1

Quality Space

THERE is more to sentience than sensory qualities. Even if there comes a day when all the conceptual tangles wrapped around the qualities of sensory experience have been resolved, the job of understanding sentience would still be only half done. Lurking beyond the thickets of qualia lies another problem that is just as big and nasty. So even after those thickets are cleared, philosophers still could not rest; they would still have work to do. They would still have gainful employment. To make this case I need to explain (in this chapter) how one might come to understand the qualities of sensory experience, and then (in the next) show that even when that task is done there are aspects of sentience that remain untouched.

1.1 FOUR ASSAYS OF QUALITY

A *quality space* is an ordering of the qualities presented by a sensory modality in which relative similarities among those qualities are represented by their relative distances. Qualities that are relatively similar to one another are closer to one another than are qualities that are relatively less similar. The machinery required to derive such multidimensional orders is rather elaborate (see Clark 1993*b*), so it is easy to lose track of what the end-product will do for us, and why we might want some. Let us start then with a synopsis of the scope and limits of quality space.

The label itself turns out to be ambiguous. It helps to distinguish two different senses in which one might describe the 'qualities of experience'. As in so much of contemporary philosophy of mind, the distinction can be traced back to some prominent discussions by Wilfrid Sellars (1963: 93–4, 192–3), who noted the difference between ascribing a property to something experienced—something seen, felt, or heard, for example—and ascribing a property to the experiencing of it. Sellars urged that predicates such as 'red', 'square', 'loud', 'smooth', 'sweet', and 'cold' are in their initial and paradigm uses ascribed to the objects seen, heard, felt, tasted, or touched. The logical subjects of these predicates are, in the manifest image, physical objects, physically in front of the sense organs. A 'sensation of red' is first and foremost a sensation of an *object* that is red; 'red' characterizes the thing seen, not the seeing of it. But Sellars noted that once we have acquired the notion of a sense impression—which is not something seen, but instead a seeing of something—these predicates can acquire a second, distinct use.

Imagine, he suggests, that the notion of a sense impression is introduced as a theoretical term in an explicit theory. To do any work for us these hypothetical entities must be endowed with properties that allow them to play the appropriate role in the generation of perceptual episodes. What properties? Sellars suggested that they are constructed by analogy with the sensed qualities themselves. Our initial analogical model for a sense impression of a red triangle is precisely: a red triangular wafer. Clearly we cannot attribute 'red' to the sense impression in the same sense in which we attribute it to the wafer. So, Sellars suggests, we limit the application of the analogy to those structural patterns of resemblance and difference found among the colours. We invent 'counterpart properties' red*, green*, and so on, whose relations of resemblance and difference to one another are structurally similar to those found among the colours of objects, but which can be attributed to sense impressions. A sensation *of* a red apple is not itself red; instead it has some counterpart property, red*, in virtue of which the apple in front of the eyes *looks* red.

This distinction is not original with Sellars; it is clearly related to Thomas Reid's distinction of two different senses in which one can speak of the properties of 'ideas'. And the distinction has been noted, in various guises, by many authors since Sellars. An etymologically memorable and hence useful terminology is provided by Galen Strawson (1989). He distinguishes between 'phenomenal' properties and 'qualitative' properties, and I shall adopt this terminology. Phenomenal properties are all those that characterize how things appear. They are attributed to things in front of the sense organs. If the apple one spots, picks, and bites looks red, feels cold, and tastes sweet, then red, cold, and sweet are phenomenal properties. They characterize aspects of the appearance of the apple. Those appearances depend in part upon internal states of whoever happens to be apple-picking: upon the 'sense impressions', 'sensations', or 'experiences' of the apple-picker. Qualitative properties are properties of such internal states, in virtue of which things out there, dangling in front of one's grasping fingers, appear as they do. So the sense impression of the apple has a 'red qualitative character': a character in virtue of which the apple *looks* red. One might also use the Sellarsian notation: red*. The qualitative property is a property of an internal state in virtue of which something else—the apple at one's fingertips—has a phenomenal property: it looks red.

