

[[Comments are most welcome! I would much appreciate if you happen to read this draft and let me know whether you find my claim and the case for it convincing — even a few sentences would suffice!]]

## Are Phenomenal Zombies Really Conceivable?

Murat Aydede

Department of Philosophy  
University of British Columbia  
maydede@interchange.ubc.ca

[[*Introduction...*]]

- I -

Zombies, as conceived by philosophers these days, are creatures that are physically indistinguishable from normal people that nevertheless completely lack phenomenal consciousness. The kind of zombie I want to focus on is one that is molecule-by-molecule identical to a healthy, normal, adult human being living in a world physically like ours — indeed this might be our own actual world. To make things more concrete, pick any such person that you actually know. Let this be John. John is not a zombie.

Now consider an exact physical replica of John, call him Zhon. Note that because John and Zhon are physically alike, they are also behaviorally and functionally alike. So if you were to encounter Zhon, you could not distinguish him from John. Under the imagined circumstances so far, you would normally expect Zhon to be conscious as well. But let's stipulate that Zhon has no conscious experiences whatsoever — he's never had them, nor will he ever have them. So Zhon, according to this stipulation, doesn't know — indeed cannot know — what it is like to have conscious experiences of any kind. There is nothing it is like to be Zhon. Zhon lacks conscious phenomenology altogether. If Zhon were to be a metaphysically possible creature, he would be a zombie.<sup>1</sup> So this is the notion of zombie I would like to focus on.

According to many philosophers, Zhon *is* a possible creature, and that is because Zhon is a conceivable creature. This gives us the argument from zombies against physicalism. Physicalism is the doctrine that says: all that exist is physical through and through, including conscious minds and their conscious experiences. The zombie argument, as we might call it, is a species of conceivability arguments:

1. If Zhon is conceivable, then Zhon is possible.
2. Zhon is conceivable.
3. Hence, Zhon is possible.

---

<sup>1</sup> Zhon is a phenomenal zombie since he lacks conscious experiences. Whether Zhon could lack conscious thoughts is left open for the purposes of this paper.

Now since the choice of Zhon was arbitrary, we can, of course, generalize the argument to all zombies *like* Zhon — that is, to zombies that are physically indistinguishable, in relevant respects, to healthy, normal, adult human beings. As far as I can tell no one actually thinks that zombies can exist in our world: they are allowed to be nomologically impossible. But why does the mere metaphysical possibility of zombies count against physicalism? It is easy to see why. Consider John, an actual human being with conscious experience, and Zhon, a phenomenal zombie. They are physically type-identical. But despite this, if we grant the possibility that Zhon lacks consciousness whereas John has it, we are in fact saying that the physical make-up of John doesn't necessitate his having consciousness. In other words, we are asserting that John's consciousness does not metaphysically supervene on his actual physical constitution (and its physical history).<sup>2</sup> Zhon's metaphysical possibility, therefore, implies something actual about John, namely that John's consciousness is actually over and above his physical constitution. But if this is true, then not all that exist is physical. Thus physicalism is refuted. So, zombies are serious business. Accordingly, no physicalist can accept the metaphysical possibility of zombies like Zhon. But since the zombie argument is valid, a physicalist must deny one or both of its premises.

I am a physicalist. I deny both. I have argued against the first premise elsewhere.<sup>3</sup> Here I want to argue against the second premise. There are at least two ways of arguing against the conceivability of zombies. One is to show that zombie scenarios are inherently inconsistent. This is the surest way if successful. My argument will not take this form. For reasons I have elaborated elsewhere I grant that there is no formal contradiction in the idea of a phenomenal zombie — even when combined with the full information-theoretic picture I will sketch later. So in this sense, and only in this sense, I grant that phenomenal zombies are conceivable — following Chalmers 2002 we may say that they are negatively conceivable. But of course, it is the positive conceivability of zombies that is the intuitive driving force behind the conceivability arguments against physicalism. The notion of having a positive conception of a situation will have to remain at an intuitive level for present purposes. But positive conceivability requires more than a mere lack of formal inconsistency. Intuitively, to say that P is positively conceivable is to say that one can imagine a situation in some detail in which P is revealed to be true. This imagination need not be purely perceptual: it would ideally involve the resources of one's conceptual system — enriching, disambiguating, and interpreting the imagined situation as necessary. Furthermore, positive conceivability must be ideal. Ideal conceivability is, as Chalmers puts it “conceivability on ideal rational reflection.”<sup>4</sup> Intuitively, if P is ideally positively conceivable, then any arbitrary details of a situation in which P is revealed to be true can be further elaborated without inconsistency. Hence I will read premise (2) as stating that

---

<sup>2</sup> I won't worry in this paper about global vs. local supervenience. We may assume whatever global features (history, environment, etc.) of our world that consciousness necessitates according to a physicalist who rejects the possibility of zombies. Likewise, we may assume that a zombie will be a creature with all the non-local physical properties (including historical properties) deemed for consciousness.

<sup>3</sup> See Aydede & Guzeldere (2005) where we argue that the conceivability of *phenomenal zombies* doesn't entail their possibility — for special reasons. We remain neutral on the more general claim: for any P, if P is ideally conceivable, then P is possible.

<sup>4</sup> Chalmers (1999:477). For more discussion of various forms of conceivability, see Yablo (1993) and Chalmers (2002).

Zhon is ideally positively conceivable (assume henceforth the relevant notion is ideal positive conceivability — I'll just use 'conceivable' most of the time).

I will argue that phenomenal zombies such as Zhon are not conceivable. Alas, arguments of this sort tend to be non-demonstrative. But I believe my case is the strongest in the literature. It is also the most direct: I will describe the relevant physical aspects of Zhon in sufficient detail so that when you keep this description fully in mind, you won't be able to conceive Zhon as lacking phenomenal consciousness, or so I will claim... or so I hope.

