12

Perceptual Indiscriminability and the Concept of a Color Shade

Leon Horsten

12.1 INDISCRIMINABILITY AND COLOR SHADES

We shall be concerned with visual indiscriminability, and more specifically on perceptual indiscriminability of colors. But, as usual, the discussion is intended to carry some more generality. It intends to contain some lessons for the philosophical theory of perceptual qualia in general.

According to the received view, perceptual indiscriminability is a nontransitive relation. Until about a decade ago, there was a nigh consensus in the philosophical literature that the received view is correct. Today, this consensus has dissolved. This has prompted me to make an attempt at reevaluating the question of the transitivity of perceptual indiscriminability.

In the first sections of this chapter, recent challenges to the thesis of the nontransitivity of perceptual indiscriminability will be critically examined. Such challenges can take different forms. Some of them concentrate on philosophical arguments for the nontransitivity thesis; some concentrate on alleged empirical evidence for the claim that indiscriminability is nontransitive.

Fara's challenge belongs to the first kind. She has criticized Wright's influential philosophical argument that was intended to establish that indiscriminability is nontransitive. We shall investigate to what extent Wright's argumentation can be upheld in the face of Fara's reply.

Raffman has emphasized that the question of the transitivity of the relation of perceptual indiscriminability is ultimately empirical in nature. And she believes that the alleged empirical evidence in favor of the nontransitivity thesis is far from

I am indebted to the participants in the Arché Vagueness Conference (St. Andrews, 2007) for insightful comments and helpful suggestions. My thanks go especially to Paul Egré (who commented on this chapter), Diana Raffman, Patrick Greenough, Crispin Wright, Nathan Salmon, and Anthony Everett. Section 12.3 and section 12.6 are based on earlier joint work with Rafael De Clercq, to whom I am also grateful for discussions specifically relating to the problems with which the present chapter is concerned. It should not be assumed that Rafael De Clercq agrees with the overall position that is articulated and defended in this chapter.

conclusive. She argues that a contextualist interpretation of the data can be given which leaves ample room for the thesis that perceptual indiscriminability is a transitive relation after all. We shall see that whereas her arguments are not conclusive, there is much to be learned from her considerations.

If perceptual indiscriminability is transitive, then at least a criterion of identity for color shades is readily obtained by Frege's familiar method of abstraction. But if perceptual indiscriminability is a nontransitive relation after all, then we cannot rely on the method of abstraction to yield an identity criterion for color shades—at least not in any straightforward manner. The question then becomes acute whether the concept of a color shade is coherent in the first place, and, if so, which form it could take.

It shall be argued that in this situation, it must be conceded that the concept of a color shade is to some extent theoretical in nature. But since the concept of a color shade is invoked in the first place to make sense of our color discrimination judgements, it must respect these as much as possible. We shall therefore inquire how a concept of a color shade can be obtained which is, in the face of nontransitivity of indiscriminability, maximally faithful to our color indiscriminability judgements.

But faithfulness to indiscriminability judgements is not the only requirement. The resulting concept of a color shade should also be well in agreement with the way in which the concept of a color shade is used in natural language. So we must also ask ourselves whether this requirement, too, is satisfied.

In what follows, I have in mind the notion of indiscriminability in color for a given agent which is kept fixed throughout the discussion. This agent is assumed to have capacities for discriminating between colors of objects that are or fairly average for a human being. Two objects are said to be indiscriminable in color by the agent if she cannot discriminate between them with respect to color on the basis of a direct color comparison. Thus the relation of discriminability with which we are operating coincides with what Goodman called the *matching relation.*

Indiscriminability with respect to color does not exclude that there are other ways in which the agent can distinguish the colors of objects. Suppose, for instance, that there is a nontransitive triad of objects x, y, and z. On the basis of a direct color comparison, x cannot be discriminated from y, y cannot be discriminated from z, but x can be discriminated from z. Then our agent can *reason* that since x is discriminable from z whereas y is not, x and y must have a slightly different color. Still, in our sense of the word, x and y remain indiscriminable. In other words, the concept of indiscriminability as it will function in our discussion is assumed to be *phenomenal.*

Indiscriminability will also be taken to be a *judgemental* relation. For two objects to be indiscriminable in color for our agent, she has to be able to judge them to be indiscriminable when she visually compares them. Likewise, for these objects to be discriminable for her, she has to be able to judge them to be discriminable on the basis of a visual comparison.

12.2 WRIGHT'S NONTRANSITIVITY ARGUMENT

Perceptual indiscriminability is nontransitive if it is possible to have three items x, y, z such that x is perceptually indiscriminable from y, y is perceptually indiscriminable from z, and x is perceptually discriminable from z.

This entails that precisely one of the following two theses must hold:

The Transitivity Thesis Perceptual indiscriminability is a transitive relation.

The Nontransitivity Thesis Perceptual indiscriminability is a nontransitive relation.

Until fairly recently, it was common philosophical practice to accept the nontransitivity thesis (and thus to reject the transitivity thesis) on the strength that we can imagine a process of gradual change in which a series of unnoticeably small changes finally add up to a noticeable change (in respect of a given quality).

Several authors have sought to show that perceptual indiscriminability is nontransitive in this way. Here we discuss an influential philosophical argument by Crispin Wright (Wright 1975, 345–7).

Wright presents his proof as a reductio showing that the nontransitivity thesis follows from the possibility of phenomenal continua. The argument can be paraphrased as follows. Suppose that indiscriminability is transitive. Then consider a process of change in respect of some observable property (think of it as a determinable such as color, position or pitch). The process is composed of stages between which there is no seemingly abrupt transition, and is non-recurrent in that for two distinct stages x and y, with x preceding y, there is no later stage z such that z is more like x (in respect of the observable property) than y is. Take any two stages D_i and D_j such that D_j is discriminable from D_i and yet close enough to it to guarantee that all stages lying in between are either indiscriminable from D_i or indiscriminable from D_j. In other words, the intermediate stages will appear to have the same determinate of the determinable as either of the two surrounding stages (e.g. the same shade of color). They cannot be indiscriminable from both D_i and D_j since being-indiscriminable-from is supposed to be a transitive relation. As a result, the region between D_i and D_j will divide into two adjacent sub-regions, one consisting of stages indiscriminable from D_i, the other consisting of stages indiscriminable from D_j. Since indiscriminability is supposed to be transitive and since D_i is discriminable from D_j any stage belonging to the first sub-region will likewise be discriminable from any stage belonging the second sub-region. However, if this is true, then, contrary to what we have been assuming, a seemingly abrupt change must occur between D_i and D_j.