The proposed technical distinction uses the words 'phenomenal' and 'qualitative' in two of their accepted senses, but one must immediately acknowledge that those same words are used in other contexts in entirely different ways. Sensation being such an interesting subject to our species, we have evolved an extraordinarily intricate and flexible terminology for signalling, classifying, characterizing, analysing, describing, savouring, and stimulating it. So, for example, it is possible to find natural language usages that invert the proposed technical distinction: we speak of the 'phenomenal character' of experience and use 'phenomenally red' and similar locutions to ascribe properties to some mental states. Conversely, we speak of the red colour and sweet taste of the apple as obvious examples of the qualities of sensation. Even the philosopher's technical term 'qualia' is ambiguous between what I have called phenomenal properties and qualitative properties. C. I. Lewis (1929) used the term in both senses, for example. The redness of the red patch is for many philosophers the paradigm example of a quale; yet to the innocent eye it appears to be attributed to something out there in front of the eyes. Smooth, cold, and sweet are other qualia crowding in as one hefts the apple into biting range; yet they too characterize how the thing in the hand appears.

Some theoreticians identify qualia with what I have called 'phenomenal properties', and the going hypothesis is that these are properties represented by one's sensory experiences (Lycan 1996; Tye 1995). Yet qualia are typically thought to be properties of mental states, not of apples; one would be considered lunatic if one tried to resolve their problematic character by a careful study of arboreal botany. Qualia at least provide one subject properly studied in (and perhaps confined to) the halls of philosophy of mind. Theoreticians favouring this usage speak of qualia or qualitative character as properties ascribed to mental states (see Shoemaker 1990, 1994*a,b*). On one version, qualitative character consists of those properties of mental states in virtue of which they represent phenomenal properties (Harman 1990); on another, qualitative character is independent of the representational properties of mental states, but is accessible via introspection (Block 1990, 1995). The debate is further confused by the fact that the term 'qualia' itself has become invested and infested with theoretical commitments, so that to many participants the question of whether qualia exist is now synonymous with the question of whether intrinsic, non-intentional, introspectible properties of mental states exist (see Dennett 1988). Hence, sadly, the term 'qualia' can no longer be used as a neutral term describing a problem; now it entails a commitment to one kind of solution. This ruins the word for my purposes; the neutral term 'qualitative properties', introduced above, should not be assumed to have any of the defining characteristics of the now infamous 'qualia'.[1]

[1] Likewise any use I make of the words 'quale' or 'qualia' in what follows should be interpreted in the old-fashioned sense of C. I. Lewis, James, and Goodman; not as the infamous new 'strange qualia' (Lycan 1996: 110).

Like ‘qualia’, the label ‘quality space’ is ambiguous. The properties whose order it presents might be phenomenal properties, or they might be qualitative properties. In the former sense quality space is an ordering of the qualities that things appear to have; in the latter it is an ordering of properties of sensation, in virtue of which things out there appear as they do.

The distinction between phenomenal and qualitative properties helps to clarify the status of quality space. In fact the manufacturing process that yields this product has at least four distinct stages, or four different interpretations for the ‘qualities’ thereby spaced: (*a*) stimulus classes; (*b*) phenomenal properties; (*c*) qualitative properties; and (*d*) psycho-physical properties. Stimulus classes are used as the entering wedge to identify phenomenal properties, but, as will be seen, those properties cannot be defined in terms of stimulus parameters. And qualitative properties, if all goes well, ultimately are identified with neural ones. The crucial step is the step from (*b*) to (*c*), from phenomenal properties to qualitative ones; and this is genuinely a change in subject-matter, as the properties have distinct logical subjects. But let us trace the genesis of the product.

1.1.1 Classes of Stimuli

It is important to recognize from the outset that the occupants of quality space are not stimuli, but rather the qualities that stimuli present. In any ordering of hues, orange is between red and yellow (it is a binary combination of the two), but this ordering is not an ordering of stimuli. It has nothing to do with the fact that a 600 nanometre wavelength is between 580 and 620. In fact there may be no stimulus metric in terms of which an arbitrary stimulus for orange is ‘between’ arbitrary stimuli for red and yellow respectively. Physically disparate stimuli may all present the same hue quality. Metamers, or combinations of wavelengths that have the same effect on the visual system, will do so. Even though orange is between red and yellow, there is no determinate sense to the suggestion that all the metamers that present orange are somehow ‘between’ all those metamers that present red and all those that present yellow.