Before I begin, let me make two further points. First, I've said "I will describe the relevant physical aspects of Zhon." My description, however, won't be at the physical level as ordinarily understood. There are different levels of descriptions of an organism — indeed of anything. We can describe an organism at the level of subatomic particles, or at the level of atoms and molecules, or at the level of functionally discrete packets of molecules and their organization, or at the level of molecular biology, or physiology, etc. The differences between the levels may not be neat, but that there are different levels of description seems obvious. As you go up in the hierarchy of levels, you lose details but gain generality, and regularities not visible from lower levels start to appear. The higher levels supervene on lower levels in such a way that, at least in principle, the higher levels are explainable by or reducible to lower levels. For present purposes, what matters is that in this hierarchical system inter-level relations are, intuitively, not metaphysically problematic or puzzling, at least not in the same way in which phenomenal consciousness is thought to be problematic in relation to any physical level below it. Let's call *any* level that metaphysically supervenes on a (purely and entirely) physical level that is suitably low in the hierarchy of levels, a *naturalistic level*. On this understanding, the zombie argument amounts to the claim that no naturalistic level entails the existence of phenomenal consciousness. Premise (2) of the Zombie Argument (that Zhon is positively conceivable) implies that there is no naturalistic level whose entertainment, however detailed it might be, will rationally compel an ideally rational (and phenomenally conscious) thinker to attribute phenomenal consciousness to Zhon.<sup>5</sup>

---

<sup>5</sup> The intended claim is:  $\forall n \forall t [E(t,n) \rightarrow \sim A(t,zhon)]$ , where 'n' ranges over naturalistic levels involving Zhon, 't' ranges over ideally rational and phenomenally conscious thinkers. E(t,n): t entertains the full canonical description of n. A(t,zhon): t finds it rationally compelling to attribute phenomenal consciousness to Zhon. This claim may be too strong in the light of the fact that we don't quite know the extension of 'ideally rational thinker'. Formally it is falsified if there is one such thinker entertaining the relevant naturalistic description who finds it rationally compelling to attribute consciousness to Zhon. If we don't make strong demands on what it takes to be an ideally rational thinker, falsifying the claim will be easy. So I would take myself to have succeeded if most philosophers (qua ideal rational thinkers) who entertain the description I'll provide below will find it rationally compelling to attribute consciousness to Zhon. Given that my argument is not meant to be apodictic, I will be satisfied even with a weaker result: most philosophers who entertain the description I'll provide will find themselves incapable of imagining that Zhon lacks a conscious phenomenology.

Compare [[elaborate this further]]: "... possession of a concept such as 'knowledge' or 'water' bestows a conditional ability to identify the concept's extension under a hypothetical epistemic possibility, given sufficient information about that epistemic possibility and sufficient reasoning. That is, possession of these concepts in a sufficiently rational subject bestows an ability to evaluate certain conditionals of the form  $E \rightarrow C$ , where E contains sufficient information about an epistemic possibility and where C is a statement using the concept and characterizing its extension, for arbitrary epistemic possibilities. And conceptual analysis often proceeds precisely by evaluating conditionals like these." Chalmers & Jackson (2001:??).

Most claims about the conceivability of zombies are made with a low level physical description in mind; for instance, with a stipulation that goes like: “Let P be a complete *microphysical* description of John” (or Zhon, or our world, etc. — depending on the argument’s setup). [[Refs]] Where the microphysical level is presumably meant to be subatomic, atomic, or molecular level. It may even be at the cellular or physiological level. But hardly above that. We are then asked to imagine to hold that level in mind and deny phenomenal consciousness. But this is cheating. How are we supposed to entertain a description at the molecular level, with zillion complex molecules all put together in zillion different ways, and functioning as a dynamic system in incredibly complex ways. Presumably the idea in pumping the relevant zombie-friendly intuitions is that the sheer complexity won’t matter: if you can entertain, say, the incredibly complex molecular structure of even a single protein, you can entertain them all — more of the same, all put together in all these complex ways, etc. (remember all ideal conceivability requires is that any arbitrary detail can be further described without generating an inconsistency). Now ask yourself how phenomenal consciousness is supposed to emerge from this mind-boggling molecular mess. Magic! So keep this molecular jungle in mind and deny consciousness to it. So easy!

The mistake isn’t that it is impossible to entertain such a complexity at the microphysical level — although this is certainly true. Rather, the mistake is that the microphysical level is the wrong level with respect to which we are supposed to carry out the thought experiment about zombies. The right level is the highest naturalistic level with sufficient detail. At this level, complete detail would certainly help but is not required. In what follows I will provide an information theoretic description of what is going on when we consciously sense our environment. There is no one level here, but a mixture of descriptive levels that are all naturalistically kosher. I will claim that if you keep the provided naturalistic description of a normal adult human being firmly in mind, you will find it impossible (or at least extremely difficult) to deny phenomenal consciousness to him. Put differently, you will find yourself incapable of intellectually moving from the actuality of John to the possibility of Zhon — or so I hope...

This brings me to the second point I would like to make before I begin this task. It is important, for my argumentative purposes, that it is Zhon that we’ll find it impossible to conceive. Remember, Zhon is a complete physical replica of John who is in fact a fully mature normal healthy human being that you know well. The zombie argument is cast in terms of Zhon, *not* in terms of animals who we suspect to be conscious, or martians who are supposed to be conscious in ways somewhat similar to us, or robots that seem capable of engaging in actions similar to ours, or most importantly, actual or imaginary human beings who are abnormal, or otherwise informationally/functionally *not* equivalent to normal adult human beings functioning under normal circumstances (e.g., blindsighters, or Block’s superblindsighters 1995). This is important for at least two reasons. First, the anti-physicalist defenders of the Zombie Argument cast their argument, as a matter of fact, to make a case primarily for creatures like Zhon whose “sensory-perceptual” systems are *informationally* type-identical to John’s. I will ask no less: imagine Zhon with exactly such an informational capacity. Second, the only noncontroversial conception of phenomenal consciousness is derived from our own case — it is a first-person conception. When someone claims that Zhon is conceivable, he is claiming that Zhon lacks *this* sort of mental life, where the demonstrative is supposed to pick out the relevant first-person aspects of one’s own conscious experiences. At least for the polemical purposes of this paper, whether we can start to build, intuitively, a first person conception of what it is like to

completely lack phenomenal consciousness is irrelevant.<sup>6</sup> As allowed by the defenders of the Zombie Argument, I want to focus on *us* (e.g., John), and *our* exact physical replicas (e.g., Zhon, hence on our exact informational equivalents whose informational travails are accomplished by physically/functionally type-identical mechanisms).

In the following section, I will describe, in a sketchy way, an imaginary sense-modality, voluvision, and an actual one, color vision, from an information theoretic perspective with sufficient relevant detail and will ask you to generalize it in obvious ways to all other sensory modalities that we happen to possess and will ask you to keep this description firmly in mind. I will then raise the question whether you can conceive Zhon, described in such terms, as lacking phenomenal consciousness.<sup>7</sup> Along the way, I will examine some of the implications of the ideas developed during this exercise. I will briefly touch upon what I take to be essential to have a sensory modality that gives rise to conscious experiences, and if we have time, suggest ways in which some further philosophical puzzles about experience can be resolved given this framework.

