

## Chapter 13

# The Character of Color Terms: A Materialist View<sup>†,\*</sup>

This paper investigates the character of predicates like:

(A)  $\lambda x(x \text{ is red})$  and

(B)  $\lambda xy(x \text{ appears red to } y)$

where  $x$  stands for a visible object and  $y$  for a perceiving subject (the reference to a time may be neglected).<sup>1</sup> I take here “character” in the sense of Kaplan (1977) as substantiated by Haas-Spohn (1995, 1997). The point of using Kaplan’s framework is simple, but of utmost importance: it provides a scheme for clearly separating epistemological and metaphysical issues, for specifying how the two domains are related, and for connecting them to questions concerning meaning where confusions are often only duplicated. All this is achieved by it better than by any alternative I know of.<sup>2</sup>

Therefore using this framework seems especially relevant to color talk where metaphysical and epistemological issues are more difficult to tell apart or may even

---

<sup>†</sup>This paper was originally published in: W. Künne, A. Newen, M. Anduschus (eds.), *Direct Reference, Indexicality, and Propositional Attitudes*, Stanford: CSLI Publications, 1997, pp. 351–379. It is reprinted here with kind permission of CSLI Publications.

\*I am very much indebted to Wolfgang Benkewitz, Martine Nida-Rümelin, and Ulrike Haas-Spohn; to a large extent the ideas of this paper have emerged in long lasting discussions with them. I am also indebted to Galen Straws on for various helpful remarks concerning style and content. The research was supported by grant No. Sp 279/4-2 of the Deutsche Forschungsgemeinschaft.

<sup>1</sup>It may seem excessively correct to use  $\lambda$ -abstraction here. But I do so only because it will help me later on to avoid awkward English.

<sup>2</sup>The credit equally goes to Stalnaker. In fact, the epistemological usefulness of the framework stands out much more clearly in his work; cf. in particular Stalnaker (1978, 1987). Despite their mutual claims of distinctness the work of Kaplan and that of Stalnaker are so closely related that I feel justified in speaking of one framework; for the precise nature of this relation cf. Haas-Spohn (1995, sects. 2.1 and 3.9).

seem to coincide.<sup>3</sup> And it should help in particular with my more specific goal, namely to clarify the epistemological and metaphysical status of such statements as:

- (1)  $x$  is red if and only if  $x$  would appear red to most English speaking people under normal conditions.
- (2)  $x$  appears red to  $y$  if and only if  $x$  (appropriately) causes  $y$  to be in a neural state of the kind  $N$ .
- (3)  $x$  is red if and only if the reflectance spectrum of the surface of  $x$  is of the kind  $R$ .

Indeed, I shall argue that (1) is analytic only in one reading and merely a priori in another reading. Moreover, I shall argue that after having set aside epistemological worries there is no good reason why one should not be able to be metaphysically conservative and to believe that (2) and (3) are necessarily, though a posteriori true for some  $N$  and some  $R$ , i.e., to sustain physicalism concerning colors and a type-type identity theory concerning color experiences; this is why I have characterized my views in the title as materialistic. Those who share these views anyway might still find it interesting to see how they fit into a broader theoretical framework; and those who oppose these views have to face the whole framework which appears to be successful on other scores. In any case, the framework should help both sides to more easily locate and clarify their divergence. Indeed, it was the main intention of this paper and the twin paper by Nida-Rümelin (1997) to exemplify this potential of clarification.

The paper starts with a presupposition and will then present six claims, the last three being the ones about (1)–(3) I have just indicated.

What I presuppose is simply the general adequacy and power of the framework of Kaplan and Stalnaker; I briefly recall its essentials as I use it here.<sup>4</sup> According to this framework, the right way of doing semantics for a given natural language is to recursively specify the character of all well-formed expressions of that language. The character of an expression is a function which assigns to each context the intension the expression has in that context, where the intension is a function from the set of index worlds or, more generally, from the set of indices into the set of categorically appropriate extensions. Thus, if a possible utterance of an expression is defined to be just that expression in a possible context, then the character of that expression may be represented by a two-dimensional scheme the rows of which show the intensions of all of its possible utterances.

There is wide agreement that intensions are suited for treating metaphysical modalities, in particular metaphysical necessity, but also counterfactuals, causation, and so forth. However, the two-dimensional scheme is also capable, though this is

---

<sup>3</sup>Almog (1981) is carried by the same enthusiasm concerning this framework. It is the only example I know of which explicitly takes this approach to analyzing color talk. But we differ in details, as will be seen below; moreover, in (1984) Almog withdrew his theory presented in (1981) and developed a new one without saying, however, how it applies to color talk.

<sup>4</sup>Cf. also Haas-Spohn (1997).

less accepted,<sup>5</sup> to generally account for epistemological modalities, apriority, linguistically expressible belief, and so on. How does it do that?

A preliminary point is that each context determines its associated index.<sup>6</sup> Thus, each possible utterance of an expression as such has not only an intension, but also an extension, namely the value of the intension at the associated index; in particular, each utterance of a sentence has a truth value. Following Stalnaker, I call the function which assigns to each context the extension an expression has in that context the diagonal of the expression; this function is, so to speak, the diagonal of the two-dimensional scheme that represents the character of the expression. It is this diagonal which does the epistemological job. The general reason is Stalnaker's, and it is very simple. Namely, whenever a speaker utters a sentence or a hearer hears one, they are not fully informed about the actual context; but in any case the speaker believes she says something which is true in the context and the hearer, if he accepts the utterance, believes he hears something which is true in the context. Thus, their epistemic attitudes are directed to possible contexts, that is, have sets of possible contexts (or the corresponding indicator functions) as their objects,<sup>7</sup> and it is the diagonal of the uttered sentence which represents their belief. Clearly, the belief expressed by speaking and acquired by listening is a belief *de dicto*. Consequently – and this is important – the diagonals of sentences are more specifically to be taken to represent the corresponding beliefs *de dicto* (cf., however, footnote 15 below).<sup>8</sup>

If this epistemological account is to work, a crucial hypothesis is required to hold: There is a stock of philosophical arguments showing that the intensions of sentences are (almost) never the objects of belief.<sup>9</sup> But for context-independent sentences having the same intension in every context the diagonal essentially coincides with the intension; and clearly, sentences built from context-independent expressions are in turn context-independent. Therefore, if the diagonal is to perform its epistemological job, most expressions must be context-dependent.

---

<sup>5</sup> Kaplan, for instance, does not fully believe in it; cf. his skeptical remarks in (1977, sect. XXII).

<sup>6</sup> If an index consists only of an index world, the index associated with a context is just the world of that context. The same holds for less simply conceived indices – as long as for each index parameter there is a corresponding context parameter (that this is so is a substantial semantic claim).

<sup>7</sup> This idea is briefly indicated in Lewis (1983, p. 230), and further developed in Haas-Spohn (1995, ch. 2), and Spohn (1997a).