In recent years, the nontransitivity thesis has been called into doubt. These challenges tend not to consist of arguments that directly support the thesis of the transitivity of indiscriminability. Rather, they consist in the first place of attempts to undermine the alleged evidence for the nontransitivity thesis. In the next section, I shall scrutinize Fara's critique of Wright's philosophical argument. In section 12.4, we shall turn to Raffman's attempt to deflect the charge that we have conclusive empirical support for the nontransitivity thesis.

12.3 UNNOTICED APPARENT CHANGES

According to Fara, Wright's proof relies on two assumptions: the possibility of phenomenal continua and the finiteness of human discriminatory powers (Fara 2001, 931). The first assumption is needed to deny the existence of a seemingly abrupt transition from one stage to another. The second assumption allows for perceptually indiscriminable stages in the process. According to Fara these two assumptions 'are, taken individually, not implausible [but] they are in so much tension with each other that it is utterly unreasonable to accept them jointly when neither has anything remotely like adequate support' (Fara 2001, 931).

Closer inspection of Wright's argument reveals that the first assumption, concerning the possibility of phenomenal continua, is not necessary for his argument. In fact, it is not hard to see that Wright's argument, reduced to its essentials, is exceedingly simple. Aside from a plausible physical assumption, the finiteness of our powers of discrimination is all that is needed for the argument. Let there be given an observable physical quantity Q. Suppose that the value of this quantity can be expressed as a real number. (Thus, the quantity Q can be regarded as a determinable with specific values as determinates). And adopt the physical continuity assumption that the value of Q varies according to some smooth continuous function (in the mathematical sense of the word!) through time. Let r_i refer to the value of quantity Q at time i. Now we assume finite discriminability in the sense that (i) there are r_a, r_b such that the subject can discriminate between them and (ii) there is a $d \in \mathbb{R}$ such that if $r_i - r_j < d$, then a given person is unable to perceptually discriminate between Q at i and Q at j. Now consider a finite chain $r_a = r_0, r_1, \ldots, r_n = r_b$, such that for each r_i in the chain, $r_{i+1} - r_i < d$. The foregoing assumptions entail that such a chain exists. Moreover, finite discriminability entails that a subject perceiving the chain will not notice 'an abrupt change', which means that the change in Q will be perceived as continuous in Wright's (phenomenal) sense. Elementary mathematical considerations show immediately that this chain must contain a violation of transitivity of indiscriminability. After all, since each element in the chain is indiscriminable from the next with respect to Q, transitivity would imply that the first element is indiscriminable from the last. However, by assumption r_a is discriminable from r_b with respect to Q.

So in the final analysis, all rests on the assumption of finite discriminability.[1] Fara does not find this assumption evident. In her discussion of the phenomenon of 'slow motion', she writes (Fara 2001, 928):

> ... we have two competing explanations of what is going on when the hour-hand of a clock looks to have moved over some long [time] interval, but also seems to have looked still during every sufficiently short sub-interval. The first explanation is that when we judge the hour-hand to look still, say for every twenty-second period, it does in fact look to be in the same position at the end of each period as at the start. The alternative explanation is that when we judge the hour-hand to look still, although there is at least one twenty-second period for which it does

[1] One may wonder at this point why Fara believes that the phenomenal continuity assumption is in tension with the finite discriminability assumption. A diagnosis is offered in De Clercq and Horsten (2004, section II).

not look in the same position at the end as at the start, we do not notice this. *Noticing* the change in an apparent position requires not only that there be an apparent change, but also that we believe there to be one.

[emphasis in the original]

In other words, according to one explanation of what happens when the hour-hand of a clock changes unnoticeably, there is no apparent change because there does not appear to be a change: at least at a conscious level, things look exactly the same before and after the change. This explanation seems plausible enough. However, Fara's sympathy lies with the other explanation: the *apparent* position of the hour-hand of a clock—the position it appears to have—changes constantly, i.e. even within time intervals that are so short that we are unable to tell ('notice') whether there has been a change.

Neuro-psychological research has born out that we routinely respond to visual stimuli of which we have no conscious awareness.[2] In the situation that Fara describes, too, it may be that we are behaviorally able to respond in some way to the change of the position of the hour-hand in a twenty seconds period even though we are not consciously aware of a position change. So, in this sense, it might be said that the position change may not completely escape our attention even if it escapes our conscious attention.

Elsewhere in her paper, Fara argues that accepting the nontransitivity of 'looking the same as' does insufficient justice to the phenomenal character of looks (Fara 2001, 932). After all, if 'looking the same as' is transitive, then looks can simply be taken to be equivalence classes of the relation; and if 'looking the same as' is nontransitive, then one must either maintain that there are things which look the same (in some respect) but nevertheless do not have the same look, or that there are things which look different but have the same look, or both.

However, if this objection is justified, then it might be argued that Fara stands guilty of a similar charge. Fara's concept of indiscriminability is in the end a subjudgemental and subconscious relation; it can fail to obtain in a comparison test without the agent being conscious that it fails to obtain. By separating the notion of apparent change from the notion of noticed change, Fara may then be said to deprive the notion of apparent change of its phenomenal nature. At best, the notion of apparent change becomes one that is determined by our (partially unconscious) response-behavior rather than by the contents of our explicit consciousness. Nothing can prevent Fara from abstracting a notion of color shade from such a behavioral relation. But *this* notion of color shade will fail to qualify as fully phenomenal.