A stimulus is best considered an occasion, a particular episode of irritation of transducer surfaces. The various ambient energies involved can differ physically even though all those episodes appear the same to the creature suffering the irritations. The ordering of qualities does not order that hotchpotch of varieties of ambient physical energies. Relations of matching and discriminability are instead used to order the qualities that those stimuli present.

Matching is the simplest of such relations, since its relata are particular stimuli—particular occasions. But even matching has its subtleties. For example, in colour vision the relation has not just two terms—stimuli *x* and *y* match—but at least eight: stimulus *x* matches stimulus *y* for observer *O* in state of adaptation *A* under ambient illumination *I* viewed with surround *S* at angle α with angular subtense β. Change any of these terms, and the match is disturbed. Such profligate relativism provides a second reason to deny that the ordering provided by quality space is an ordering of stimuli. Physically the 'same' stimulus might on a second occasion present a different quality. If the observer's state of adaptation or the ambient illumination are changed sufficiently, it will do so. Flood the field with red light for a while, and a second presentation of that 'same' packet of ambient energies will look distinctly greenish.

Occasions being so fleeting, a science can get under way only if some scheme of classification can be devised under which one can recognize something repeatable in the flux. The notion already mooted, of generating a second instance of an event which is physically the 'same', provides one route. To the extent that one can develop instruments to control and repeatedly generate new instances of the 'same' packet of ambient energies, one can start to experiment. In practice within psycho-physics a 'stimulus' is a class of such occasions: physically controllable and repeatable instances of the 'same' packet of photons to the eye, compression waves to the ear, gases to the nose, and so on.

It is only after one can begin to manipulate such classes of occasions that one can begin to assess the more complicated relations of discriminability. We need these more complicated notions to cope with the otherwise disconcerting variability in responses to a forced-choice task. Present the 'same' pair of stimuli to an observer repeatedly, and sometimes they will be judged to match, sometimes not. One must assess whether that *distribution* of responses differs from one produced by chance alone. Such assessment is invariably statistical, requiring repeated trials. It follows that the terms for the relations of discriminability are classes of stimuli: occasions collected by the underlying physics of the ambient energies. Other workable terms for similarity have at least this minimal complexity: they too relate classes of stimuli. For example, triadic relations of relative similarity (*x* is more similar to *y* than to *z*) are unworkable if the terms are restricted to single, unrepeatable occasions. In comparing relative similarities of three colour patches, one may typically first compare *x* to *y*, and then compare *x* to *z*; but if *x* is allowed to range only over occasions, this second comparison is not a comparison of the 'same' *x* as found in the first.

Given some particular classes of repeatable stimuli that look orange, red, and yellow, one might be able to devise some physical interpretation for the claim that 'orange is between red and yellow' in stimulus terms.

Unfortunately, tomorrow one is likely to find other, disparate classes of stimuli that are metamers for those used yesterday, and which fail to satisfy those physical relations. Furthermore, the stimuli used yesterday might today present different qualities, since today it is overcast, or we are testing a different observer, or we finished repainting the walls of the shop. Relationships between the qualities cannot be cashed out directly in stimulus terms, and attempts to define 'orange' in stimulus terms are doomed to failure.

To the question 'What are the occupants of quality space?' the natural answer is 'qualities', and what these considerations show is that while we might be able to *label* a point in quality space with some stimulus specification—some class of stimuli which happened to present that quality—we cannot *identify* the quality with that class. A finite class of occasions might help to pick out the quality, but it cannot be used to define the identity of that quality. So from the very beginning, the relations of matching, discriminability, and relative similarity among classes of stimuli are used to order something other than those stimuli themselves. Discriminations among stimuli serve to order the qualities that the stimuli present.

How this works still seems like magic. It is not magic, but it is still wondrous to behold. Order emerges from indifference, to use Nelson Goodman's apt phrase (Goodman 1972: 423–36). In appearance we find structure. The various similarity relations—matching, discriminability, and relative similarity—have formal features that generate the order of qualities. Matching, for example, is non-transitive; *x* might be judged to match *y*, and *y* *z*, but the differences accumulating between *x* and *z* might be sufficient to reject a match between them. An ordering of qualities that respects the principle that more similar qualities should be placed closer to one another than less similar ones is obliged to put the quality presented by *y* between those presented by *x* and by *z*. Other more direct methods and more powerful inferences are possible; if you hunger for such, you can find satiation in *Sensory Qualities* (Clark 1993*b*).

