D. (2007). Truest blue. *Analysis, 67*, 87-92.
- Campbell, J. (1993). A simple view of colour. In J. Haldane & C. Wright, eds., *Reality, Representation and Projection*. Oxford: Oxford University Press.
- Chalmers, D. J. (2006). Perception and the fall of Eden. In T. Gendler & J. Hawthorne, eds., *Perceptual Experience*. Oxford: Oxford University Press.
- Cohen, J. (2004). Color properties and color ascriptions: A relationalism manifesto. *Philosophical Review, 113*, 451-506.
- Cohen, J. (2006). Colour and perceptual variation revisited: Unkown facts, alien modalities, and perfect psychosemantics. *Dialectica, 60*, 307-19.
- Cohen, J. (2007). A relationist guide to error about color perception. *Nous, 41*, 335-353.

- Cohen, J. (2009). *The red and the real*. Oxford University Press.
- Cohen, J., Hardin, C. L., & McLaughlin, B. P. (2007). The truth about “the truth of true blue.” *Analysis*, 67, 162-166.
- Conover, W. J., Johnson, M. E., & Johnson, M. M. (1981). Comparative study of tests for homogeneity of variances: with applications to the outer continental shelf bidding data. *Technometrics*, 23, 351-361.
- Cook, R., Kay, P., & Regier, T. (2011). The World Color Survey. Retrieved November 15, 2011, <http://www.icsi.berkeley.edu/wcs/>.
- French, R. M., Mareschal, D., Mermillod, M. & Quinn, P. C. (2004). The Role of Bottom-Up Processing in Perceptual Categorization by 3- to 4-Month Old Infants: Simulations and Data. *Journal of Experimental Psychology: General*, 133, 382-397.
- Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 10, 234-257.
- Gonzalez, C. & Madhavan, P. (2010). Diversity During Training Enhances Detection of Novel Stimuli. *Department of Social and Decision Sciences. Paper 116*.
<http://repository.cmu.edu/sds/116>.
- Hacker, P. M. S. (1991). *Appearance and Reality*. Oxford: Basil Blackwell.
- Hardin, C. L. (1988). *Color for Philosophers*. Indianapolis: Hackett.
- Hardin, C. L. (2003). A spectral reflectance doth not a color make. *Journal of Philosophy*, 100, 191-202.
- Hardin, C. L. (2005). A green thought in a green shade. *Harvard Review of Philosophy*, 12, 29-39.
- Heit, E., & Feeney, A. (2005). Relations between premise similarity and inductive strength. *Psychonomic Bulletin & Review*, 12, 340–344.

- Jordan, G. & Mollon, J. D. (1995). Rayleigh matches and unique green. *Vision Research*, 35, 613-620.
- Kalderon, M. E. (2007). Color pluralism. *The Philosophical Review*, 116, 563-601.
- Kuehni, R. G. (2004). Variability in unique hue selection: a surprising phenomenon. *Colour Research and Application*, 24, 158-162.
- Lutze, M., Cox, J., Smith, V. C., Pokorny, J. (1990). Genetic Studies of Variation in Rayleigh and Photometric Matches in Normal Trichromats. *Vision Research*, 30, 149- 162.
- Macpherson (2012). Cognitive penetration of colour experience: Rethinking the issue in light of an indirect mechanism. *Philosophy and Phenomenological Research*, 84, 24-62.
- Maud, B. (1995). *Colours: Their nature and representation*. NY: Cambridge University Press
- McDowell, J. (1994). *Mind and World*. Cambridge, MA: Harvard University Press.
- Malkoc, G., Kay P., & Webster, M. A. (2005). Variations in normal color vision. IV. Binary hues and hue scaling. *Journal of the Optical Society of America A*, 22, 2154–68.
- Mclaughlin, B. P. (2003). The place of color in nature. In R Mausfeld & D. Heyer, eds., *Colour Perception: Mind and nature*. Oxford: Oxford University Press.
- Miyahara, E. (2003). Focal colors and unique hues. *Perceptual Motor Skills*, 97, 1038–1042.
- Mizrahi, V. (2007). Color objectivism and pluralism. *Dialectica*, 60, 283-306.
- Palmer, S. (1999). Color, consciousness, and the isomorphism constraint. *Behavioral and Brain Sciences*, 22, 923-943.
- Pautz, A. (2006). Color Eliminativism. Retrieved January 4, 2012, from <https://webspace.utexas.edu/arp424/www/elim.pdf>.
- Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77, 353–363.

- Raftopoulos, A. (2005). *Cognitive Penetrability of perception: Attention, strategies, and bottom-up constraints*. New York: Nova Science.
- Roberson, D., Davies, I., & Davidoff, J. (2000). Color categories are not universal: Replications & new evidence from a Stone-age culture. *Journal of Experimental Psychology: General*, *129*, 369-398.
- Scheffrin, B. E., & Werner, J. S. (1990). Loci of spectral unique hues throughout the life span. *Journal of the Optical Society of America A*, *7*, 305-311.
- Tye, M. (2000). *Consciousness, color, and content*. Cambridge, MA: MIT Press.
- Wade, T., Jongman, A., & Sereno, J. (2007). Effects of acoustic variability in the perceptual learning of non-native-accented speech sounds. *Phonetica*, *64*, 122-144.
- Watkins, M. (2002). *Rediscovering colors: A study in pollyana realism*. Kluwer.
- Watkins, M. (2005) A posteriori primitivism. *Philosophical Studies*, *150*, 123-137.
- Webster, M. A., & Kay, P. (2005). Variations in Color Naming Within and Across Populations. *Behavioral and Brain Sciences*, *28*, 512-513.
- Westphal, J. (1987). *Colour: A philosophical introduction*. Basil Blackwell.
- Westphal, J. (2005). Conflicting Appearances, Necessity and the Irreducibility of Propositions about Colours. *Proceedings of the Aristotelian Society*, *105*, 219-235.
- Wuerger, S. M., Atkinson, P. and Cropper, S. (2005) The cone inputs to the unique-hue mechanisms. *Vision Research*, *45*, 3210– 3223.
- Wright, W. D., & Pitt, F. H. G. (1934). Hue discrimination on normal colour-vision. *Proceedings of the Physical Society (London)* *46*, 459–454.

Figures

Figure 1. Examples of the worksheets for the present experiment



Figure 2. *Standard Deviations (SD) for the red and green matching and naming worksheets across participants. A significant difference appears between matching and naming tasks. The difference remains when green matching and naming tasks are compared but does not remain when red matching and naming tasks are compared. No significant difference appeared between the matching worksheets or between the naming worksheets.*