<sup>8</sup> As will become clear, this marks a basic difference between this paper and Nida-Rümelin (1997). Nida-Rümelin holds that in the special case of utterances of sentences like “the sky is blue” the normal speaker expresses the phenomenal, as she calls it, as well as the non-phenomenal belief that the sky is blue; this agrees with her diverging explanation of the character of color terms. By contrast, I think that also in this special case the belief primarily expressed is only the belief *de dicto* (which roughly, though not fully corresponds to what she calls the non-phenomenal belief), and that the ascription of any further beliefs to speakers on the basis of their utterances is licensed only by additional background assumptions which may or may not hold.

<sup>9</sup> The best known references are, of course, Putnam (1975), Kripke (1979), and Burge (1979).

To arrive at the same conclusion in a slightly different way: A sentence is *a priori* (true) if and only if its diagonal assigns truth to all possible contexts; if neither it nor its negation is *a priori* true, it is *a posteriori* or informative. Clearly, a context-independent sentence which is necessary in one context must also be *a priori* and in fact analytic, i.e. true in all contexts and at all indices.<sup>10</sup> However, many necessary sentences are informative and not *a priori*, let alone analytic. Therefore, again, most expressions must be context-dependent.

Hence, for this hypothesis to hold true usual context-dependence as in indexicals and demonstratives is not enough. One has to interpret Putnam's hidden indexicality of natural kind terms as dependence on the context world as it is here understood,<sup>11</sup> and one has to find context-dependence in other predicates, for instance in Burge's examples, and in names as well.<sup>12</sup> For the same reason it will be crucial to find out whether the color terms (A) and (B) are hidden indexicals, i.e. dependent on the context world; this is the only way to tell whether their metaphysics and their epistemology can be treated distinctly in our framework.

So, how then do we determine the character of a given expression? First, we find out what is known *a priori* about the extension of the expression; in principle, we can do this with good old Cartesian methodical doubt. Having done this, we know the diagonal of the expression; we thus have one entry in each row of the two-dimensional scheme. From that entry we project the entire row, that is an extension for all of the other indices. The vehicle for doing this is what may be called the essentiality convention pertinent to the expression. This convention specifies for each context what is essential for the extension of the expression and thus allows to project it onto other indices. It must indeed be assumed that the linguistic community has such an essentiality convention for each of its referring expressions.<sup>13</sup>

A final preliminary point: My epistemological talk is quite loose in an important respect. Usually, belief, *apriority*, informativity, etc. are notions applying to individual subjects; something is believed by, or is informative to, a given individual. On the other hand, I have explained a character to be that of a given natural

---

<sup>10</sup>Or, equivalently, a sentence is analytic iff its necessity is *a priori*. This is Kripke's notion of analyticity in (1972).

<sup>11</sup>How this may be done is explained in Haas-Spohn (1997).

<sup>12</sup>Kaplan was skeptical of the generality of the epistemological strategy (which he had invented for demonstratives) precisely because he denied the context-dependence of names. And Almog withdrew his (1981) precisely because he had there misidentified the context-dependence of names; cf. Almog (1984, pp. 10f.).

<sup>13</sup>For details see Haas-Spohn (1995, sect. 3.5). However, the point is easily explained with Putnam's "water"-example: It is a convention of the English speaking community that "water" is a natural kind term denoting a substance, if there is a single substance underlying most of what we call "water", or any mixture of a few substances, if there are few substances underlying most of our "water"-paradigms, or anything sharing certain superficial characteristics, if no underlying physical structure can be found. This is the English essentiality convention for "water" as Putnam (1975) describes it; and the context world then tells which of the possible cases for which the convention is prepared becomes relevant and thus what is water in other possible index worlds.

language (or, more precisely, of a given and maybe changing state of that language). This entails that all the epistemological notions just derived from the character must be taken as applying to the given linguistic community as a whole and not to any of its subjects; the a priori is that of the linguistic community; informativity is measured by communal standards; etc. Such communal epistemic states are certainly a vague matter, but not worse than meanings and languages; and when talking about the latter, we certainly cannot avoid talking of the former.

There is a certain tension between the individual and the communal notions. Indeed, the tension is irreducible, since I take the communal epistemic state not as a kind of average of all the individual epistemic states or as something like Putnam's stereotype, which may be assumed to be embodied in most or all competent individuals, but rather as a kind of sum of the individual states, as consisting of what is recognized by the community as the best knowledge available to it, which need not be embodied in any individual. If, nevertheless, one wants to stick to the sketched framework, the conclusion is that it has to be doubled, i.e. to be developed on a communal as well as on an individual level, including an explanation of how the two levels relate.<sup>14</sup>

However, all this seems unnecessarily complicated for the present purpose. Therefore I will be deliberately sloppy concerning the two levels, or, rather, my account will explicitly refer to the communal level while pretending – although this is, strictly speaking, false – that it equally applies to the individual level.<sup>15</sup> It seems to me that my account is not essentially affected by this sloppiness; but this is a claim I do not attempt to prove here (even though Nida-Rümelin 1997 may throw doubt on it).

So much about the framework I am presupposing. How does all of this apply to color talk? I shall unfold this in a series of claims:

*Claim 1:* Color terms like (A)  $\lambda x(x \text{ is red})$  are hidden indexicals.

This looks implausible. Our standard example for a hidden indexical is “water”, and at first sight “red” seems to be quite different from “water”. We all might say to the very best of our knowledge: “This is water”, and we might still be wrong, because the alleged sample of water may differ in essential aspects from other samples; water has a hidden nature. On the other hand, if we all say to the best of our knowledge: “This is red”, then that object is red. There seems to be no hidden nature to be found in red things which would separate between genuine redness and fake redness.

However, this is not quite true. Though redness seems to have an overt nature, it does not show it under any circumstances. One's individual color judgment can be

---

<sup>14</sup>This is elaborated in Haas-Spohn (1995, sects. 3.8–3.9); it is here where the crucial difference between Kaplan and Stalnaker unfolds.

<sup>15</sup>In particular, this remark modifies my claim that the diagonal of a sentence represents the corresponding belief de dicto. This is correct only if “belief de dicto” is taken in the unusual communal sense; the beliefs which individual speakers express by utterances are, strictly speaking, not these diagonals. What they do express can be correctly accounted for in the just mentioned doubling of Kaplan's framework.

mistaken; and there is in principle also the possibility of collective error. The light may be strange; there is a whole set of optical tricks and delusions; there is collective madness; and so forth. Thus, the colors show their seemingly overt nature only under normal conditions, and the point is that these normal conditions have a hidden nature. This is most easily and clearly demonstrated, with a familiar type of argument, for the normal conditions concerning illumination.

There is not only daylight and twilight, but also twinlight. Twinlight looks as white and bright as daylight, and under twinlight all the things familiar to us look the very same color as under daylight. Thus, without modern physics we could not tell apart daylight and twinlight, and perhaps even present physics does not yet help. Now imagine that in some possible world there is a kind of objects which we have not encountered so far; let us call them modaleons. In daylight modaleons look deep blue, in twinlight they look glaring red.