12.4 APPEARING IN CONTEXT

Raffman has articulated a contextualist position on the basis of which she criticizes Wright's argument for the nontransitivity of indiscriminability (Raffman 2000). She

[2] See Weiskrantz (1986).

is concerned with the relation of perceptual indiscriminability of objects, in respect of a perceptual property (Raffman 2000, 155). This should be interpreted loosely. It is intended to include situations where someone is asked to discriminate visual images or parts of visual images with respect to color. Raffman's arguments are not intended to establish but only to make room for the hypothesis that perceptual indiscriminability is a transitive relation (Raffman 2000, 154–5). She says of her her argumentation that much of it is 'speculative [...] and in need of empirical test' (Raffman 2000, 155).

The contextualist hypothesis can roughly be formulated in the following way:

The Contextualist Hypothesis When objects x and y are mentally compared with respect to perceptual property P with the aim of reaching a discrimination judgement, contextual factors can and typically do influence the agent's perception of x and of y.

This hypothesis seems eminently plausible, and is supported by a large body of empirical data, which go under the rubric of 'contrast effects'.

The contextualist hypothesis is used by Raffman to undermine arguments for the nontransitivity of perceptual indiscriminability such as Wright's as logical fallacies. Suppose first x and y are compared by someone with respect to color; then y and z are compared; and then x and z are compared. And suppose that the person's discrimination judgements are no, no and yes, respectively, thereby forming a prima facie case for nontransitivity of the relation of perceptual indiscriminability. Then the contextualist will point out that, to a first approximation, in the first comparison, x constitutes the context of y, and in the second comparison, z constitutes the context of y. If the contextual hypothesis is correct, this causes the person's perception of y in the first comparison to differ from the person's perception of y in the second comparison. So one is not allowed to 'carry over the middle term' in the argument for nontransitivity (Raffman 2000, 161).[3]

In a contextualist framework, the transitivity thesis takes the following form. *If objects are compared in view of a perceptual property P in such a way that there is no contextual disturbance of perception, indiscriminability is a transitive relation.* According to Raffman, this hypothesis is not excluded by what is presently known. And we have seen that Raffman expresses the hope that empirical evidence might be brought to bear on it.

It is clear that even in experiments set up to determine whether one participant can discriminate between the color of stimuli, her answers will display a statistical distribution. Hardin emphasized the importance of this phenomenon for the philosophical discussion about the (non-)transitivity of indiscriminability.[4] It will happen that even for a single pair of color patches, she will sometimes answer that she can discriminate them, whereas at other times she will judge them to be indiscriminable in color. So it must be conceded at the outset that our assumption that an agent always makes

[3] Schroer (2002) develops a similar, but less fully articulated, line of reasoning.
[4] See Hardin (1988).

the same discriminability judgement makes us guilty of grossly oversimplifying the situation. We shall have to come back to this: in section 12.6, the implications of retracting this assumption will be (all too briefly) discussed.

Raffman considers possible counter-evidence to the contextualist version of the transitivity thesis. She considers circles, presented against a uniform background color, and divided into 3, 4, 5 or more equally large sectors, where adjacent sectors receive very similar colors. In particular, she considers such a circle divided into 3 sectors (Raffman 2000, 163ff). Participants are asked to compare the three sectors two by two with respect to color. The advantage of such a set-up is that the 'context' of y, for example, when compared to x is the same as when compared to z, for in each act of comparison the whole circle remains firmly in view of the participant.[5]

Raffman contends that for circles divided in three sectors, solid empirical evidence pointing in the direction of nontransitivity has not been forthcoming. It appears to be impossible to construct a circle divided into three sectors x, y and z of slightly different colors in such a way that the agent systematically judges x and y as well as y and z indiscriminable, but also systematically judges x and z discriminable.

Here the statistical nature of the distribution of discriminability answers may play a substantial role. The statistical distribution is presumably just not sufficiently sharply peaked to come to a reliable conclusion in such a triad situation. This point is not without importance. It entails that reports such as the following just cannot be accepted without further ado:

> Suppose I focus on just patches 1–3 and claim sincerely to attend simultaneously to the color-appearances of all three patches. ([. . .] I can attest from my own case—as you can probably attest from yours—that we can [focus on three patches at a time].) At a given instant, I claim, patches 1 and 2 looks the same to me, as do patches 2 and 3, but patch 1 and 3 look different.
>
> (Mills 2002, 395)

At this point, it is important that a background condition for success is firmly kept in mind. This condition was highlighted by J. A. Burgess (Burgess 1990, 209), but it has not received much attention in the literature. For the experiment to succeed, x, y and z have to be perceived as *uniformly* colored. This forces us to design the experiment carefully. A first question is whether the three sectors are separated from each others by lines of a uniform color which clearly differ from the colors of x, y and z. And a second question is whether the three sectors really touch each other or, alternatively, are 'pulled apart' some distance against a uniform background. If the answer to both these questions is no, then it can legitimately be questioned whether the background condition is satisfied. When one tries to construct a pie consisting of three contiguous sectors that forms a counterexample to the transitivity hypothesis, the participants in the experiment have some tendency to see it as a phenomenal continuum. To the extent that that is the case, the participants will be hesitant to affirm that each of the three sectors are uniformly colored. This problem can be mitigated by pulling the sectors apart against a uniform background color. Schroer notes that the sectors should

[5] Raffman admits that she is 'helping [her]self to the notion of a visual context more or less unexamined' (Raffman 2000, 159), but let us grant her that.

not be pulled apart far (Schroer 2002, 265–7). If the sectors occupy parts of the visual field of the participant that are not very close to each other, then the participant cannot compare them directly. So on account of physiological limitations she will have to shift her focal visual attention from one sector to the other. And this entails that in her act of comparison, she has to rely on her memory, which is much less precise than occurrent appearances are.