- II -

The spokesperson of the Cyborgs Corporation announces the availability of a breakthrough new technology. She says it is the first brain-implantable product of a larger project for developing cybernetic organisms (cyborgs) with new and enhanced sensory capabilities. On the screen we see a device fitted on the forehead of a cyborg that has hardwired connections to the brain on several points on the skull. The spokesperson calls the device “voluvisor.” According to the briefing, the voluvisor provides sensory information about the 3-dimensional volumetric structure of one’s field of voluvision, which is a sensory field roughly coinciding with the width and height of the visual field of a normal adult human but with only 30 to 50 yards of depth sensitivity depending on the volumetric conditions. The voluvisor uses a combination of various types of advanced sonar and short range radar technologies as well as passive detectors for multiple portions of the electromagnetic spectrum, and is fully integrated to the user’s cognitive and sensory motor systems through brain implants. In other words, it does not piggyback on the genetically traditional sensory organs and systems of the brain: it provides an independent sensory modality with its own systems of neural networks set up in ways similar to the standard sensory neural networks of the natural brain.

Roughly, what the voluvisor does is to provide information for any two discriminable points (atomic volumes) in one’s 3D voluvisual field whether or not they are filled with the same or similar “matter” according to a fixed multidimensional similarity matrix — without giving any (extractable) information about what, if any, the filling material is. It is, basically, a multidimensional tomographic device. Furthermore, it does this in real time preserving the egocentric 3D coordinates of the field, in a way similar to standard vision that gives us surface

---

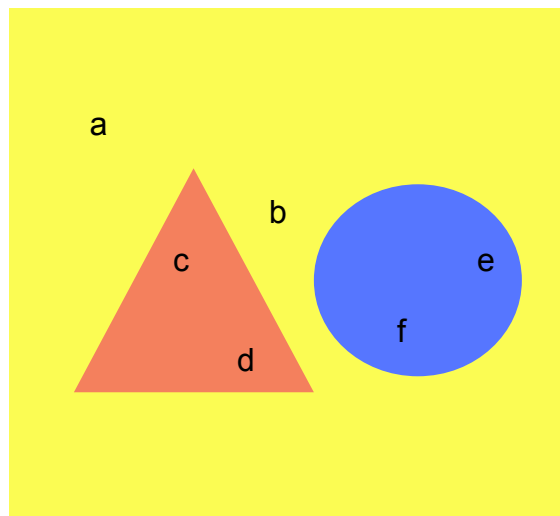
<sup>6</sup> Here I have in mind attempts like Ned Block’s superblindsighters (1995), or Kirk’s or Searle’s thought experiments about someone who starts to lose sensory modalities slowly without losing too much functionality as imagined from a first-person perspective [[refs]].

<sup>7</sup> In what follows, Zhon will mostly drop out of the picture. The point of introducing the zombie idea through Zhon was to make vivid what/who we are asked to conceive: it is a creature with fully (informationally) functional sensory modalities and a “cognitive” architecture exactly like that of a normal adult human being.

information of visible objects in real time. In other words, at any giving moment, the voluvisor provides similarity information between any two or three points in one's voluvisual field and the coordinate information changes as the spatial relation of the perceiver changes with respect to these points.

According to the spokesperson, the novelty in this technology came with the breakthrough in the development of information exchange interfaces between the brain's neural systems and the microchip technology. The sonar and radar (along with other detection) technologies involved in the voluvisor were not themselves all that new. We could already build systems that could give us this information on a screen with color coded imagery, or even better, in dynamic 3-D holographic displays with sound effects. But we had to process this information through standard vision and hearing. But even then, we couldn't get the full tomographic information through these displays for obvious reasons. The challenge was to build a cybernetic device that could give us the full volumetric information in sensory form to be consumed both by motor systems in the direct guidance of behavior and by higher-level cognitive systems. The cognitive systems would use this information to build conceptual representations of the voluvisual field for learning and planning, much like the way we consume the information provided by our regular sensory modalities, especially vision. So a quick and rough comparison with vision will help.

Consider the following picture.



[[ *Note for the non-color printout:* The triangle is colored solid orange, the circle solid blue, and the background is yellow.]]

The letters represent whatever physical property is detected in color vision in each arbitrarily chosen discriminable points on the picture. Your visual experience gives you the following information (among others):

$a=b$ ,  $c=d$ ,  $e=f$ ,  $a \neq c$ ,  $a \neq f$ , and so forth.

Intended reading: the just discriminable region named *a* instantiates the same property (or: just about the same property) as the just discriminable region named *b*, and so forth. In other words,

your experience gives you all the sameness/difference information about *any* two discriminable points. Furthermore, in some intuitive sense, it tells you, for instance, that *a* is more similar to *c* than to *e*. We get this kind of similarity information for *any* three points, and the sameness/difference information, in the limit, for *any* two points. Let's call this sameness/difference/similarity information between any two or three points "differential information." Note that this is a *lot* of information. Obviously under fixed conditions, the resolution of our visual experience is finite. Still, the differential information is about *any* two or three (just discriminable) points.

The notion of information I am using here comes from the formal communication theory developed primarily by Shannon and Weaver (1948, 1949) and later appropriated by Dretske (1981) and others into a form designed to deal with informational *content* of signals. We can characterize this notion as follows:

A physical event (signal) *s* carries the *information* that *a* ( $\neq s$ ) is *F* (*a* being the source) =<sub>def.</sub> Necessarily, *a* is *F* if *s* occurs and certain appropriate channel (background) conditions obtain, but it is not necessary that *a* is *F* if only these channel conditions obtain without the signal. (Cf. Dretske 1981:??)

Do we get any further (extractable) information about the nature of the physical properties thus detected in color vision? It appears not. Simplifying things to hues only (ignoring luminance and saturation, among other parameters),<sup>8</sup> it is plausible to say that color vision is a way of getting differential information of a set of surface properties without getting useable information about what those properties are. It says something like: whatever detected surface properties are instantiated at any two points they are the same or different, or similar with respect to a third point. But nested in this information about the picture above, there is more interesting and useful information about shapes. For instance, a complete specification of differential information about any three points will imply that there is a circular shape and a triangular shape on the painted surface, and that they are spatially related to each other in a certain way. Call the information implied by the differential information *nested information*.<sup>9</sup> Ordinarily we are interested in the information nested in the differential information — this is the information about the familiar world of figures, sizes, shapes, relative positions, objects, events, people, etc. Vision as a *perceptual* mechanism (as opposed to a *sensory* modality) has many complicated built in algorithms to extract this nested information. Consequently, our visual world is not the world of mere surfaces presented to us in the form of differential information; rather it is the world of shapes, objects, events of various kinds. This is perception. Nevertheless the differential information is the sensory fabric of visual perception.<sup>10</sup>

---

<sup>8</sup> For instance, I am ignoring depth information, surround and contrast effects, color constancy. These phenomena complicate things considerably. There is also the question of whether the phenomenal similarity can be mapped onto an objective similarity matrix. But I believe the basic picture that will emerge is basically right.

<sup>9</sup> The terminology is Dretske's (1981).

<sup>10</sup> This is not meant to be a claim about how visual perception actually works. Rather it is a claim about the mode the world is perceptually presented to a perceiver: the perceived world of objects and events is presented as embedded in differential information.