In contrast to what Nida-Rümelin (1997) prefers from her point of view, it would not be appropriate, I think, to say that modaleons change color when the index world changes normal light. When talking counterfactually about changing light we would not say, for instance, that sun-flowers would be orange if a huge red filter were fixed between the sun and the earth; rather we would say that they look orange under these circumstances, though they still are yellow. Similarly, we would say that modaleons, which are actually blue, would still be blue, but look red if the world were filled with twinlight.

Consider now different context worlds with different normal light; for all we know the context world we live in may be filled with daylight or with twinlight. If the foregoing is granted, then the modaleon case clearly shows the extension of color terms to vary with the context world. Viewed from a context world filled with daylight, modaleons are blue, whichever index world they inhabit; viewed from another context world filled with twinlight, however, modaleons are not blue, but red. So, this example shows the hidden nature at least of the normal lighting conditions and thus at the same time the context-dependence of the predicate  $\lambda x(x \text{ is red})$ .

This remote reason for the context-dependence of color predicates of the type (A) vanishes, if we turn to color predicates of the type (B); how things look to us at a given moment does no longer depend on such normal conditions. Thus, we might expect that terms of type (B) are not context-dependent; this would also conform to the traditional view that we cannot be mistaken about which color something looks to us at a given moment. But contrary to this I contend:

*Claim 2:* Color predicates like (B)  $\lambda xy(x \text{ appears red to } y)$  are hidden indexicals.

The reason is basically that there are what I take to be clear cases falling under the heading “inverted qualia”. For better explanation I have to introduce a very coarse piece of current color perception theory. As is well known, the human retina contains a lot of cones each of which is equipped with one of three kinds of pigments. All three pigments are sensitive to large parts of the visible spectrum, but in varying degrees. The maximal sensitivity of the pigments lies, respectively, in the red, the green, and the blue segment of the spectrum. So, the pigments are called R-, G-, and B-pigments; and accordingly, the cones containing them are called R-,

G-, and B-cones. A decisive link between the activity of the cones triggered by the incoming light and the color sensation is now provided by the so-called opponent process theory. According to this theory, the activity of the R- and the G-cones is compared closely behind the retina. The more the activity of the R-cones outweighs that of the G-cones, the more reddish is the color impression; and vice versa. Moreover, the activity of the R- and the G-cones is summed up and compared with the one of the B-cones. Again the more the sum outweighs the activity of the B-cones, the more yellowish is the impression; and the more the activity of the B-cones preponderates, the more bluish is the impression.<sup>16</sup> It is important not to get confused here about the classifications underlying the labels R, G, and B. The pigments so labelled are classified according to their chemistry.<sup>17</sup> By contrast, the opponent process theory offers a functional criterion for classifying cones as R-, G-, and B-cones; they are so classified according to their subsequent wiring.<sup>18</sup>

One of the many explanatory achievements of the opponent process theory is that it can explain dichromatism or red-green blindness. The explanation is simply that for some reason both the R- and the G-cones contain the same pigments so that their activity is always the same and no impression tends to be reddish or greenish.<sup>19</sup>

Now, Piantanida (1974) had a special hypothesis about dichromatism. Obviously, red-green blindness may come in two forms; either the R-pigments are contained also in the G-cones, or the G-pigments are contained also in the R-cones. Piantanida conjectured, very roughly,<sup>20</sup> first that both forms are due to genetic defects, secondly that these defects are located on different genes and are thus statistically independent, and thirdly that there is consequently a slight chance of suffering from both defects. For male persons this chance is about 1.4 per thousand. Would such a male be color-blind? No; his discriminatory powers are precisely as fine-grained as ours, only his reddish and greenish impressions are reversed. Such persons are called pseudonormal. Obviously, it is very difficult, if not impossible without violating bodily integrity to find out about pseudonormality, even for the pseudonormals themselves. But perhaps you, dear reader, are one of those! It is not so unlikely; for instance, about 58,000 of the 40 million male Germans would be pseudonormal, if Piantanida is right!

I do not know the scientific fate of Piantanida's hypotheses, and I cannot assess their scientific plausibility. But clearly, they make perfect sense, they are testable,

---

<sup>16</sup>The details are quite complicated, however, and empirical research is extremely difficult; cf., e.g., Boynton (1979, chs. 7 and 8).

<sup>17</sup>In fact, there occur not only the three normal forms, but also a number of chemical variations; cf. Boynton (1979, ch. 10).

<sup>18</sup>Due to their symmetrical role the issue of distinguishing R- and G-cones is quite subtle, however; cf. Nida-Rümelin (1997) for more detailed considerations.

<sup>19</sup>Note that this explanation presupposes the independence of the classifications of pigments and cones which I have just stated.

<sup>20</sup>For details, see also Boynton (1979, pp. 351–358).

and they might well turn out to be true.<sup>21</sup> The crucial point is how we should talk about pseudonormals. I find it very clear that the right way to talk about them is just as I did, namely that their reddish and greenish sensations are reversed; thus, red peppers look green to them and green peppers look red to them. I would not know how to conclusively refute those who refuse to talk that way, but it will become still clearer in the course of the paper that this is indeed a meaningful way of talking.<sup>22</sup> One may also sense an ambiguity and think that it is equally appropriate to say that red peppers look red to pseudonormals, that is, look as red things look to them. I shall discuss this alleged ambiguity in a moment; but the primary sense of “looks”, and the one I am presently referring to, is the one in which red peppers look green to pseudonormals.

Now I am finally prepared to explain the context-dependence of the term (B)  $\lambda xy(x \text{ looks or appears red to } y)$ . Take a situation in which someone with G-pigment in his R-cones and R-pigment in his G-cones looks at a ripe tomato. Viewed from our actual context world where most English speaking people have R-pigment in their R-cones and G-pigment in their G-cones, that person has a deviant color perception, and the situation must be described as one in which the ripe tomato appears green to him. Viewed from a context world, however, in which most English speaking people have their pigments reversed,<sup>23</sup> that person is perfectly normal; and the situation must be described as one in which the tomato appears red to him. Thus, to conclude, the truth value of “that tomato appears red to this person” as applied to one and the same situation varies with the context – whence the context-dependence of appearance terms.

Is that meant to say that you may be mistaken when you, well educated, fully attentive, and absolutely sincere, as you are, say: “This tomato looks red to me”? Yes, precisely. Unbeknownst to you, you may be pseudonormal, and your utterance may thus be false. The point of the argument is simply that the application of  $\lambda xy(x \text{ appears red to } y)$  is relative to a standard of normal vision, that the context world sets this normality standard, that the nature of this standard is unknown, and that no one knows for sure whether he conforms to that standard or not.

This seems to make the doubtful presupposition that there is a standard of normal vision. Is it not possible that Piantanida’s statistics is wrong and that, say, a third of the population is pseudonormal? Surely; in fact, if one looks at perception experiments, one sees a surprisingly large variation in human color perception.<sup>24</sup> But I do

---

<sup>21</sup> Hilbert (1987, p. 92), seems to be the first to have mentioned pseudonormality in the philosophical literature; but apparently only Nida-Rümelin (1993, ch. 4 and 1996) fully realized its philosophical significance.