Let us assume in the sequel that the experiments are set up in such a way that the participants perceive the monochromatic colors as uniformly colored. Then one may ask: *how could empirical evidence bear out that the nontransitivity hypothesis is false?* In any actual mental comparison of two objects x and y with respect to P, representations of x and y are necessarily present in a special way. When a person compares x and y, she is focused on x and y in a manner that she is not focused on z. Suppose, perhaps contrary to the facts, that it were possible to construct a three-sector circle in such a way that it would reliably generate prima facie evidence for the nontransitivity thesis. Then the contextualist is free to conjecture that in experiments of this sort, concentrating on x and y (in a comparison of x and y) versus concentrating on y and z (in a comparison of y and z) makes a difference in the perception of y. Thus the prima facie evidence for nontransitivity would on closer inspection disqualify as genuine evidence for nontransitivity of the indiscriminability relation.

Just this move is executed by Raffman when she considers variants of this experiment with circles divided in at least five sections (Raffman 2000, section III). A knee-jerk reaction to the difficulty of constructing a three-sector circle which generates prima facie evidence for the nontransitivity hypothesis is to divide the circle into more sectors—as many sectors as it takes! One might, for instance, divide the circle into five sectors such that sector 1 is very close in color to sector 2, which is in turn very close in color to sector 3, and so on, until one arrives at sector 5, which is not very close in color to its neighboring sector 1. Raffman thinks that on closer inspection, even prima facie evidence of this kind does not refute the transitivity hypothesis (Raffman 2000, 169):

> . . . it may be that in any series whose adjacent members are indiscriminable (would be judged the same in a same / difference comparison), [. . .] at any given time at least one member looks different in its two hypothetical comparisons with adjacent items, even in the case the entire series is viewed simultaneously.

But there is no reason why a contextualist could not take this line even concerning circles divided into three sectors.

In fact, this kind of move is always open to the contextualist. It even provides a scheme for reacting to diachronic comparison experiments. Suppose we modify the experiment so that at a given time, the whole visual field is occupied by exactly one monochromatic paint chip, and the agent is asked whether she can discriminate the color presently in her field of view from the color she saw two seconds ago. Even though the color sample is now not presented against a visual background, it is presented against a 'visual memory context'. Presumably visual memory contexts can influence the content of a color perception in a way similar to the way in which visual contexts influence colors perception of objects.

Raffman's attempt to give *hard* empirical content to the contextualist defense against the prima facie five-sector counterevidence against the transitivity hypothesis is unsuccessful. She suggests the following diachronic comparison experiment (Raffman 2000, 167):

> Suppose you are given a new task, involving just the pairs #3 / #4 and #4 / #5: you are to attend first to #3 and #4, then to #4 and #5, this time with the aim of judging whether #4 looks the same or different in the two pairings—in the two attendings, one might say. In other words, rather than making two comparisons, of #3 to #4 and #4 to #5 as in the serial scenario, your new task is to perform a single, cross-contextual comparison. As I will put it, your task is to compare #4 viewed in an act of focused attention to #3 and #4, with #4 viewed in an act of focused attention to #4 and #5. First you attend to the pair #3 / #4, then you shift your attention to #4 / #5, keeping #4 in view continuously, and you make a same / difference judgement of #4. Isn't it possible that #4 should look different?

It is not clear that Raffman's diachronic comparison experiment is well-conceived in the first place. The memory-image of #4 may not be sharp enough for an average participant to compare with the occurrent experience of #4 (Wright 1975, 336), (Schroer 2002, 267). But even if memory-images are sufficiently crisp for the task, there is no reason why the contextualist should feel cramped by consistent responses that in the two comparison acts #4 looks the same, if such responses were forthcoming. She is free to conjecture that after the second comparison task, the memory of #4 in the first comparison task is influenced by the subsequent focus on #5. This is after all, for all we know, possible. And if it is the case, then 'no difference'-responses to the diachronic task have little import for the transitivity hypothesis.

At this point it becomes evident that conclusively refuting the contextualist defense of the transitivity hypothesis on the basis of empirical evidence, is impossible. This is just a consequence of the fact that was noted earlier, namely that in an explicit color comparison between two items x and y, these items are present in a special manner; they are present in a way that is unlike the way in which features of the context are present in consciousness. This always leave the contextualist free to deny that it is permissible to 'carry over the middle term'.

One might say that the situation with the transitivity hypothesis is no different from that with scientific hypotheses in general. One may hold as fast to them as one likes. But there is a point at which rescue missions begin to look somewhat ad hoc. It seems that if empirical circumstances would force Raffman to ascribe contextual effects to focusing acts, i.e. when empirical evidence would force her to question whether in the five-sector circle experiment #4 looks the same when compared to #3 as when compared to #5, this point has been reached.

For all that has been said, the transitivity hypothesis may still be correct. It may be correct even in the face of the sort of hypothetical prima facie empirical counterevidence that we have discussed. Still, as Raffman has emphasized, it remains at least in part an empirical matter. For it may well be that the empirical data look much more favorably upon the contextualist theory than in the pessimistic scenarios that we have been considering.

12.5 COLOR SHADES: PHENOMENAL OR THEORETICAL?

Raffman was right to insist that it is at least to some extent an *empirical* question whether perceptual indiscriminability is a transitive notion. But we have seen that it is not as easy to test the transitivity hypothesis as it appears at first sight. If we want to empirically determine color shades, how should we proceed? What are the possible outcomes of such an investigation? And what is their significance?

As mentioned before, color shades are supposed to make sense of our sense experience. But the discussion of Fara's position entails that we ought to be more precise. We must decide whether we want color shades to make sense only of what we are consciously aware of, or whether it should also make sense of what we are unconsciously or behaviorally aware of. This is purely a matter of decision. But it appears best in line with the historical development of this enterprise if we restrict ourselves to *conscious* sense experience, which the agent is able to make explicit in discriminability judgements.

From a contextualist point of view, the situation then looks as follows. Contextual factors typically influence indiscriminability judgements. So when indiscriminability judgements are combined in arguments, the context must be kept as uniform as possible.