But the differential information about hues is only one dimension of a three dimensional discriminability space in vision.<sup>11</sup> At every surface point, normal color vision gives information about hues, their saturation, and brightness somehow synthetically fused into a unity with three coordinates in the discriminability space. We can still discriminate between any two points with, say, exactly the same hue and brightness but with different saturation (similarly with respect to other combinations). In fact, we discover how many dimensions this space has by analyzing what discriminations we can make (through multi-dimensional scaling experiments — see Clark 1996). Compare the synthetic fusion of color vision to the analytic combination of musical chords — sounds at different frequencies simultaneously heard.

Our visual system can give us this differential information not only for *simultaneously* presented points in an expanse, but also for *diachronically* presented points across scenes. What kind of a detection system is capable of doing that? Think of this question as an engineering problem that the physiology of our visual system has managed to solve. This ability requires a way of *positively* categorizing or identifying the synchronically and diachronically different instances of a property (or properties) as the same or different. From an engineering perspective, one way (perhaps the most obvious way) to do this is to use the same physical label (response) or the same value of a continuously varying physical parameter (at least for each discriminability dimension) for the different instances of each property detected. It is likely that these values are the attractors (maxima/minima) of a multidimensional vector space realized in a neural network in the brain (as Paul Churchland has been claiming for a while [[refs]]).

From a physical perspective, the detected surface properties may be quite complex or heterogeneous whose physically type-distinct instances forming an equivalence class inducing the same detector response. Indeed this seems to be the case with the Surface Spectral Reflectances (SSRs) detected in color vision: SSRs corresponding<sup>12</sup> to colors constitute disjunctive sets (mostly because of metameric matches — see Appendix 2). The SSR of a surface is determined by detailing what percentage of visible light is reflected at each wavelength. The SSR of a surface is not dependent on the nature of the ambient light or on the nature of the observer's visual apparatus. The molecular structure of a visually discriminable surface point will absorb and therefore reflect the incident light in different quantities at each wavelength. Thus, one way to detect whether two points on a surface has the same or different molecular structure at a given time under the same lighting conditions is by devising a way of profiling these reflectances. This could be done in a number of ways — some are more reliable than others. The way our visual system does this, simplifying quite a bit, is by computing the value of a function whose arguments are responses of three types of transducers (cones in the retina). These cones respond to the same wavelength (with the same intensity) in characteristically different ways. The output values of the function are predetermined and form a continuum that can be represented, in the case of hues, in a two dimensional discriminability

---

<sup>11</sup> This was first proposed by A. H. Munsell in 1905, whose representational scheme has come to be known as the Munsell color solid (see Appendix 1). See the appendix for illustrations. In what follows, it is useful to keep in mind that the term 'space' is used for both the three-dimensional *physical* space one's visual or voluvisual mechanisms report about (their field), and for the abstract multi-dimensional *discriminability* space whose dimensions are fixed according to what and how many physical parameters are detected in this physical space. The discriminability space can vary widely from one sensory modality to another.

<sup>12</sup> I'll just use the term 'corresponding' to remain neutral about the ontology of colors. According to primary quality theories of colors (such as Hilbert's 1987), colors just *are* (sets of) SSRs. See below.



space with four primaries (corresponding to four unique hues).<sup>13</sup> So every point in a surface reflecting detectible light appears to have one of these values. It turns out that this way of doing it creates equivalence classes of different molecular structures being registered as the same; and identical molecular structures being classified as different under different lighting conditions. For instance, (as Austen Clark says somewhere) registering two surface points under the same illumination as different will almost always correspond to detecting real differences in SSRs. In other words, the informational value of registering two points as different is a lot greater than the informational value of registering two surface points as the same.<sup>14</sup> The upshot of such a system is to register the differential information between any two or three points in one's visual field, and this is normally sufficient to give us the information we are really interested in (ordinarily); namely, the information about lines, edges, corners, curves, boundaries, surfaces, shapes, volumes, objects, events, etc. This information is nested in the differential information about points on a three dimensional *surface* grid in physical space (basically a 2 1/2D sketch).

Like every actual and possible sensory detection system, voluvision works on the same principles. But it gives differential information about points in a three dimensional *volume* grid. It does this by registering similarity relations between any three points in one's voluvisual field according to a five dimensional discriminability space whose axes mark roughly the density, viscosity, temperature, "transparency" (brightness/intensity), and the "material kind" filling the atomic volumes (let's assume it is delivered synthetically)<sup>15</sup>. The "kind" dimension has six primaries corresponding to the response characteristics of the bouncing sonar and radar beams interacting with the atomic structures and the molecular bondings of the medium filling the atomic volumes. Some of the voluvisual kinds approximate folk kinds such as metallic, wooden, glassy, misty, void/empty, etc. The "transparency" (= brightness/intensity) information is information about the penetrability of sonar and radar beams for media after approximately 30 yards of depth in the voluvisual field. After roughly 30 yards, different material filling the volume starts to resist the penetration of the increasingly weaker beams to different degrees and become more or less opaque accordingly.

This information about any three physical points (atomic volumes) in one's actual voluvisual field is available *only* as *differential* information along the values of the five parameters in the five-dimensional discriminability space (as can be confirmed by multi-dimensional scaling experiments done on voluvision capable subjects). The information delivered doesn't in fact tell you whether you are discriminating temperatures, densities, different chemical structures, etc., under these descriptions. These descriptions are *theoretical* from the sensory voluvisual perspective, so to speak, just as the description "surface spectral reflectance" is theoretical (third-person) from the first-person perspective of "color" vision. But, again, nested in this differential information is the much more interesting and useful information about lines, edges, corners, curves, boundaries, surfaces, shapes, volumes, objects, events,

---

<sup>13</sup> This would be a plane perpendicular to the main vertical axe of the color solid. See Appendix.

<sup>14</sup> See Clark (19XX).

<sup>15</sup> This is optional. I've opted for it to make it more like color vision. But they need not be synthetically fused. Some or all might be analytically presented. In hearing, for instance, although simultaneous sounds at different frequencies can be heard as distinct sounds (hence, analytically), the auditory information about the amplitude of a sound is nevertheless synthetically fused with the information about its pitch. Discriminability doesn't require analytic information delivery.

people, and so on. Proper algorithms to extract this nested information is built, at least partly, into the voluvisor; the rest of the categorization is done by the conceptual system in the brain.

Unlike the normally sighted people, someone with voluvision is literally capable of (volu)seeing the whole objects and their interior make-up and structure all at once, but from a spatial perspective. For instance, by just voluseeing an apple one can come to know whether it is rotten inside or whether it is bitten on the “other” side. A physically constructed maze of an appropriate size and proportion might prove challenging for the normally sighted. But someone capable of voluvision can solve it by just scanning around for a few seconds. On the other hand, like in normal vision, one can focus one’s voluvision on subregions of one’s voluvisual field, and increase or decrease its resolution. So, normally, peripheral voluvision is less detailed (peripheral in three spatial dimensions including the depth relative to the voluvisor’s vantage point).