<sup>22</sup> The point is more fully argued by Nida-Rümelin (1996).

<sup>23</sup> Clearly, this is a possible context world. Which kind of biochemical substance is in which kind of so-and-so connected cones of most English speaking people is a contingent matter about which we need not have any knowledge.

<sup>24</sup> Cf. Boynton (1979, ch. 10), and Hardin (1988).



not need this presupposition, just as Putnam need not presuppose that water, or jade, for that matter, is just one substance.<sup>25</sup> On the contrary, our essentiality convention for appearance terms responds flexibly to various empirical findings.

For further explanation, I would like to relate this point to the familiar view due to Chisholm (1957, ch. 4), that appearance terms have three different readings, a phenomenal, a comparative, and an epistemic reading. This is my

*Claim 3:* The phenomenal, the comparative, and the epistemic interpretation of  $\lambda xy(x \text{ appears red to } y)$  are not three different readings; they rather reflect the context-dependence of this term by being appropriate in three different kinds of contexts.

Let me briefly recall these three interpretations:

- (E) According to the epistemic interpretation, “ $x$  appears red to  $y$ ” says as much as “in the absence of counter-evidence,  $y$ ’s encounter with  $x$  tends to produce  $y$ ’s belief that  $x$  is red”.<sup>26</sup>
- (C) According to the comparative interpretation, “ $x$  appears red to  $y$ ” means “ $x$  looks to  $y$  in the way red things usually look to  $y$ ”.
- (P) For the phenomenal interpretation, finally, there is no such paraphrase; there “ $x$  appears red to  $y$ ” holds only if  $y$  has a specific common type of qualitative experience.

We have seen<sup>27</sup> that according to our essentiality convention for “water” the essential properties for being water depend on the actual properties of the many “water”-paradigms we have in the context world – whence the context-dependence of “water”. The very same is true of “appearing red”, as these three interpretations reflect:

Imagine *case 1* which I take to be actually obtaining: In this case there are few people with deviant perceptual capacities; there are few color-blinds and few or no pseudonormals. There may be variations; the sensitivity of the pigments may slightly differ in different people; the neurons comparing the activities of the cones may not respond in a completely uniform way; and so on. But on the whole most people have a roughly equal functional and physiological arrangement of the visual apparatus including higher brain regions. In that case, we would apply  $\lambda y \forall x(x \text{ appears red to } y)$ <sup>28</sup> only to those normal people whose visual system is in a certain state; we could apply it also to some deviant people, if their deviation is as simple

<sup>25</sup> Cf. Putnam (1975, pp. 239–241).

<sup>26</sup> Or in Pitcher’s more careful words: “ $y$  causally-receives, by means of using his eyes in the standard visual way, the (perceptual) belief, or an inclination to have the (perceptual) belief, or a suppressed inclination to have a (perceptual) belief, that  $x$  is red”; cf. Pitcher (1971, pp. 85–95).

<sup>27</sup> Cf. footnote 13 above.

<sup>28</sup> This is to replace the awkward colloquial phrase “is appeared red to” introduced by Chisholm (1957, p. 62), by a less awkward formal phrase.

as that of pseudonormals. But we would not further extend the application. In that case, i.e. in such context worlds, the appropriate interpretation of  $\lambda_{xy}(x$  appears red to  $y)$  is the phenomenal one in which it involves a particular phenomenal quality.

Now compare this with *case 2*. Its simplest version is that there are so many pseudonormal persons that they cannot be dismissed as deviant; there are just two normal kinds of visual systems. In that case, each group can claim with equal right that ripe tomatoes, for instance, look red to its members; it would have no point if the members of either of the two groups insisted that tomatoes look red only to them. Thus,  $\lambda_{xy}(x$  appears red to  $y)$  does not involve a certain phenomenal quality in this case. This is particularly clear from the fact that in this version objects appearing red to one group produce the same phenomenal quality as objects appearing green to the other group. It is still clearer in cases where there are many human visual systems which even the most advanced future science is unable to match; then the phenomenal qualities experienced by our fellows would be just as foreign to us as those of the bat. Still, color talk miraculously runs as smooth as it does. So, these would be cases or contexts in which the comparative interpretation of  $\lambda_{xy}(x$  appears red to  $y)$  is appropriate;  $\lambda_y \vee x(x$  appears red to  $y)$  would then be applicable to all beings having qualitative experiences which somehow enable them to discriminate and classify red things as we do, even though this ability would remain mysterious.<sup>29</sup>

There is the even less demanding *case 3*, the absent qualia case. It seems perfectly imaginable that some individuals behave in the very same way as we do without having any phenomenal experience at all. Why should computers be able to pass the Turing test only if they have built in sensations? Think also of such things as blind-sight where people with a specific brain damage behave towards objects similarly as normal people do, but are unable to report any conscious visual experience.<sup>30</sup> If this is imaginable, it might turn out to hold in the context world.

---

<sup>29</sup> It may be that I have overestimated human uniformity and that human vision is so varied as to rather fall under Case 2; this is an empirical question (possibly undecidable due to vagueness). However, Strawson (1989) argues, I understand, that Case 2 yields the appropriate description of the meaning of "red" in any case. This is the main point where I do not agree. In an important argument (sect. 6) he considers Monet and Renoir color vision (which is analogous to normal and pseudonormal vision) and asks whether the meaning of "red" changes when English gets smoothly translated into the language of a population with Renoir vision (or when the share of Renoir vision among English speaking people slowly increases from 0 or 1 to 99 or 100 percent). His answer is: surely not; and the reason seems to be that there cannot be meaning changes which nobody noticed. However, if meanings are explicated as characters there can be unnoticed meaning changes, as is carefully explained in Haas-Spohn (1997, sect. IV). Think again of "water" (which is less confusing than "red") and of Putnam's twin earth: It makes a difference whether we travel there before or after being able to distinguish between H<sub>2</sub>O and XYZ. If we travel there after having this ability, XYZ never gets into the extension of English "water". But if we travel there before (and do this very often and develop a close interchange with twin earth), then the character and indeed the extension of English "water" has changed; at the outset XYZ did not belong to it and later on it does. Strawson apparently does not observe this difference.

<sup>30</sup> Cf., e.g., Weiskrantz (1980).

You are presumably quite sure that you have phenomenal experiences. But perhaps you are one of the very few gifted people; the normal case may be to have no visual experiences at all, but to talk as if one had some. But we would still have our usual color talk. And we would still have beliefs; in some mysterious way our beliefs are pushed this way and that way by our encounters with the things in the world. This then is the way things appear to us. So, in this extreme case at least, only the epistemic interpretation of  $\lambda xy(x \text{ appears red to } y)$  seems appropriate.

So, what is essential for  $\lambda xy(x \text{ appears red to } y)$  depends, according to our linguistic essentiality convention, on how the context turns out to be; and the three interpretations just mark three significantly different kinds of contexts. They are thus integrated into a single character of  $\lambda xy(x \text{ appears red to } y)$ .