Raffman advises us to keep the context constant (as much as possible) throughout the indiscriminability experiment by making sure that all the stimuli remain clearly within the visual field of the participant throughout the successive comparison tasks. For concreteness, let us suppose that the participants are presented with sectors of a circle which are slightly differently colored. To ensure that the sectors are perceived as uniformly colored, they are pulled apart a bit. And the sectors are presented against a uniformly colored background. To conclude, let us suppose that we are working with a large but finite number of color stimuli ('paint chips').

In this way, we obtain a total indiscriminability graph in a fairly straightforward way. But on closer inspection even this much is far from clear, for the following reason. Surely there will be a maximum on the number of sectors that can be presented in one experiment (Raffman 2000, 170). And it seems likely that this number will be less than the number of color stimuli that are of a slightly different color from which we want to abstract the color shades. Hence not all distinct stimuli can be present in one run of the experiment. Instead, many runs will have to be done with many sextuples (say) of colored sectors. Even if the transitivity hypothesis can explain each run of the experiment individually, the question remains how the results of the different runs of the experiment should be patched together. It will not be sufficient to assume without further ado that if one stimulus is present as a sector of the circle in different runs of the experiment, the appearance in one run of it is the same as the appearance of it in another run. After all, the same stimulus was offered in different contexts, and the contextualist hypothesis will predict that contextual differences typically change the appearance of the stimulus.

In sum, it is not at all clear how the results of the sequence of experiments can be taken to give rise to one global graph on the class of stimuli. It cannot be excluded

that the simultaneous focal presence of one color may influence the appearance of the other—this is unavoidable. And this would mean that the contextualist may still cry foul: it may be that it is impermissible to combine the individual discrimination judgements into a discriminability graph. But from an *operational* point of view, this appears the best we can do.

We shall not pursue this problem further. Instead we shall for the sake of argument assume that it can be dealt with so that in the end, an indiscriminability graph is obtained. So let us suppose that an experiment is set up and carried out in this way. Then there are in general two possible outcomes. We shall examine them in turn.

First, it may turn out that the empirical data can be explained very well by the transitivity hypothesis in the framework of a version of contextualism which does not appear ad hoc. Then, it seems, color shades can be abstracted in the familiar way from explicit indiscriminability judgements that are corrected as much as possible for contextual effects. Abstraction does not yet tell us what the nature of color shades consist in. But it does give us a criterion of identity for color shades:

The Innocent Criterion The color shade of paint chip x is numerically the same as the color shade of paint chip y if and only if x and y are perceptually indiscriminable with respect to color.

In these circumstances, and given the supposition that the transitivity hypothesis can be upheld in the light of the data, there will be sharp cutoff points between color shades.

A marked advantage of the innocent criterion is that it ensures that color shades harmonize perfectly with the indiscriminability relation. Color shades are admitted into our ontology in the first place in order to make sense of our visual experiences. So if we can best make sense of the discrimination experiment on the basis of a global transitive graph on the stimuli, then we are in the fortunate situation that the abstracted collection of color shades will be completely faithful to our (corrected) discrimination judgements.

One may wonder whether a theoretical element has nevertheless entered into the contextualist defense of the transitivity hypothesis. After all, on this account there are no 'pure' indiscriminability judgements; it is still the case that all we have are objects that are indiscriminable *in a context.* But this appears unproblematic. It just seems hard to deny that context is a parameter that plays a role in the discriminability of objects with respect to color—even though until recently philosophers have hardly recognized it.

A second possible outcome is that empirical evidence makes it hard even for a contextualist to maintain the transitivity hypothesis without resorting to seemingly ad hoc maneuvers. Then it is impossible to abstract color shades from indiscriminability judgements in the familiar way. Either some patches that cannot be distinguished in color will be judged to have different color shades, or some patches that are discriminable in color will be judged to have the same color shade, or both.

This is what Fara finds objectionable (Fara 2001, 909). In his earlier writings, Wright also deemed this unacceptable (Wright 1975, 352).[6] Color shades ('looks') were supposed to make sense of our sense experience first and foremost. If the concept of a color shade does not *fully* respect our indiscriminability judgements, then it is not a phenomenal concept. Wright went so far as to say that in such a situation we have a transcendental concept (Wright 1975, 357).

Yet what are our options, in the situation under consideration? We have access to color shades only via color comparisons. In this sense, indiscriminability judgements are prior to the color shades themselves.

First, one can deny that color shades (or sense data in general) exist. This is not an uncommon stance. Armstrong takes this position (Armstrong 1961). Fara does not explore this option; she merely mentions in passing the possibility to give up on 'looks' altogether (Fara 2001, 916, fn 13). Raffman does not embrace this position either, although she is careful not to rely anywhere in her argumentation on the assumption that color shades exist (Raffman 2000, 160–1). Perhaps from a contextualist point of view denying the existence of shades *should* appear as an attractive option. For would it not be natural to say, from a contextualist perspective, that color shades (or 'looks') are essentially relative to a context? This would mean that it makes sense to say that *in a given visual context* in which two objects x and y are present, the color shade of x is the same as the color shade of y. But it does not make sense to ask whether the color shade of an object z presented in one visual context is the same as the color shade of u presented in another visual context. (As far as I am aware, the prospects of this position have not yet been fully explored in the literature.)

A second option consists in conceding that the concept of a color shade is to some extent a *theoretical* concept. This line was taken by J. A. Burgess (Burgess 1990, 218–19):[7]

> ... theories that respect the phenomenology of perception are just that: theories. This means that they might not only be required to postulate properties that are (in some sense) not presented in experience in order to do justice to some facts of experience [...]; it also means that they might need to discard some apparent data as illusory.

The concept of a color shade is not clearly a philosophers' concept, but it is also not clearly a pre-theoretical concept that is as pure as driven snow. It appears to be a *low level* theoretical concept. This should not be taken to contradict the fact that it is intimately tied to experience. Indeed, it can and should be upheld that its aim remains first and foremost to make sense of our indiscriminability judgements.