With only normal sight, you cannot tell whether this carafe contains any liquid; or if it does, whether it is wine or water, or whether the carafe is full or half-full, etc.



With voluvision, you can. Not only that, you can also tell what is behind:



Suppose you are equipped with voluvision and taking a focused look at the carafes from an angle so that the normal sight cannot see the ones behind the first. Let's assume your focus has a depth of 3-4 yards centering your voluvisual field — call this volume the *focus cone* — so that the three carafes are wholly within it. Under the circumstances, you are getting all the differential information about the aforementioned five physical parameters between any three points in your cone. Let's call these points *atomic volumes* at the maximal resolution of your scanners. These parameters are, to repeat, density, viscosity, temperature, brightness (transparency), and material kind, making up the axes of the five dimensional discriminability space synthetically fused. At every atomic volume, you are detecting the values of these five physical parameters and are capable of making reliable discriminations between these volumes as to whether they are the same/different/similar (synchronously and diachronically) with respect to these parameters. By getting this differential information, you are, of course, also getting the nested information about various boundaries, shapes, volumes, structures, objects, events, and so on.

From an information-theoretic perspective, there is nothing mysterious about getting only differential information about certain physical parameters without getting extractable information about what the nature of these parameters are that are being compared for similarity. Simple signals can carry information about quite complex properties or events without making the information of this complexity extractable, i.e., available for further exploitation by the consumer systems downstream. To illustrate, let me elaborate an example that Dretske uses.<sup>16</sup> It is possible to carry *all* the information encoded by a picture of a scene with a simple/primitive signal, say a buzzer system. Suppose you have got a digital camera taking pictures of the traffic at a busy road intersection. The informational value of each picture taken is determined, roughly, by how many possibilities eliminated at the source, the road intersection. Suppose a certain picture is the picture of a certain scene at the intersection. What is the likelihood of that exact scene (at least from camera's perspective) happening again? Very low indeed. So the informational value of that picture, the amount of information it carries, is extremely high. Now

---

<sup>16</sup> Dretske (1981:138–39).

think of the buzzer system's set up: it is setup in such a way that it will buzz when and only when the exact same scene depicted in the picture occurs. Actually, as Dretske notes, the computer recognition programs that rely on whole-template matching procedures approximate this kind of transition from one form of coding to another. But, note, both structures will carry exactly the same information (thus exactly the same amount of information). However, intuitively, the buzzer's buzzing carries the information carried by the picture in a way that is not extractable, whereas the picture carries it, intuitively, in an extractable form.<sup>17</sup> The notion of differential information implies that no further information about the nature of each property detected at points on a surface is available in an extractable form to consumer systems downstream. What is extractable by the consumer systems downstream is what is nested in the differential information about these surface properties. There comes a point your visual system's most specific information about spacetime points bottoms out as nothing but differential information.<sup>18</sup>

This distinction between information being presented in extractable vs non-extractable form is absolutely crucial for understanding sensory modalities as information entry venues. In fact, although I have no space to elaborate here, I can make a stronger claim: every nomologically possible sensory modality *must* utilize differential information in the service it provides to the organism. No organism, or a machine for that matter, unless it has God's powers, can receive sensory information from its environment all in extractable form down to fundamental particles and properties. The point at which a sensory system registers (chunks) the information about various physical magnitudes only as the same/different/similar determines the point at which the information is carried in a non-extractable form. This is the most specific information that the signal can carry about the environmental properties it detects (Dretske 1981 says that the signal carries this information in digital form and the information nested in it in analog form).<sup>19</sup>

---

<sup>17</sup> For further discussion, see Kulvicki (2004), and Aydede & Guzeldere (2005).

<sup>18</sup> It might seem strange that a visual percept can detect instances of two surface properties in such a way that will make it possible to classify these instances as instances of the same property or different properties without telling anything about what these properties are. If this were true, it would indeed be strange. But it isn't. The visual percept does carry information about what these properties are whose instances are being compared. They are SSRs and this information is delivered by the visual percept. But it is delivered in a non-extractable format. We extract this information only with the help of relevant scientific theories. When the relevant parts of psychophysics and neurophysiology become part of the background channel conditions. See the next fn.

<sup>19</sup> [[To be elaborated elsewhere: The notion of differential information is a complicated one for important reasons, and the picture I have provided here in this paper isn't entirely correct, and in some ways even misleading. Consider the information you get about surface properties, a and b (in the above picture). If there are elements in your experience, a\* and b\*, that carry information about a and b, then intuitively the most specific information your experience has about a and b can be expressed as:

- a\* carries the information that a is yellow7 (the SRR set corresponding to yellow7)
- b\* carries the information that b is yellow7 (the SRR set corresponding to yellow7)

This is not differential information. The differential information that a=b you get from your experience is information that is therefore extracted *out of* this more specific non-differential information. This has important implications for a variety of issues such as whether the differential information implies the awareness of elements internal to one's experience (or its implementation in the brain), and the similarity matrix provided by a quality space vis-à-vis the informational/representational content of this quality

The notion of differential information also provides a principled way of determining the set of secondary qualities for any sensory modality: just determine (by doing empirical science — psychophysics basically) *what* physical properties or quantities are presented in the form, and only in the form, of differential information in that modality. These will be, or will determine with additional premises, the secondary qualities associated with that modality.<sup>20</sup>

So much for science fiction.<sup>21</sup> The description I have provided so far of how voluision is supposed to work as an independent sensory modality is almost entirely from a third-person perspective in the mixed language of psychophysics, information-theory and a bit of engineering (so is the description of “color” vision — more or less). I take this description and its further elaboration to be naturalistically kosher. The burning question is: what about the first-person qualitative phenomenology? What does it feel like to engage in voluision? This I cannot tell you, never tried it myself. But there is no problem in this since I cannot tell you what it is like to detect SSRs through normal vision either — something I’ve experienced extensively.

But there is a more fundamental question: how can we be sure that there will be a distinctive qualitative phenomenology for voluision? Put another way, the question is whether it is positively conceivable that you, equipped with a voluvisor, are looking at the carafes above and getting all the differential information in real time in your focus cone, and yet it doesn’t feel like anything — “it is all dark inside”. At this point I would like to remind the reader that what is at issue is positive conceivability, which requires a bit more than a mere lack of formal inconsistency. To repeat, to say that P is positively conceivable is to say, intuitively, that one can imagine a situation in some detail in which P is revealed to be true.