The various similarity relations holding among classes of stimuli are used to order the qualities that those stimuli present. It follows that the relations among the qualities themselves are at some remove from relations among stimuli. The occupants of a quality space for colours will be, not packets of photons, but colour qualities: red, yellow, green, blue, orange, and so on. In such an ordering we find orange between red and yellow. This fact is at some remove from relations that we can state in stimulus terms. It does not imply that one packet of photons is 'between' two other packets. The relations of discriminability and relative similarity that suffice to place orange between red and yellow are not relations among the qualities themselves. One does not discriminate the quality orange from the quality red; instead discrimination always has as its terms classes of stimuli. What sort of fact is it that orange is between red and yellow? It is not directly about discriminations. Perhaps it could be called a *structural* fact. It is a fact about properties of the relationships of discriminability, matching, and so on holding among classes of stimuli that present the qualities in question.

Another implication of this somewhat indirect relationship between points in a quality space and particular stimuli should be mentioned. Attempts to order the qualities presented by stimuli are perforce limited to those stimuli used in the attempt. A quality space is always an ordering of qualities presented by a particular gamut or sample of stimuli; it depicts the order within that population, but foreign stimuli may follow other laws. In the lab one seeks occasions that are precisely controlled and replicable, and so coloured lights might always be presented through a reduction tube, and surface colours will always be presented as patches of a fixed size and angle, under constant ambient illumination, with neutral surrounds, and using procedures to ensure that the subject between trials returns to a neutral state of adaptation. A lab whose results change when it repaints its walls would not have its grants renewed. With a particular gamut of such stimuli we might find a colour space of just three dimensions: hue, saturation, and brightness. But these three dimensions will undoubtedly prove inadequate to describe the dimensions of variation in visual appearance found in a larger and more diverse sample of naturally occurring stimuli. Brown is found nowhere in such a space, for example, as its appearance depends critically on the presence of a bright surround. Even the number of dimensions of the quality space will depend upon the sample chosen. Glossy surfaces, reflections, translucency, transparency, shadows, and mists all require dimensions of variation in appearance beyond the three sufficient for coloured surfaces or lights presented in the lab.

1.1.2 Properties of Appearance

Now to pull some strands together. The science can start only after it forms classes of repeatable stimulus occasions. These classes might be used to label or identify a point in quality space, but these points are not identical to any such collection of occasions. The occupants of quality space are not stimuli. Instead they are something distinct: the qualities those stimuli present. These are, precisely, phenomenal properties, or properties of appearance. Matching, discriminability, and relative similarity do not proceed along any straightforward physical metric, but instead along a metric of similarity of appearance. Two stimuli match because they look the same, sound the same, taste the same, and so on. So in this first interpretation of quality space, the qualities that are its occupants are phenomenal properties. The dimensions of hue, saturation, and brightness are dimensions that suffice to order the variations in appearance among precisely controlled coloured lights in a reduction tube, or two degree coloured patches in a neutral surround. (Technically the dimension called ‘brightness’ for coloured lights is called ‘lightness’ for coloured surfaces.) Those variables describe how the stimuli *look*. Similarities and differences in those appearances can be captured with three independent terms. They delimit the gamut of phenomenal properties presented by that population. Granted, as soon as we loose the strictures of the lab, and consider some naturally occurring stimuli, we must be prepared for some additional complexities. Three dimensions prove inadequate to describe a reflection of the blue sky seen in a brown mud puddle, for example. But whatever dimensions we add will still be dimensions of variation in appearance: dimensions that order phenomenal properties.

On this interpretation a colour ‘quality space’ is very close to a colour order system. One way to reveal the variation in colour appearance is to construct a gamut of stimuli that present varying appearances, and then physically order the stimuli in such a way that stimuli that are close to one another present qualities that are relatively similar to one another. Colour sample systems are the best example, but arrays of tuning forks provide an auditory version. With the colour sample system one attempts to print pages whose inked patches are such that two patches are relatively close to one another on the page only if they look relatively similar in colour. The enterprise is by no means trivial; it requires sophisticated understanding of the chemistry of inks and dyes and of the rules of their combination. Even when it is successful it should be clear that the colour quality presented by a particular patch cannot be defined by the particular combination of inks that happen to present it.