Whether it is fruitful to develop a theoretical concept of color shade depends on what use it can be put to. In ordinary language, color shades play an important communicative and judgemental role. We routinely make judgements about color shades, for instance when we say 'this shade is the same as that shade' while pointing at parts of the surfaces of two objects. If we develop a theoretical concept of color shade that respects indiscriminability judgements as much as possible, we may hope to validate

[6] He has since then sought to qualify his position on this issue. See below, section 12.7.

[7] Linsky defends a similar position (Linsky 1984).

many of the judgements of this kind. In communicating our visual experiences to others, we use the concept of a color shade. We can construct a theoretical model of what such color shades could be.

In the quotation above, Burgess appears to take a realist view of color shades. Such a view is certainly tenable, but it is not forced upon us. One can also take color shades to be useful theoretical fictions. Which of these two stances is more appropriate is, as in all forms that the realism debate takes, difficult to adjudicate.

12.6 APPROXIMATING BY EQUIVALENCE RELATIONS

Let us proceed on the assumption that we want to recognize the existence of color shades in our ontology. More specifically, let us assume that we want to find a criterion of identity for color shades in the face of apparent nontransitivity of the indiscriminability relation. Then we are faced with two requirements that are pulling in opposite directions. On the one hand, we want color shades to respect the indiscriminability relation as much as possible. On the other hand, we need an equivalence relation to base our abstraction on.

In this predicament, we should try to strike a good compromise. Being phenomenal or observational is a matter of *degree* in this context, and should not be taken to be a matter of principle. We should look for an equivalence relation that is somehow as close as possible to the nontransitive indiscriminability relation. From such an equivalence relation, a criterion of identity for color shades can be obtained. The advocates of color shades in the light of nontransitivity have been aware of this task. But some of them have underestimated the subtlety of the problem.

A first proposal was made by Nelson Goodman (Goodman 1966, ch. IX). He thinks that the innocent criterion should be replaced by (roughly) the following criterion:[8]

Goodman's Criterion The color shade of paint chip x is numerically identical with the color shade of paint chip y if and only if for every paint chip z, x is perceptually indiscriminable with respect to color from z if and only if y is perceptually indiscriminable with respect to color from z.

Even if indiscriminability is nontransitive, the right-hand-side of Goodman's identity criterion is an equivalence relation. If Goodman's criterion is correct then discriminability of paint chips x and y is a sufficient, but not a necessary condition for the paint chips to be of a different color shade. So the effect of Goodman's criterion is to loosen the tie of color shades with indiscriminability in exchange for having an equivalence relation.

Goodman's criterion will result in what Wright calls *Goodman Shades*. Wright has observed that under fairly general circumstances, Goodman Shades will be very finely grained (Wright 1975, 354). Indeed, in many circumstances, each pair of paint chips

[8] Burgess (1990) also defends a version of Goodman's criterion. Linsky (1984) shows sympathy for it, but he has reservations and does not quite endorse it. See Linsky (1984, 379).

will be judged by the criterion to be of different color shades. This restricts the usefulness of Goodman's criterion. For surely the color shades that we attempt to communicate and quantify over in ordinary language are not so fine grained.

Williamson has proposed methods for finding equivalence relations which are much closer to the nontransitive indiscriminability relation (Williamson 1986, 1990). We are given an indiscriminability relation G which is reflexive and symmetrical but nontransitive. Our task is to approximate G by an equivalence relation. We can approximate G from above by taking the transitive closure of G. Or we can approximate G from below by considering maximal equivalence relations $G^{-} \subset G$. Either way, we obtain equivalence relations that are in a qualitative sense close to G: we call these *qualitatively* best equivalence-approximations.

In the present context, the approximation from above does not make sense. For if the indiscriminability graph is connected (as in most cases it will be) then the transitive closure is the total graph on the underlying collection of paint chips. And that would mean that we have only one color shade. In other words, compared to Goodman's criterion we would have landed in the other extreme.

Qualitatively best equivalence-approximations from below seem more promising. They are of the same kind as Goodman Shades: they ensure that discriminability is a sufficient but not in general a necessary condition for having different color shades. The difference is that qualitatively best equivalence-approximations are more coarsely grained than Goodman Shades. Consequently, qualitatively best equivalence-approximations from below are in general *more* faithful to the indiscriminability relation than Goodman Shades.

One drawback of qualitatively best equivalence-approximations from below is that they are in general not unique. For all nontransitive indiscriminability graphs, there exist more than one best equivalence-approximations from below. This problem can be mitigated by choosing the most coarse grained best equivalence-approximation from below. But even imposing this as an extra requirement does not always ensure that there is a unique best approximation. This entails that a conventional element is inherent in the proposal of choosing a qualitatively best equivalence-approximation as the basis for an identity criterion for color shades. Sometimes all the best equivalence-approximations are equivalent to each other up to a simple transformation. In such cases, the conventional element does not play a deep role in the theory of color shades. But there are indiscriminability graphs for which the equivalence-approximations are not equivalent up to a simple scaling factor. In such situations, the conventional component plays a deeper role.

Williamson's methods can be somewhat improved upon by taking a quantitative point of view (De Clercq and Horsten 2005).[9] Our task again is to approximate the indiscriminability graph as closely as possible by a transitive graph. Taking a quantitative view of the matter, we consider those transitive graphs that result from G by cutting and/or pasting a minimum number of edges from/in G. Equivalence

[9] The method that we are about to discuss can be fairly straightforwardly extended to infinite domains. See again De Clercq and Horsten (2005).

approximations in this sense are *quantitative* approximations. It is easy to show that often the quantitatively closest approximation to G results from cutting *and* pasting edges. In other words, often a closest equivalence approximation is one that partially overlaps with G. It is also clear from this that quantitatively best approximations that only cut edges are qualitatively best approximations from below (in Williamson's sense).

Nontransitive indiscriminability graphs are more likely to have a unique quantitatively best equivalence-approximation from below than to have a unique qualitatively best equivalence-approximation from below. But uniqueness can still not be guaranteed in all circumstances. So the conventional element will play a smaller role, but it cannot be guaranteed that it plays no role whatsoever. But in general, quantitatively best equivalence-approximation will be more faithful to the indiscriminability relation than qualitatively best equivalence-approximations. So quantitatively best equivalence-approximations can be said to be more 'phenomenal' than Williamson's qualitative approximations.