So suppose you are equipped with a fully functional voluvisor and are looking at the carafes getting all the differential information in real time in your focus cone. You can answer questions about whether the carafes are full, or how full they are, or with what (among many

---

space (here the ongoing debate about what to make of color similarity judgments comes easily to mind), even the notion of similarity involved is problematic... But none of this is directly relevant for my polemical purposes in this paper.]]

<sup>20</sup> See Kulvicki (200?). [[ Further elaborate if you have time: Another benefit of this way of looking at the sensory experience is that it promises to address a puzzle as ancient as the mind-body problem itself (some think it is the same problem). Sellars especially is known to have made a heavy weather out of this — labeling the puzzle as the puzzle of “ultimate homogeneity” (of a pink ice-cube, say), known sometimes under the label, “the grain problem.” Our sensory experiences, color experiences in particular, seem to present us things with properties that are instantiated homogeneously through and through. How can we be *presented* with properties that are ultimately homogenous (grainless) in a world that is essentially particulate, that is, “grainy” all the way down to its fundamental particles and their properties? If this is a problem, we have solution in our hands with the notion of differential information... See my XXXX). ]]

<sup>21</sup> The voluvisor, although not actual, is nomologically/empirically possible. In fact we already have various stripped down versions of the imagined scanning and detection technologies in, for instance, increasingly more efficient and convenient medical scanning technologies (PET, MRI, OCI, X-ray, NS, Echo, among others) with computed tomography (CT), various airport security scanning technologies, ultrasonic imaging, sonar technologies in subterranean and deep sea cartography, and literally on and on... The real difficulty is whether we can ever invent technologies that will provide broadband communication interface between the brain’s own neural networks and the signal processing microchips of the scanners or detectors. This may prove to be technologically impossible due to the brain’s immense complexity. But there is nothing empirically or nomologically impossible in any of this.

other similar questions) with the same ease and confidence that you can answer the questions about whether they are all the same color, or which one has the same color as bananas', etc. When asked how you know these things, you answer "well, I volusee it (to be that way)". Moreover, the voluvisual information is available to you in a way that allows you to acquire new concepts: say, for some environmental reason, it becomes important for you to recognize media with a certain temperature, viscosity, and density of a certain volume, you call the instances "flurg". It is a perceptual kind for you just as blue circles are for the normally sighted. You can apply these concepts *directly* on the basis of your voluvision. When asked, how you know this is a flurg, you answer: "I volusee it (as one)". You appeal to voluvisual information to justify your actions, and sometimes explain and predict others'. You imagine and plan your movements by appeal to the information made available to you in voluvision, and so on.

And yet when asked whether voluseeing feels like anything, you answer:

*No, it 's all dark in there, there is nothing it is like to volusee! Yet I am not merely guessing whether the first carafe is half-full with water when I say so. Neither do I make an automated inference from the fact that in the past these kinds of thoughts popping up in my mind have proved to be perfectly reliable. No, I do have this information immediately available to me which I directly rely on for the epistemic, cognitive, and practical tasks you've just mentioned, but really it doesn't feel like anything — qualitatively it is as blank as it gets!*

Is this positively conceivable? You are getting the full differential information between any three points in your focus cone simultaneously along the five aforementioned parameters, and this information is immediately available to be consumed not only by your motor systems but also by your epistemic and cognitive faculties — some of which are mentioned above.<sup>22</sup> Yet the delivery of this information has no phenomenology attached to it.

Here is another way to see the mystery. Close your eyes while standing in front of a slowly changing but clearly visible scene, and try to positively imagine getting the full differential information about SSRs (and, of course, the information nested in it) in *informationally* the exact same way in which you would be getting it if your eyes were not

---

<sup>22</sup> There are two visual pathways in the brain (the dorsal and the ventral streams originating from the primary visual cortex) where the same sensory information, after its delivery, is used mostly but not entirely independently of each other. One of the functions of the dorsal stream is to use the incoming sensory information in what we might call mundane behavior in auto-pilot — e.g., as we navigate around objects, the fine-tuning of our behavior seems to be initiated and controlled before the relevant sensory information becomes conscious through the ventral stream. Still the auto-piloting system is under voluntary control (it can be interrupted or modified by conscious decision), which means that it is at least partly under the supervision of the other stream to some degree. Also, the dorsal stream has only a two-second window for auto-piloting, but it's not capable of conceptual learning, planning, and recognition/categorization (it is really nothing more than a two-second moving wall in the processing of the egocentric spatial information). It is plausible to think that voluvision would have a similar organization. So being in command of motor behavior in auto-pilot is not probably sufficient to make any information stream (and its motor use) phenomenally conscious. This is why the more epistemic and cognitive tasks are important in the explanation of why there has to be a phenomenology associated with certain forms of information delivery — what is cognitively done with that information is of crucial significance.

closed and you were looking at the scenery. This is what it is like to be a visual zombie, but being a voluvisual zombie is like this too.

When challenged how you know whether, say, the second carafe is full, you advert to your “experience” for epistemic warrant, namely to a voluvisual state of yours that either carries or purports to carry the relevant differential information and the information nested in it. Note that in this sense things may voluvisually seem to be a certain way even though things aren’t actually that way. You can have voluvisual “hallucinations.” These are states that purports to carry voluvisual information even though they don’t actually carry it — you have informational access to these states, and you can report and describe them. Consequently, you can describe how things voluvisually seem to you without committing yourself whether things are the way they seem, even when things are in fact the way they voluvisually seem to you. Hence you not only have direct access to differential information (when you do) but also to its actual or purported delivery. This is a form of “introspection” of your voluvisual state. How could you report and describe your voluvisual state in the absence of complete qualitative phenomenology? What is there to report? Informational content? But if the delivery and use of this barrage of differential information about any three points in your focus cone is phenomenologically blank, then it is a mystery how you can tell whether you are getting any information at all (purported or actual). Or, suppose, I ask you to *imagine* getting such and such voluvisual information, maybe about a would-be maze of a certain shape. I might ask you to tell me how many right turns you need to make from the start in order to exit the maze. If it is conceivable that the actual delivery of voluvisual differential information has no phenomenology, I suppose, it is equally conceivable that merely imagining the delivery of such differential information would have no phenomenology either. But then it is not at all clear how you can tell whether you’re imagining the delivery of such and such voluvisual information as opposed to so and so. Do you have any positive conception of how you can tell the difference?

Let me make the claim explicit (as applied to voluvision):

- (V) You<sup>23</sup> cannot (ideally positively) conceive that in any world in which
- someone, equipped with voluvision, is getting the full voluvisual differential information, and
  - that this information is available to be used by one’s more cognitive and epistemic faculties that are capable of extracting the information nested in this enormous barrage of differential information so that learning, planning, reasoning and acting on that information are made possible,
  - but that one has absolutely no qualitative phenomenology associated with voluvision (i.e., the voluvisual differential information is presented to one without any phenomenology whatsoever).