1.1.3 Qualitative Properties

If we had to stop after printing an array of colour patches, the project would be relatively uninteresting, at least to psychologists and philosophers. But the plot thickens: these arrays of colour patches serve as something like illuminated manuscripts, revealing some of the secular mysteries of vision. Study of the dimensions of variation in phenomenal appearance provides surprisingly potent insights into what must be going on inside a creature’s visual apparatus.

A paradigm example of this sort of reasoning was provided by Thomas Young in 1801. The premiss was a fact long known about colour mixing and matching: that the appearance of any coloured light can be matched by mixing and adjusting the intensities of just three distinct ‘primaries’. The primaries had only to meet the requirement that no one of them could be matched by any combination of intensities of the other two. So three dimensions suffice to describe the variation possible in the appearances of coloured lights. What Young inferred from this was the trichromatic character of human colour vision: that there are at least three, and probably no more than three, distinct types of colour sensitive receptors ('particles') in the retina (see Boring 1942: 111). These we now identify as three classes of cones. Young recognized that there might be more than three, but that the facts of colour-mixing and -matching required only three. More than three would crowd the retina unnecessarily, so Young plumped, parsimoniously, for the minimum required.

Qualitative properties are properties of the internal states of the creature—those sensory states without which the creature would not see (or sense) what it sees (or senses). To revert to Sellarsian terminology: a sense impression of red is endowed with some property in virtue of which that sense impression is an impression of red, and not of some other colour. This property is not redness, exactly, but it is tied, by an analogous structure of resemblances and differences, to colour properties—to those properties *of which* the sense impression is a sense impression. The sense impression has the qualitative character red*, and in virtue of being red*, is a sense impression *of* red, and not (say) of blue.

What this manœuvre requires (and what Young recognized in 1801) is that the structure of similarities and differences among the qualitative properties of sensation must be sufficient to account for the structure of similarities and differences among phenomenal properties. By no means can there be *less* structure among qualitative properties than among the phenomenal ones, since such a finding would render miraculous the creature's capacities to discriminate. There could be more structure, but it is unparsimonious to suppose so. In technical, information-theoretic terms, the channel between transducer irritations and the eventual muscle twitches that indicate a discrimination has some capacity that can be measured in bits. More discriminations require more bits. If anywhere within this channel there are intermediary states that throw away information, conflate incoming *Ps* with incoming *Qs*, then such information can never reliably be regained, and the channel capacity of the whole must suffer. Sensory states are intermediary states within such a channel. Hence the information content of those sensory states must at least match that of the discriminations. And so there must be at least as much structure among the similarities and differences of qualitative properties as there is among phenomenal ones.

So now, if this reasoning is sound, 'quality space' receives a second interpretation. It tells us not just about the similarities and differences among the properties of appearance. It also tells us something about properties of the creature's states of mind. That order is now an ordering of the qualitative properties of sensation.

Qualitative properties are not properties of the things seen, heard, felt, or tasted, nor are they properties that characterize how those things appear. Instead they are properties of states of mind that help to explain why things appear as they do. So knowledge of them is not 'knowledge by acquaintance', or at any rate it is not knowledge that might be confirmed in some relatively direct fashion by looking, listening, or touching, no matter how carefully. Instead it is closer to 'knowledge by description' (see Shoemaker 1994*b*: 310). We make inferences about what the qualitative character of sensations must be in order to explain aspects of our relatively direct acquaintance with how things look. It is gratifying to note that other authors have come to similar conclusions on independent grounds. Dretske (1995: 41–4, 53) describes knowledge of the qualitative character of sensations as 'displaced perception'. By attending carefully to the variations in shades among inked patches of paper in front of one's eyes, one can come to appreciate something about the variations in qualitative character among visual states that are found somewhere behind those eyes. The array of colour patches—the colour sample system—allows indirect perception of the qualitative character of one's own visual states. In such fashion this illuminated manuscript casts light into the workings of one's own visual apparatus.