Why then did they appear to be three different readings and thus to uncover an ambiguity? The reason, it seems to me, is that “appear” and “look” are conjoined not only with “red”, but with many other phrases as well. In fact, the usual claim associated with these readings is that the scheme “looking *F*” (and not its instantiation “looking red”) has three different readings, depending on what is taken as *F*; and this claim is usually accompanied by quite determinate opinions concerning which reading is appropriate for which kind of *F*. My claim 3 interprets this determinateness as a (maybe unreflected) certainty about the actual context world and the interpretation of “looking *F*” pertinent to it.

For instance, it seems very likely that we live in a context world where “appearing red” carries the phenomenal interpretation. Again, circumstances seem to be such that a comparative interpretation is most appropriate for phrases like “appearing square” or “looking like a capital A” which are about simple forms possibly appearing in many different ways. Finally, as things stand, the epistemic interpretation seems applicable not only to perceiving beings without phenomenal experience, but also to us for phrases like “appearing to be a car” where the appearance is phenomenally too complex and varied and best reduced to the proximate epistemic effect.

However, for all these instantiations of “looking or appearing *F*” it seems possible to imagine cases which show the same context-dependence as I have displayed it for “appearing red”. Imagine, for instance, beings who have phenomenal experiences, but who see only letters, maybe in Garamond. Thus, if a car is approaching them, they read “car” written in Garamond in the relevant place of their visual field (strangely, these beings are tuned to English); and one may refine the example by giving meaning to the size and color of the letters in their visual field. For such beings a phenomenal interpretation of “appearing to be a car” seems appropriate.<sup>31</sup> Hence we find the three interpretations not only across the various instantiations of the scheme “looking *F*”, but indeed within each locution of this type; and this, so I

---

<sup>31</sup> This example came to my mind when reading Cresswell (1980, pp. 129–131), where he invents similarly weird examples for arguing, contra Jackson, that there is no difference between “looking red” and “looking like a tomato”, i.e. that both may be equally given a phenomenal and a comparative reading. This argument further illustrates my present point.

have argued, is better accounted for by giving this locution one context-dependent meaning. Only if one neglects this context-dependence do there seem to be three different readings or meanings.

Let me summarize the point of claim 3 in a somewhat different way: There are two extreme views to be found in the literature. Some think that subjects have certain types of sense impressions, qualitative experiences, or however one may call them, that we can refer to the subjects' having them, maybe even in a direct or rigid way, and that we in fact do so with such expressions like  $\lambda y \vee x(x \text{ appears red to } y)$ .<sup>32</sup> Perhaps the most famous expression of the opposite view is found in Wittgenstein (1953, sect. 293) where he ponders about how we could talk about the alleged beetles in our boxes when everyone can look only into his or her own box and where he says:

Here it would be quite possible for everyone to have something different in his box. One might even imagine such a thing constantly changing. – But suppose the word “beetle” had a use in these people’s language? – If so, it would not be used as the name of a thing. The thing in the box has no place in the language-game at all; not even as a *something*, for the box might even be empty. – No, one can “divide through” by the thing in the box; it cancels out, whatever it is.

My claim 3 proposes a middle course agreeing, in a way, with both views. Wittgenstein is right: Our communication with the very language and the very color expressions we have obviously works well, however the empirical facts turn out to be and even if our boxes are empty; and a theory of meaning for this language has to do justice to this fact, as my account tries to do.<sup>33</sup> But under happy circumstances, we have the same kind of beetles in our boxes, and then we do talk about our beetles. Surely, it is very likely that the circumstances are happy; confirming this is not impossible, and with such findings as the opponent process theory we are indeed beginning to confirm this.

I turn now to the relation between the two kinds of color predicates, i.e. between (A)  $\lambda x(x \text{ is red})$  and (B)  $\lambda xy(x \text{ appears red to } y)$ . Again, we find two extreme views both of which hold that the one predicate is definable by or reducible to the other. On the one hand, those maintaining an objectivist account of colors insist that  $\lambda x(x \text{ is red})$  denotes an objective property, i.e. a primary quality of objects by themselves.<sup>34</sup> They may add that  $\lambda xy(x \text{ appears red to } y)$  should be explained by how subjects respond to objects' being red; the comparative and the epistemic interpretation of  $\lambda xy(x \text{ appears red to } y)$  precisely are attempts to give that explanation. But since the phenomenal interpretation turned out to be appropriate at least for some contexts, e.g., those where vision is realized in most humans in a relatively uniform way, this additional claim does not seem defensible. On the other hand, those

---

<sup>32</sup>Cf., e.g., Nida-Rümelin (1997). As she emphasizes in sect. 3, this does not necessarily require to posit objects like impressions, sensations, etc.

<sup>33</sup>This is not to say, however, that my application of Kaplan's framework to color talk would have a specifically Wittgensteinian character.

<sup>34</sup>Cf., e.g., Jackson and Pargetter (1987).

defending a subjectivist account of colors say that the colors of objects are secondary qualities which can only be explained by how the objects appear to us; they would thus cite statement:

(1)  $x$  is red if and only if  $x$  would appear red to most English speaking people under normal conditions

as a definition or an analysis of being red.<sup>35</sup> In order to assess this, we have to inquire into the modal status of (1).<sup>36</sup> This is the content of my:

*Claim 4:* (1) is a priori in English<sup>37</sup>; but it is analytic only in one reading and not analytic in another reading.

We find out about the apriority of (1) by evaluating it in each context (and its associated index). There (1) seems to be true; I cannot imagine any context world in which the two sides of the equivalence in (1) would differ in truth value. In any case, (1) is true in all three kinds of contexts for which the three interpretations of  $\lambda xy(x$  appears red to  $y)$  are, respectively, appropriate. This just reflects the point noted earlier that the colors of things have an overt nature given normal conditions; in no context can an object which appeared red to most of us under normal conditions turn out not to be red, and vice versa.<sup>38</sup>

We find out about the analyticity of (1) by evaluating it in each context and each index; (1) is analytic if and only if it is true in all of them. Consider, for instance, the actual context with a ripe tomato before us and transfer the tomato just as it is into some counterfactual index world  $i$ . Thus, this tomato is as red in  $i$  as it is here. Now let us assume that most English speaking people in  $i$  are pseudonormal, that is, have R-pigment in their G-cones and G-pigment in their R-cones; this is definitely a possible counterfactual supposition. Since I take it that the actual context is one in which pseudonormals are deviant people with red-green-inverted sensations,

<sup>35</sup>This is explicitly done, for instance, by McGinn (1983, pp. 5–14) – with the exception that he would not restrict (1) to English speaking people. This minor difference is cleared up in footnote 37.

<sup>36</sup>In Spohn (1997c) [here: ch. 12] I have again used the Kaplanian framework for a related inquiry into the modal status of reduction sentences for dispositional predicates in general and also into the epistemology and ontology of normal conditions. This inquiry may further illuminate the following discussion.