This may be an appropriate place to illustrate the differences between the different strategies for abstracting color shades from nontransitive relations. As a basis for this illustration, we use the indiscriminability graph

$$G = \langle V, E\rangle = \langle\{e_1, e_2, e_3, e_4\}, \{\overline{e_1e_2}, \overline{e_2e_3}, \overline{e_3e_4}\}\rangle.$$

It scarcely needs to be mentioned that this is a highly simplified indiscriminability graph. In an actual experiment, we shall want a much larger domain of paint chips. And it is very unlikely that the associated indiscriminability graph will be serial in the way that G is.

If we follow Goodman's criterion, then there is a one-to-one correspondence between the paint chips and the shades, for the transitive graph E^g to which it gives rise is the totally unconnected graph. Since G is connected, the qualitatively best equivalence-approximation from above E^+ is the total graph. There are two qualitatively best equivalence-approximations E_1^-, E_2^- from below, of which the edges are, respectively:

$\overline{e_2e_3}$;

$\overline{e_1e_2}, \overline{e_3e_4}$.

We see that according to E_1^- there are three color shades ($\{e_1\}, \{e_2, e_3\}, \{e_4\}$) whereas according to E_2^- there are only two ($\{e_1, e_2\}, \{e_3, e_4\}$). So if we want the most course grained qualitative approximation from below, we must choose E_2^-. Also, we see that from a quantitative point of view there is only one best approximation: E_2^-. The latter makes only one 'mistake' against G, whereas E_1^- makes two mistakes and E^g and E^+ contain even more discrepancies with G.

One advantage of the innocent criterion and of Goodman's criterion is that they specify a *rule* for deciding, at least in principle, whether two paint chips are of the same color shade. On finite domains, best equivalence-approximations (both in the qualitative and in the quantitative sense) are of course decidable in principle. But

quantitatively best equivalence-relations even on a decidable infinite nontransitive graph need not be decidable at all. And even in the finite situations to which we confine our discussion, best equivalence-approximations do not give us a rule in any natural sense of the word for deciding equality of color shade for paint chips. Also, and related to this, most discriminability graphs are such that their best equivalence-approximations cannot be succinctly expressed in natural language.

For some indiscriminability graphs, the quantitatively best equivalence-approximation both cuts and pastes edges.[10] In other words, sometimes the quantitatively best equivalence-approximation is an equivalence relation that partially overlaps the original graph. When closeness to the agent's discriminability judgements is our primary consideration, it seems that one should in such a situation opt for such an overlapping approximation.

But this would entail that we count some patches that are discriminable in color as having the same shade of color. Surely many participants in the debate will find this objectionable. But is it really? Are we not sometimes willing to count certain objects as having the same shade even if we can discriminate them in color? Don't we sometimes say 'yes, I see a slight difference between them, but they are the same shade'? So if fitting common language usage is an important consideration, then it is not clear that our concept of color shade is not permitted to function in this way. Admittedly such considerations fall short of being decisive. Perhaps in such cases what is meant is that the relevant objects are of *roughly* the same shade. So we shall in the sequel refrain from challenging the thesis that it is part of the meaning of the concept of a color shade that discriminability entails difference in color shade.

As mentioned before, indiscriminability with respect to color is in practice not an all or nothing affair. Participants are typically more sure of some of their indiscriminability judgements than of others. This can be captured by assigning *weights* to the edges in the discriminability graph (De Clercq and Horsten 2005, 388). At the very least we should allow participants to opt out of *some* indiscriminability judgements. This would be captured by setting the set of weights equal to $\{0, 0.5, 1\}$, where an edge with weight 0.5 corresponds to a pair of paint chips for which the participant is unsure whether they are indiscriminable with respect to color.

Assigning weights to edges of course limits the scope of the non-uniqueness problem even further. But it will still not dissolve the problem completely. So what should we do when we have a weighted indiscriminability graph for which there exist two or more quantitatively best equivalence-approximations? As the reader will expect, familiar maneuvers present themselves. For instance, it is possible to maintain that two paint chips are only of the same (a different) color shade if the corresponding edge belongs to all (no) quantitatively best equivalence-approximation. Where some equivalence-approximations disagree, we could say that there is no matter of fact whether the relevant paint chips are of the same color shade.

Suppose that we have a global nontransitive indiscriminability graph of which there are more than one quantitatively best equivalence-approximations. Then how do we individuate color shades?

[10] A simple example is given in De Clercq and Horsten (2005, 377).

One option is to take the stance that color shades are low-level theoretical constructs seriously and simply abstract them from one particular quantitatively best equivalence-approximation—perhaps from one that is particularly easy to describe. Another option would be to argue that the 'belonging to' relation between color stimuli and color shades is to some extent *vague*. A supervaluation idea could be applied in the following way. Take any two color stimuli *a* and *b* from the global indiscriminability graph. Say that *a* and *b* (determinately) belong to the same color shade if *a* and *b* are indiscriminable according to all quantitatively best equivalence-approximations. *a* and *b* (determinately) do not belong to the same color shade if *a* and *b* are discriminable according to all quantitatively best equivalence-approximations. And in the remaining case, it is indeterminate whether *a* and *b* belong to the same color shade.

If one wants to remain as close as possible to the way in which the concept of a color concept functions in natural language, then the latter might look like the best option. But this position posits vagueness in the world, and arguments have been formulated which purport to show that the only vagueness that can exist is linguistic vagueness (Evans 1978). But in the present situation the vagueness involved—if there is any—cannot be easily shifted to the meaning of expressions. The reason is that natural languages do not contain simple names for all color shades. Perhaps natural languages do not even for every color shade contain a definite description that singles it out.[11] So if one is persuaded by Evans' arguments against worldly vagueness, then it is not easy to see in which way an appeal to vagueness can solve the non-uniqueness problem.