Suppose Otto is looking at colour patches in a lab, and the patches and conditions are controlled in such a way that the variations in appearance of the patches can be described in just three dimensions: hue, saturation, and lightness. Otto sees a red patch. Now we propose that colours have locations in a colour quality space. If Otto sees the redness of the patch, must he therefore see it as a location in colour quality space? Boghossian and Velleman (1991), considering this question, say

> either one already sees colors in terms of their locations in a space of co-determinates—in which case, the appearance of one color alludes to the others—or one doesn't yet see colors in terms of their locations in color space—in which case, their appearances furnish no grounds for drawing the similarities and differences constitutive of such a space. (Boghossian and Velleman 1991: 104)

They quite rightly attack the first alternative, on the grounds that the appearance of a particular shade of red does not imply that there exist neighbouring shades that are yellower or bluer. So we are left with the other alternative, and the problem of understanding what locations in colour quality space have to do with seeing a particular colour. As they say, 'If one doesn't already see colors under characterizations locating them in such a space, then one sees nothing on the basis of which locations could be assigned to them' (Boghossian and Velleman 1991: 104).

But notice what happens to this argument if one can distinguish between phenomenal properties and qualitative properties. The appearance presented to Otto by a particular red patch has a particular location within the order of phenomenal properties, but it need not present itself *as* having that location. To describe the gamut of colour appearances presented to Otto while he participates in our experiment, we require three dimensions of variation in colour appearance. The patch 'has' a location in the sense that it has some determinate value in each of those three dimensions. Otto can discriminate this patch from others we present him in the lab, and if we simply try to catalogue his discriminatory abilities in the domain of surface colours, we find we need three dimensions—three variables—to *describe* the data. Hence to describe a particular appearance of a particular patch, we must give some determinate value for each of those three dimensions. Otto need not see that particular shade *as* having those three 'coordinates', but if he does see that particular shade, he sees the shade that *has* those coordinates—the one that occupies that particular place within the order of phenomenal qualities. That is all it means to say that the appearance has a location; it need not be a content carried in the appearance itself.

Furthermore, the distinction between phenomenal and qualitative properties opens the possibility that Boghossian and Velleman deny. Appearances might furnish grounds for assigning a perceptual state a location within quality space even though that location is no part of the appearance itself. What is going on in Otto when he sees the red patch? He is in some sensory state *S* in virtue of which the patch looks as it does. We attribute qualitative properties to the internal state *S*. For *S* to have that particular qualitative character is for it (for example) to have a property in virtue of which the stimulus appears to have a particular hue, a property in virtue of which the stimulus appears to have a particular saturation, and a property in virtue of which it appears to have a particular brightness. Call these three properties 'differentiative' (see Clark 1993*b*: 70 ff.); what they help to differentiate are things in front of the eyes. If *S* varied in any of its three differentiative properties, the stimulus would look to have a different hue, or saturation, or brightness. That is all it means to say that *S* has a location in colour quality space.

If we require three dimensions to describe Otto's capacities to discriminate, then qualitative states within Otto must also have at least three independent dimensions of variation. If there were less, his discriminations would suffer, since the channel capacity would fall. A switch with just four wires cannot connect six people. It cannot happen. But we do not need to suppose that Otto can detect these states or these capacities within himself, or that he is even aware that he has such keen discrimination of the patches in front of his eyes. His own qualitative states never trot across the field of view. Perhaps, if he has read Shoemaker and Dretske, he can perceive them indirectly, but even then he must stay alert, and pay attention to these distinctions.

1.1.4 Psycho-Physical Properties

There is one more step before our product is complete and we can send it out the door. We must find a neurophysiological interpretation for the dimensions of variation in qualitative properties.

Colour vision is one modality in which one can readily anticipate making this step. It has not yet been made; identifications are at best tentative, and the data are confusing. But the 'opponent process' theory has with good reason become something of the 'standard model' in colour vision, and it yields a sketch of how these identifications are likely to proceed.