<sup>37</sup>This explains my restriction of (1) to English speaking people; the primary standard for how “being red” works in English is the English speaking community. The situation changes as soon as there are established translations between English and other languages; then the people speaking these other languages become equally important. Cf. Haas-Spohn (1997, sect. III).

<sup>38</sup>Since I am talking about apriority in English. I am not claiming that (1) is a priori for any subject. This also entails that it would be inappropriate to object that a thing could be red in a context where there are no English speaking people for it to appear red. This is so because the existence of the English speaking community is a priori in English, similarly as my existence is a priori for me. But, of course, the existence of the English speaking community is not analytic in English. This as well as the mere apriority of one reading of (1) are examples showing that apriority in English is an independent notion reducible neither to analyticity nor to subjective apriority. And this in turn shows that such communal epistemic notions are needed; at least there is some work for them to do.

we would then have to say that most English speaking people have red-green-inverted sensations in *i*. So in particular, it is true in the index world *i* that the red tomato appears green to most English speaking people. This amounts to a counter-example to the analyticity of (1).

However, my claim 4 seems to state an ambiguity in (1). Where is it? I have just understood “most English speaking people” in an attributive way by evaluating it at the index world *i* itself. But we can take this phrase also in a referential way as referring to most English people not in the index, but in the context world *c*. The same kind of ambiguity can be found in the phrase “under normal conditions”; it can refer to conditions counting as normal in *i* or counting as normal in *c*.<sup>39</sup> To make this referential reading more explicit, statement (1) thus read says in each context *c*:

(4) *x* is red in the index world *i* if and only if *x* would appear red in *i* to most English speaking people from *c* under conditions normal in *c*.

The above counter-example does not apply to this reading. Indeed, I cannot think of any counterfactual index world *i* for which this reading would not be true. If this is correct, then (1) is analytic in this reading.<sup>40</sup>

There would be no point in arguing which reading is more natural. The important thing to note, I think, is that (1) is a priori in any case and that it is analytic only in the referential reading (of the relevant phrases), but not in its attributive reading.

Does this result support the subjectivist in any way? No. A preliminary point to note is that an analytic equivalence like (1) in its referential reading need not give a definition or analysis; as an analytic truth it just states a certain meaning relation. But since it is not so clear, anyway, what an analysis or a definition (of an already meaningful term of natural language) is beyond an analytic equivalence, we had better concentrate on the subjectivist’s claim that colors are secondary qualities, or that predicates of type (A) are dispositional or, more abstractly, relational, i.e. relative to perceiving subjects:

What the subjectivist would need is the necessary or, as McGinn (1983, p. 14), puts it, intrinsic dispositionality or relationality of type (A) terms. This, however, cannot be inferred from my claim 4.  $\lambda x(x \text{ is red})$  would be necessarily relational in a given context if and only if, viewed from that context, an object could be red in an index world only if it would stand in a certain (maybe only dispositional or counterfactual) relation to other objects in that index world. (1) seems to assert such a thing; but it does so only in its attributive reading which I observed to hold only a priori, i.e., not to be projectible from the context to all indices. Nor does the referential

---

<sup>39</sup>The distinction between an attributive and a referential use of denoting phrases was originally introduced by Donnellan (1966), however in an apparently different way. By using it as just explained I refer to its standard interpretation within the Kaplanian framework which is to be found in Stalnaker (1970) and Kaplan (1978).

<sup>40</sup>Maund (1986, pp. 173–176) makes a similar point by distinguishing a purely comparative and a referential use of “looks” and arguing that something like (1) is analytic in the first, but synthetic in the second use; however, he does not represent his distinction within the Kaplanian framework.

reading help, despite its analyticity, since it asserts only a certain relation between objects' being red in an index world and the people in the context world. Compare this with  $\hat{\lambda}x(x \text{ is a mother})$ , the standard example of a necessarily relational predicate. Here, viewed from any context world, someone is a mother in an index world if and only if someone else exists or has existed in that index world who is her child. The analogous assertion for "being red" is simply not licensed by the modal status of (1).<sup>41</sup>

What is licensed by the apriority of the attributive reading of (1) is the conclusion that  $\lambda x(x \text{ is red})$  is a priori relational. But this is no peculiarity of color terms; according to Haas-Spohn (1997), "water" and other natural kind terms, in fact all hiddenly indexical predicates and thus many predicates which unquestionably denote primary qualities are a priori relational.<sup>42</sup>

But even if the necessary relationality of  $\lambda x(x \text{ is red})$  cannot be demonstrated, it may actually hold in a given context. Maybe; I was, however, unable to imagine such a context. I shall return to this issue with my claim 6 when I speculate about our actual context world.

So far, I have considered all possible contexts for our two sample predicates (A) and (B). I have, implicitly and explicitly, discussed their diagonal and how to project their extension from contexts to other indices; in this sense I have carried out an epistemological and semantical investigation. After all this it is not so difficult to give a summarizing definition of the character of these predicates.<sup>43</sup> In the

---

<sup>41</sup> Jackson and Pargetter (1987, pp. 130f.) argue for the same point. They call (1) the dispositional truism and argue that the truism does not justify one in identifying colors with the disposition to look colored; rather, colors should be identified with the categorical base of this disposition. I agree, if one takes the difference between a disposition and its categorical base to be only an epistemological one: The intension of a dispositional predicate and the intension of the predicate describing the categorical base are the same, both predicates denote the same property; but their diagonals and a fortiori their characters are different. With this account of the difference between dispositions and their bases, my argument is the same as theirs. However, Prior et al. (1982), pp. 253ff., give an ontological account of the difference. So, there remains a disagreement. I discuss this disagreement more fully in Spohn (1997c) [here: ch. 12].

<sup>42</sup> Compare also the discussion about the notion of response-dependence and its rigid and its non-rigid interpretation which addresses the very same issues; cf., e.g., Vallentyne (1996). I think this notion nicely fits into the Kaplanian framework; its rigid and its non-rigid interpretation, in particular, corresponds to the referential and the attributive use as explained above.

<sup>43</sup> Applying the general scheme of Haas-Spohn (1995, p.151) to our sample predicates we get:  $x$  is red in the context  $c$  and the index  $i$  iff  $x$  shares in  $i$  all the properties which, according to the English essentiality convention for "being red", are essential in  $c$  for the redness of the objects to which "being red" is typically applied in  $c$  by the English speaking community; and  $x$  appears red to  $y$  in the context  $c$  and the index  $i$  iff  $x$  and  $y$  possess in  $i$  all the properties and relations which, according to the English essentiality convention for "appearing red", are essential in  $c$  for the relation of appearing red between any two objects to which "appearing red" is typically applied in  $c$  by the English speaking community. This abstract explanation is neither circular nor badly metalinguistic (cf. Haas-Spohn, 1997 and 1995, sects. 3.4–3.5). But it is less illuminating than the substantiation of its key terms for the case at hand; and I am here rather concerned with the latter.

rest of the paper, however, I shall engage in a metaphysical speculation concerning the intension of these predicates at our actual context world.