12.7 COLOR SHADES AND NATURAL LANGUAGE SEMANTICS

From the foregoing we may conclude that the question whether indiscriminability relation is a transitive relation, is at least to some extent an empirical one. But to some extent, it is also a theoretical one: the answer depends to some extent on theoretical interpretations of the evidence. Of course, it is often that way in empirical science.

If in the final analysis indiscriminability turns out to be transitive, then a concept of a color shade can be defined that is completely faithful to the indiscriminability judgements. If indiscriminability is not a transitive relation, then it is futile to look for a (non-contextual) concept of color shade that completely respects the relation of indiscriminability.

But in these circumstances it is still open to us to define a concept of color shade that is as faithful as possible to our indiscriminability judgements. This would result in a concept of color shades that falls short of being completely phenomenal, even though it will be as phenomenal as possible.

As adumbrated before, for Wright an important theoretical function of (in his view phenomenal) concepts such as that of a color shade consists in assisting us in

[11] Cf. the next section.

constructing a semantics for natural language. A concept of color shades abstracted from a best equivalence-approximation to our indiscriminability judgements could fit this bill.

Hardin doubts that color shades modeled on the basis of the indiscriminability relation are of much direct relevance for natural language semantics. The reason is that natural languages have names for only a few of them (Hardin 1988, 226–7). But considerations such as these should not be taken as decisive objections against Wright's project. Aside from color predicates, English contains the concept of a color shade. And this concept is not so easily dispensable. Many statements containing the term 'color shade' can be paraphrased by statements that contain the concept of indiscriminability instead. Thus, instead of saying *These two walls are painted in the same shade of orange*, we can equivalently say: *These two orange walls are indiscriminable with respect to color*. But the paraphrases of many statements containing the expression 'color shade' sound awkward. And for some statements containing the expression 'color shade' it is difficult to see how they can be paraphrased in terms of the concept of indiscriminability at all. An example of such a statement may be:

There are more shades of colors than we are able to produce samples of in our laboratory.

So it is not excluded that a concept of a color shade developed along the lines that have been outlined in this article may be of value for natural language semantics. Such a concept of color shades might count certain paint chips that are indiscriminable with respect to color as being of a different color shade. In this sense, the concept of a color shade coheres imperfectly with indiscriminability judgements. But this is only problematic if indiscriminability somehow gives the *rule* by which identity judgements of color shades are made. Already in his (Wright 1975), Wright thought that that this is precisely the assumption that should be given up anyway. Our semantic competence does not consist exhaustively in using expressions according to implicit rules. In his more recent publications, Wright has connected this with a softening of his stance on the observability of color predicates (Wright 1987, 246):

> 'How could the *appropriate* kind of sensitivity operate selectively among indiscriminabilia?' [...] it should now seem as if this question has rather disappeared. 'Looks red' ought certainly to qualify as observational on this count. But the kind of sensitivity to appearance which someone who understands 'looks red' [...] must have *does* operate selectively among items which, in respect to apparent color, cannot be told apart. [...] The suggestion that there is some kind of tension between such selectivity of response and its being the sort of response appropriate in the case of an observable predicate, depends on the thought that it cannot then be purely in response to appearance—either it is unprincipled or it is a principled response to *more* than the appearances. But 'unprincipled' here just means: not guided by rules correlating responses with appearances. So we should embrace the first alternative: such responses are indeed unprincipled, and no less appropriate or less purely 'to' appearances on that account.

These remarks are in perfect agreement with the theory that proposes identity criteria for color shades in terms of best equivalence-approximations to the indiscriminability relation. An identity criterion for color shades expressed in terms of equivalence-approximations points away from the picture of indiscriminability furnishing the

ingredients of a compact *rule* that language users are implicitly guided by when forming judgements of identity and difference of color shades. Indeed, we do not even have succinct expressions in natural language for most of these equivalence-approximations. But in case indiscriminability is nontransitive, best equivalence-approximations give us identity criteria for entities that cohere best, albeit not perfectly, with our indiscriminability judgements.

References

Armstrong, D. (1961), *A Materialist Theory of Mind,* Routledge and Kegan Paul.

Burgess, J. A. (1990), 'Phenomenal qualities and the nontransitivity of matching,' *Australasian Journal of Philosophy* 68, 206–20.

De Clercq, R. and Horsten, L. (2004), 'Perceptual indiscriminability: In defense of Wright's proof,' *Philosophical Quarterly* 54, 439–44.

—— (2005), 'Closer,' *Synthese* 146(2005), 371–93.

Evans, G. (1978), 'Can there be vague objects?' *Analysis* 38, 208.

Fara, D. G. (2001), 'Phenomenal continua and the sorites,' *Mind* 110, 905–35.

Goodman, N. (1966), *The Structure of Appearance,* 2nd edition, Bobbs-Merill, 1966.

Hardin, C. L. (1988), 'Phenomenal colors and sorites,' *Noûs* 22, 213–34.

Linsky, B. (1984), 'Phenomenal qualities and the identity of indiscernibles,' *Synthese* 59, 363–80.

Mills, E. (2002), 'Fallibility and the phenomenal sorites,' *Noûs* 36, 384–407.

Raffman, D. (2000), 'Is perceptual indiscriminability nontransitive?' *Philosophical Topics* 28, *Vagueness,* ed. Christopher Hill, 153–75.

Schroer, R. (2002), 'Matching sensible qualities: A skeleton in the closet for representationalism,' *Philosophical Studies* 107, 259–73.

Weiskrantz, L. (1986), *Blindsight: A Case Study and its Implications,* Oxford University Press.

Williamson, T. (1986), 'Criteria of identity and the axiom of choice,' *Journal of Philosophy* 83, 380–94.

—— (1990), *Identity and Discrimination,* Basil Blackwell.

Wright, C. (1975), 'On the coherence of vague predicates,' *Synthese* 30, 325–65.

—— (1987), 'Further reflections on the sorites paradox,' reprinted in R. Keefe and P. Smith (eds.), *Vagueness: A Reader,* MIT Press, 1997, 204–50.