‘Someone’ ranges over people like John — fully developed mature healthy humans — or creatures with similar information processing capabilities. We can evaluate claim (V) from a first-person as well as a third-person perspective. In the latter case, the person in question is someone other than yourself, say, John. Then the claim is: you cannot conceive John having all

---

<sup>23</sup> Since you, the reader, is arbitrarily chosen, this claim can be put into “it is not conceivable that...”. But I prefer the form in the text since I want to initially focus on the first-person evaluation of this claim.

the voluvisual differential information and using it in the aforementioned ways all the while not having any associated voluvisual phenomenology. Since my case for this claim relies on the first-person instance of (V), let me turn to that. In this case, you yourself cannot conceive that you are getting all the voluvisual differential information and using it in the aforementioned ways all the while not having any associated voluvisual phenomenology.

Because what I am asking you to concede relies, partly but crucially, on how well you'll perform a psychological/cognitive task, it is important for me to keep reminding you the nature of differential information, its huge amount, and how it's used — remember you have direct and reliable access to this information in a way exactly parallel to the access you have to visual differential information and the ways in which you use this information. While vividly keeping all this in mind, you are asked to affirm that you have absolutely no associated phenomenology — the delivery of this information is phenomenologically “blank”. The crucial point here is to reflect what happens when you remove any phenomenology in your imagination: you remove information! I claim that you cannot imagine getting all this differential information in the absence of phenomenology: the moment you remove phenomenology you remove differential information.

A potential complaint here is that we shouldn't equate imagination with conception. There are things, such as a thousand-sided polygon, that we cannot imagine but can conceive. But this complaint isn't justified, for at least two reasons. First, the very concept of qualitative phenomenology we have is of something whose essence makes it essentially imaginable and first-person. My conceiving of qualitative phenomenology or the lack of it is bound to involve my imagination in the following sense: when I claim to positively conceive of a situation which has or lacks phenomenology, I claim to positively conceive of a situation which has or lacks essentially *this* sort of thing, where the demonstrative refers to something imagined or at least imaginable, if not to something being actually experienced (thus *a fortiori* imaginable).

The sortal in “this *sort* of thing” is also important, and this brings me to my second point against the complaint. Note that we do not in fact have a positive conception of the *particular* phenomenology that would be associated with voluvision. Voluvision is a hypothetical sensory modality. You don't know what it is like to volusense anything, neither do I. A fortiori we cannot even imagine the *particular* phenomenology supposed to be associated with voluvision. So when I invite you to entertain (V) to see whether you agree or not, I am not inviting you to imagine what it is like to volusense and deny *its* existence while getting all the voluvisual differential information. So the complaint is doubly unfounded.

But we do know what qualitative phenomenology is — phenomenology associated with sensory experiences. This knowledge comes from first-person acquaintance with these experiences. One of the reasons why I've started my discussion of zombies with a hypothetical sensory modality such as voluvision was to avoid relying on the particular phenomenology of sensory experiences we are familiar with on a first-person basis. Nevertheless, I asked you above to assess (V) from a first-person perspective. The reason is that you know what sensory phenomenology is and how it functions. In some sense, you can even imagine, in the abstract, so to speak, what it is like sense the world in a phenomenally conscious way without that way being any *particular* way. That is why it is perfectly intelligible to ask whether you can positively conceive of voluseeing, say, the cafes in your focus cone, without having any associated phenomenology, even when you don't know what it is like to volusee at all — if it is like anything. But the intelligibility of this question doesn't rely on your imaginative skills in any *particular* sensory phenomenology.



Note that we haven't even started discussing informationally equivalent complete zombies yet. The claim that (V) makes is about the conceivability of voluvision without any proper phenomenology in people like you and me, creatures whose sensory information processing is phenomenally conscious. But if I can secure my claim (V), turning to complete zombies will be very quick. So let me repeat (V) in the first-person mode:

(V<sub>fp</sub>) You yourself cannot positively conceive getting all the voluvisual differential information in your focus cone and using it in the ways mentioned previously all the while not having any voluvisual phenomenology whatsoever.

The reason you cannot is not that you don't and therefore cannot know what it is like to volusee. In fact, the reason is not directly related to voluseeing per se. Rather it concerns the intimate relationship between the notion of differential information (including its delivery and use in the way explained), and the notion of having a sensory phenomenology. The moment you attempt to remove phenomenology, in your imagination/conception, from your voluvisual state, you remove the differential information. Thus your attempt to positively conceive yourself to be a voluvisual zombie gets spoiled.

Primarily the same reason is in force if you assess (V) from a third-person perspective — if you take the subject of voluvision as someone other than yourself, say, John.

(V<sub>tp</sub>) You cannot positively conceive of John getting all the voluvisual differential information in his focus cone and using it in the ways mentioned previously all the while not having any voluvisual phenomenology whatsoever.

The moment you deny voluvisual phenomenology to John, you are conceding that you have absolutely no positive conception of how John is getting the voluvisual differential information.

Again, at this point, I would like to remind the reader that what is under discussion is positive conceivability, which is lot more demanding than merely affirming that there is no formal inconsistency in the idea of John being a voluvisual zombie.

...

Now that we have a rich enough description at the right level of abstraction of a hypothetical sensory modality, voluvision, and an actual one, vision, I would like you to kindly grant me for present purposes that we can generalize this sort of information-theoretic account of perception to all modalities. I think it should be intuitively clear how such an expansion of the account to other modalities would go. Having a positive conception of an informationally equivalent total phenomenal zombie requires one to vividly imagine

- that one is getting the *full differential information in all sensory modalities*, and
- that this information is available to be used by one's more cognitive and epistemic faculties that are capable of extracting the information nested in this enormous barrage of differential information so that learning, planning, reasoning and acting on that information are made possible,
- all the while positively removing any phenomenology from this creature's first person psychology.