I first take up statement:

(2)  $x$  appears red to  $y$  if and only if  $x$  (appropriately) causes  $y$  to be in a neural state of the kind  $N$  by defending:

*Claim 5:* For some (possibly disjunctive) kind  $N$  of neural states, statement (2) is necessarily true in the actual context.

The parenthetical “appropriately” in statement (2) is to exclude deviant ways of causation which do not count as an object’s appearing to a subject; but I am not concerned with spelling out what is to count as appropriate here.<sup>44</sup> Of course, claim 5 similarly holds for other color appearance terms; thus it says in effect that color appearance terms are strongly supervenient on neural state terms, or, what comes to the same, that the property of having-a-so-and-so-colored-sensation is type-type identical with the property of being in a certain (possibly wildly disjunctive) neural state.<sup>45</sup>

Claim 5 consists of two parts. The first is a factual hypothesis, namely that (2) is actually true for some  $N$ , or rather that in most of our actual paradigm cases for some object’s appearing red to some subject we find in that subject an activation of a certain neural structure or a realization of a certain, possibly very complex neural state. As far as I know, brain research has not come up so far with results disconfirming this hypothesis; but perhaps I am underestimating the complexity and diversity of neurological findings. On the other hand – perhaps I am again underestimating present expertise – we do not seem to have any good theory what that kind  $N$  of neural state might be. But this only shows how poor our knowledge is; it does not show the senselessness or illegitimacy of that hypothesis.

The second part of claim 5 is a claim about our essentiality convention for  $\lambda xy(x$  appears red to  $y)$ , namely the claim that, given the factual hypothesis that we find a neural state type  $N$  uniformly realized in most of our paradigm cases of  $\lambda y \vee x(x$  appears red to  $y)$ , this state type  $N$  provides the essence of  $\lambda y \vee x(x$  appears red to  $y)$ ; that is, we would correctly apply  $\lambda y \vee x(x$  appears red to  $y)$  only to factual and counterfactual cases in which this state type is realized. So, this is rather a linguistic claim about our counterfactual talk. It is to be defended mainly against two doubts.

One doubt is whether, given the factual hypothesis, the essence of  $\lambda xy(x$  appears red to  $y)$  is really to be conceived so narrowly as to conform to no wider than the phenomenal interpretation. I have briefly discussed this already in case 1 following the three interpretations (E), (C), and (P). One possible ground for abandoning this doubt is how I said we would talk about pseudonormals; when we say that red peppers would appear green to them, we precisely assume the narrow essence. Another

<sup>44</sup>Cf., e.g., Lewis (1980c).

<sup>45</sup>Here I identify the property expressed by a predicate with its intension, so that necessary universal equivalence of two predicates is necessary and sufficient for the identity of the properties expressed. For the equivalence of strong supervenience and type-type identity cf., e.g., Kim (1984, sect. IV).



possible ground is that we say that, strictly speaking, nothing appears green or red to red-green blind people even if they should have other clues for correctly guessing the colors. Still another ground is that we refuse, as I think we should, to carry over human color talk to, say, bees upon finding that bees carve up the space of electromagnetic wave mixtures in quite a different and incomparable way than we do. So this doubt seems unfounded.

The other doubt is whether claim 5 provides a correct understanding of the phenomenal interpretation. One may rather think that it is the phenomenal quality itself which is essential for red appearances, i.e. that, necessarily,  $x$  appears red to  $y$  if and only if  $x$  (appropriately) causes  $y$  to have a red-sensation<sup>46</sup>; it would thus be a matter of contingency which kind of brain-states are correlated with red-sensations.<sup>47</sup> I have two reasons for resisting this doubt.

First, if the correlation of neural states with red-sensations is contingent in any case, then conceiving  $\lambda_{xy}(x$  appears red to  $y)$  as context-dependent and describing this dependence as I did in claim 3 loses its plausibility; it goes together more naturally with the view (endorsed by Nida-Rümelin 1997) that everyone, when claiming that something appears red to him, refers to the kind of phenomenal quality which he is just experiencing and the awareness of which leaves no room for error and thus for hidden indexicality. The consequence of conceiving  $\lambda_{xy}(x$  appears red to  $y)$  as involving a fixed kind of phenomenal quality in all contexts and indices is, however, that the few pseudonormals, if they exist, always refer to another quality than normal people do, hence use  $\lambda_{xy}(x$  appears red to  $y)$  with a different meaning (character) and speak, in a sense, a different language. The more varied version of case 2 mentioned after the three interpretations (E), (C), and (P) comes out even worse according to this view; there would be a Babylonian confusion where  $\lambda_{xy}(x$  appears red to  $y)$  would have many different meanings and people would talk many different languages.

This seems unwarranted to me. I do not know whether cases 1, 2, or 3 obtains (though I have already expressed my prejudice); but in any case I see no reason to assume such a possible multiplicity of languages. For instance, if case 1 should turn out to hold and if some pseudonormals should be identified, my prediction would be that these pseudonormals would not insist to continue speaking as before; they would rather correct themselves and agree to such things as that, strictly speaking, red tomatoes look green to them, i.e. they would submit to common usage. Or, if, to our great surprise, case 2 should turn out to obtain, my prediction is that linguistic practice would not change a bit; after this discovery, all of us would talk of

---

<sup>46</sup> Here, the unusual locution of having a red-sensation is defined as denoting the property which is caused to apply to a subject by an object iff that object appears red to it; in other words, it denotes the intrinsic, non-relational property which a subject has whenever the relational property  $\lambda_y \vee x(x$  appears red to  $y)$  applies to it (and which a subject may also have, as it turns out, without external cause).

<sup>47</sup> Certainly, this better catches the intentions of the adherents of the phenomenal interpretation. Kripke (1972) seems to think so with respect to pains (though not necessarily with respect to colors) (cf. pp. 334ff.). Clearly, Nida-Rümelin (1997) also favors this view.

things appearing red to us as we did before. This does not look like a discovery of many languages where there seemed to be only one.<sup>48</sup>

Maybe, however, the disagreement is not about the context-dependence of  $\lambda xy(x$  appears red to  $y)$ , but only about the essential properties of  $\lambda xy(x$  appears red to  $y)$  in the presumably obtaining case 1. Then I have a second reason for sticking to claim 5, namely internal realism.

Internal realism, as I understand it, asserts that truth is believable or discoverable; given a correct understanding of the “-able” – this is all-important – I believe that internal realism provides the defensible core of verificationism.<sup>49</sup> Now, it seems to me that internal realism may be strengthened to assert that essences are believable or discoverable. I have no clear argument for this claim<sup>50</sup>; but if so much is granted, my argument can proceed.

Let us imagine that we have investigated vision in human as well as in other sensing beings as completely as possible; for instance, we have constructed fabulous devices with which we can scan brain states in real time. After endless ingenious theorizing and ingenious experimenting we have come up with our final theory about vision, how visual input is processed, how consciousness comes into play, how all this leads to linguistic and other behavioral output, etc.<sup>51</sup> According to internal realism this final theory which cannot be shattered or improved by any further findings is true. The final theory will contain many equivalences of the form (2) all of which are true; an object will appear red to a perceiver if and only if a many-membered chain of events is realized each of which is a necessary and sufficient cause of the later ones. Among all these equivalences there will be one referring to a special neural state type  $N^*$  with the further characteristic that, given a subject  $x$  is in state  $N^*$ , there is no further or overriding reason whatsoever for or against  $x$ 's having a red-sensation and that, given a subject  $x$  is not in state  $N^*$ , there is no further or overriding reason whatsoever for or against  $x$ 's not having a red-sensation; that is, any reason for a divergence between being in state  $N^*$  and having a red-sensation would at the same time disconfirm the final theory. But then it would be strange to insist that the essence of having a red-sensation does not consist in the neural state type  $N^*$ , but in something else. In any case, no reasons whatsoever

---

<sup>48</sup> If these predictions would turn out false, however, this might well be reason for me to revise my position.

<sup>49</sup> I interpret the “-able” in the following way: the set of a posteriori truths and our inductive standards (taken in a broad sense) must be such that each truth is inductively supported by other truths (conditionally on arbitrarily many truths) and can thus be believed on true grounds. In Spohn (1991) I formally explicated this idea and proved it to be equivalent, in a way, with the universal feasibility of causal explanation.

<sup>50</sup> A major difficulty is here to adapt all the notions involved in the explication of internal realism to the more sophisticated Kaplanian framework. In Spohn (1991) [here: ch. 9] I have not dealt with this difficulty simply because I was not yet aware of it.

<sup>51</sup> Maybe we even have constructed a transmitter cap and a receiver cap directly connecting two brains, and our final theory says that the human under the receiving cap should experience similar sensations to the being under the transmitting cap.

could be adduced in favor of this, not even by the perceiver herself; and then it is simply false according to the strengthened form of internal realism.<sup>52</sup>

The parallel claim concerning statement:

(3)  $x$  is red if and only if the reflectance spectrum of the surface of  $x$  is of the kind  $R$  is my final:

*Claim 6:* For some (possibly disjunctive) kind  $R$  of reflectance spectrum, statement (3) is necessarily true in the actual context.

In order to see this, we do not have to do much more than putting together claim 5 and the analytic reading (4) of assertion (1). If we do this we get:

(5) In the actual context it is necessarily true for some neural state type  $N$  that  $x$  is red if and only if  $x$  would cause most of the actual English speaking people under actual normal conditions to be in state  $N$ .

Now, there are certainly many ways for people to get into a neural state of kind  $N$  and for a given object to bring this about; the actual causal story seems to be a matter of contingent physics and of contingent neurobiology. So how do we get from the above necessary truth in (5) to the necessary truth of (3)? This is achieved by the reference to normal conditions. Recall my speculation about twilight and the modaleons. Of course, an index world may be filled with twilight, and because physics is very different there, modaleons there produce state type  $N$  in us, that is, they appear red to us. But as I have already argued after claim 1, modaleons, when viewed from the actual context, would not count as red in that index world, but as blue, because under normal conditions such as daylight they would appear blue to us. Similar considerations apply to the normal conditions within the subjects like not being mad or intoxicated, and so on. Thus it is the function of the reference to normal conditions to keep the kind of causal process between visible objects and the observers as it normally is in the actual context world fixed throughout all possible index worlds. This enables us to locate, so to speak, the color of an object with necessity in the object itself; we do not have to settle for merely contingent correlations between the physical properties of an object and its color. And for all we know, it is the reflectance spectrum of the object's surface which is the relevant

---

<sup>52</sup> Let me clarify the hypothetical and the positive content of the argument: In any case, I think, the final theory will come up with some equivalence of the form " $x$  has a red-sensation iff  $x$  is  $P$ " with the characteristic just described. My positive claim is then that, according to strengthened internal realism, this  $P$  is the essence of having a red-sensation; and my hypothetical claim is that this  $P$  will actually turn out to be of the form "being in neural state of type  $N^*$ ". But the latter seems at least plausible. In any case, if we tend not to leave it open, but to positively assert on the basis of the opponent process theory that green peppers look red and red peppers look green to pseudonormals and to stick to this until receiving counter-evidence, we are on the track of searching for, and being prepared to accept, ever more sophisticated neural conditions for having red-sensations – a track which will eventually lead us to the type  $N^*$  required for Claim 5 to be true.

physical property. Of course, the class *R* of reflectance spectra characterizing redness forms an extremely wild and certainly quite vague region in the space of possible reflectance spectra. This is so because the class *R* is specified only in relation to the equally vague neural state type *N* and thus to a very complex biological contingency.<sup>53</sup>

According to claim 6, the nature of being red is hidden and unknown. Did it not seem to be overt? Yes, it seemed so. But then we observed with claim 1 that already the normal conditions have a hidden nature. With claim 2 we realized that the nature of appearing red is even more profoundly hidden. And this entails via the analyticity of (4) that the nature of being red is equally profoundly hidden. The appearance of overtness could be confirmed only under the variant of the phenomenal reading of  $\lambda xy(x \text{ appears red to } y)$  which I have criticized under claim 5.

According to claim 6, moreover, colors are not dispositional properties or, more specifically, secondary qualities of objects, contrary to a familiar view. What claim 6 does, in effect, is simply to identify redness, i.e. the disposition of appearing red with its categorical base. My general presumption is here that many, though probably not all dispositions are such that having the disposition is necessarily, though certainly not analytically equivalent with realizing the categorical base of the disposition. This looks implausible only if one confuses ontology and epistemology. One may say that being red is a dispositional concept, since it is a priori according to claim 4 that red things have the disposition to look red; and it is this disposition which determines for each context which property being red is. But this is an epistemological point which does not entail the ontological point that this property itself is dispositional.<sup>54</sup> The epistemological point is also reflected in the fact that in order to find out about, and succinctly describe, the class *R* of reflectance spectra we have to find out about, and to refer to, the human visual system and possibly to the class *N* of neural states. But again, this does not entail that the property of having a reflectance spectrum belonging to the class *R* would be relational in any way.

To be sure, what claims 5 and 6 say about the actual context world may be far from the truth, and then the a posteriori necessities may be very different; no one can claim certainty about this. But there is at least hope that the context world we live in is so nice as to allow us to stick to the claims 5 and 6 and thus to be metaphysically conservative and parsimonious, even though the epistemological picture I have drawn is much richer.

---

<sup>53</sup>In having this metaphysical position concerning colors I thus join what Hilbert (1987) calls anthropocentric realism. My only disagreement is that I would insist that metamers have the same color, because metamers look to have the same color under normal conditions (bright daylight etc.) and because the real color of an object shows itself only under normal conditions. If a color is thus constituted by the class of its metamers, the class *R* in assertion (3) is, for all we know, bound to be a wild one. To say, as Hilbert (1987, pp. 83f.) does, that only isomers have the same color would mean, I think, to revise ordinary color talk. Whether the revision would be a reasonable one is another question.

<sup>54</sup>Cf. my remarks about necessary and a priori relationality in my discussion of Claim 4.