But I have a hard time imagining this. I hope, you'll tell me, you do too...<sup>24</sup>

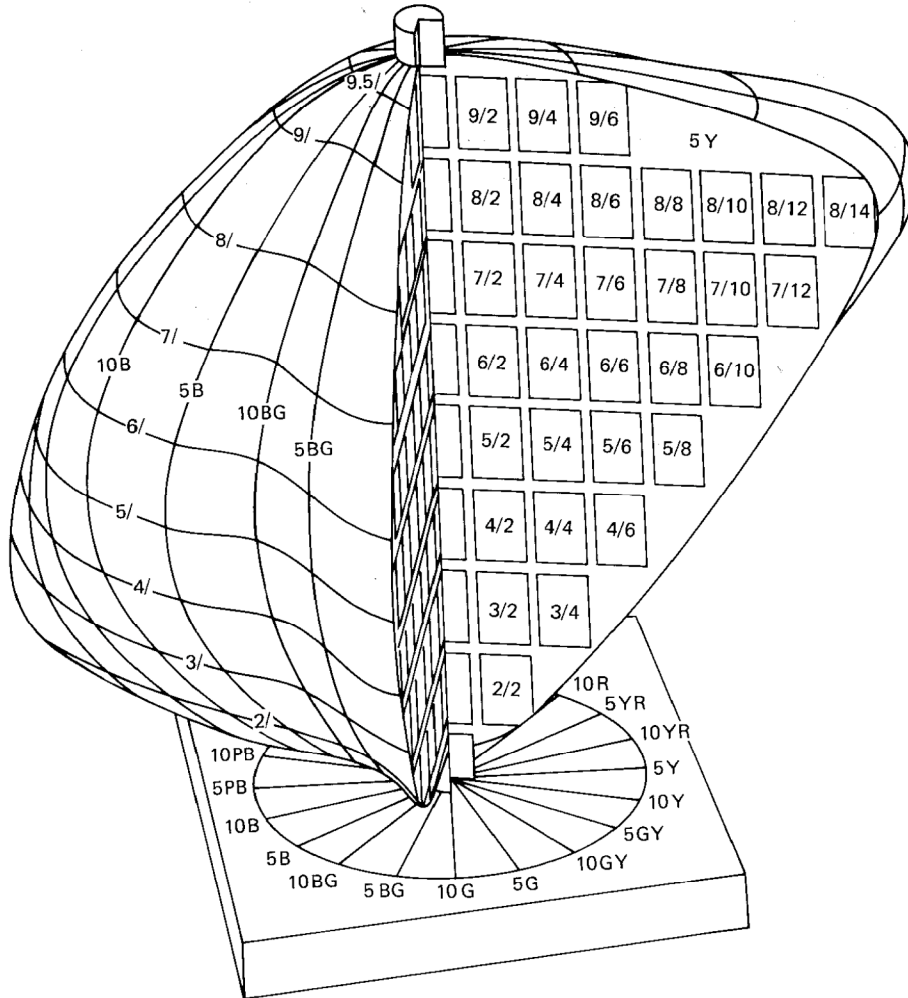
[[*References*]]

---

<sup>24</sup> For the record, I don't think that the considerations adduced to pump the intuition that informationally equivalent phenomenal zombies are inconceivable are effective for the claim that spectrum inversions are inconceivable. In fact, I think they are not only genuinely conceivable but also nomologically possible, and that the scientific picture is consistent with this being so (see Aydede & Guzeldere 2005). Furthermore, as I said before, I don't think that my efforts to pump intuitions for the inconceivability of informationally identical phenomenal zombies, if successful, amount to giving a demonstrative argument for it. I don't in fact know what a demonstrative argument to this effect would look like. To repeat: I grant that there is no formal contradiction in the idea of a phenomenal zombie even when combined with the full information-theoretic description along the lines I have tried to provide in this paper. Zombies are negatively conceivable. I hope that I have made a strong case for the claim that they are not positively conceivable. Nevertheless, the physicalist picture can be made more convincing if, on top of pumping these intuitions by giving a detailed information-theoretic account for what is going on in sensing the world, the physicalist can also give a physicalist account of why our first-person knowledge of phenomenology is bound to give rise to the illusory intuition that it cannot be reduced to anything told in completely third-person language. I think this can be done if we look at the *phenomenal concepts* from an information-theoretic point of view (see Aydede & Guzeldere 2005) — phenomenal concepts are the concepts through which we conceive (think, imagine, introspect) the phenomenal from a first-person perspective.

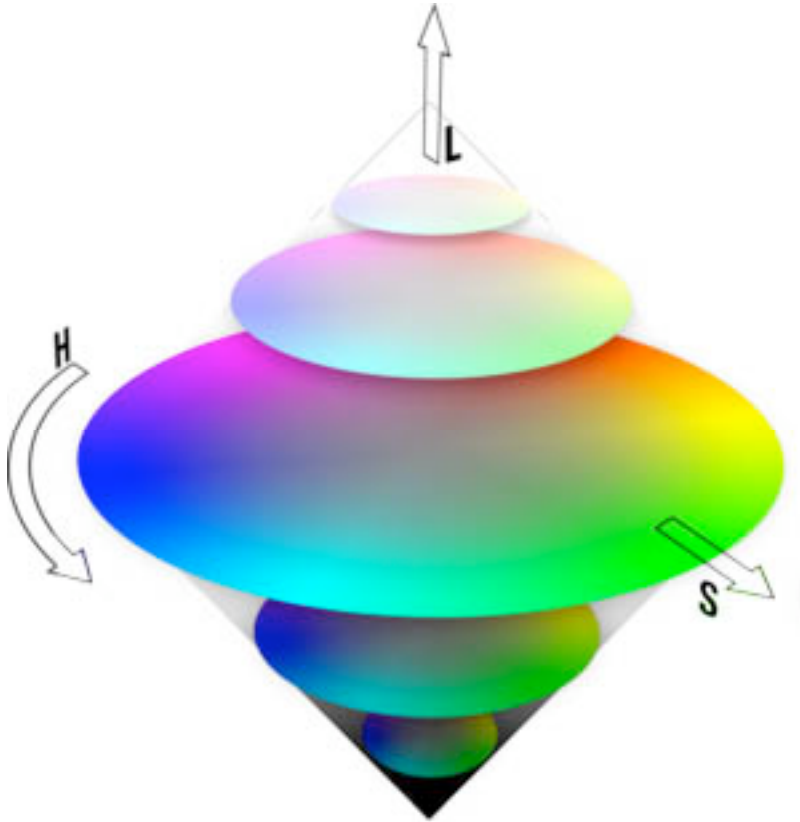
## APPENDIX 1:

The Munsell color system was devised by A. H. Munsell in 1905 and revised in 1943. It still remains popular in many industries and is standardized by many standards organizations (e.g., American National Standards Institute). The Munsell color system is a cylindrical coordinate system: A color wheel and a 2D grid around the core. The 3D shape is irregular.



## A more modern representation of the the “Color Quality Space”:

Credit: Alexandre Van de Sande (*Wikipedia*)



The **HLS** color space, also called **HSL** or **HSI**, stands for Hue, Saturation, Lightness/Brightness (also Luminance or Luminosity). While HSV (Hue, Saturation, Value) can be viewed graphically as a color cone or hexcone, HSL is drawn as a double cone or double hexcone. Both systems are non-linear deformations of the RGB color cube. The two apexes of the HSL double hexcone correspond to black and white. The angular parameter corresponds to hue, distance from the axis corresponds to saturation, and distance along the black-white axis corresponds to lightness. (Adapted from *Wikipedia*)

## APPENDIX 2

The inputs show two different spectra (metameric matches) that produce the same color experience (of purple) by producing the same ratios between the light absorptions of three retinal cone types.