To say that there are three dimensions of variation in qualitative character does not say which three. It could be any three, as long as they are independent of one another. Various coordinate schemes could carry the information required. A Cartesian system of orthogonal axes is possible. The hue, saturation, brightness system treats hue as an angular coordinate, saturation as a radius, and brightness as a vertical distance. When we seek a psycho-physical interpretation for our quality space, we seek to know how those dimensions of variation in qualitative states are actually registered by the nervous system. In a sense it is only after this is done that qualitative properties have been identified. Until then, we know at best that there are some *k* distinct dimensions of variation in the qualitative character of states in the given modality, but we do not know what those dimensions are.

The answer turns out to be quite surprising. Colours of a given brightness are not registered by separate coordinates of hue and saturation; the latter are *not* the differentiative properties on which chromatic discriminations actually proceed. Instead the gamut is described by two 'opponent' axes. One axis registers where the colour is in a series from yellow through grey to blue, and the other where it is in a series from red through grey to green. Both opponent axes carry information about both hue and saturation, but mathematically the result is the same: they describe a plane. The pair of coordinates yields a unique position in the two-dimensional order of hues and saturations of a given brightness. But no a priori theorist would ever have guessed that this specification actually proceeds with one coordinate for yellow–blue and another for red–green.

The simplest model that accounts for the trichromatic character of colour vision would be one in which chromatic systems use just three independent primaries (such as red, green, and cyan), as do printers and televisions. To pair red with green in one channel, and yellow with blue in another, seems inelegant and inefficient. The Young–Helmholtz model denied the existence of opponent processes, and of the counter-intuitive pairings of red with green and yellow with blue. But nature here trumps our intuitions: colour vision is trichromatic, but the actual axes on which discriminations turn are provided by opponent processes. Each such axis has distinct 'unitary' hue qualities at opposite ends. Those two hues are complements. Red cancels green, and yellow does blue. Mixtures of them yield some point in the achromatic scale from white to black. Such chromatic cancellation gives the theory its 'opponent' character.

We should think of red as the end-point in a series which derives from the series-generating relationship 'is redder than'. At the opposite pole of this relationship is the quality *green*. Even though we have no term in a natural language for this order-generating relation, red–green is one of the axes of variation along which the visual nervous system actually makes its discriminations. So too for yellow–blue and white–black.

1.2 THE STRUCTURE OF APPEARANCE

The moral of the story is that we might be surprised when we discover which dimensions of variation in phenomenal appearance are actually registered by the nervous system. How is it possible to be surprised by such facts? How can elements of the structure of appearance be hidden from us in this way?

It helps to return to the question of how to characterize the relations between qualities—between the occupants of quality space. Consider the fact that in an ordering of hues, any appearance of orange is found between those of red and yellow. This is not a straightforward physical fact: the various stimuli that appear to be orange (or red or yellow) form a heterogeneous physical collection. Logic forbids us from mentioning any stimulus specification in a definition of these phenomenal properties, since it is logically possible that any such stimulus could present a different quality. It could look different. Tomorrow it might.

The fact that orange is between red and yellow is not a straightforward physical fact. I shall dub it a fact of 'qualitative structure'. That structure is generated by relations of discriminability, matching, and relative similarity among classes of stimuli; the relations between qualities are consequences of that structure. I shall also use the term 'qualitative similarities' to encompass the relations among qualities. So, for example, it is a fact about qualitative similarities that places orange between red and yellow. This is not a fact that could be confirmed in any single discrimination trial or matching task. Similarities among qualities are at some remove from the relations of discriminability, matching, and relative similarity that hold among stimuli. A sensory system can register similarities between instances of orange and of red, but the qualities themselves are not objects of discrimination. Otto can sense instances that appear orange, but the quality orange could not fit within his field of view.

There are good reasons to think that all the facts that could be mentioned in defining a particular qualitative term are found among such facts of qualitative structure. The first reason is that only such structural properties generalize across individuals. Stimulus specifications do not. Differences in colour perception, even among those with 'normal' colour vision, can be demonstrated in various ways. The simplest is to pick monochromatic primaries—lights of just one wavelength—and have observers attempt to match some intermediate stimulus by mixing the primaries in different proportions. The proportions may differ vividly.[2] A mixture that matches the target for Otto may not for Sally. So if we are attempting to define the name of the hue presented by that intermediate stimulus, we cannot name what matches for Otto or for Sally.

More dramatic demonstrations are available once percipients know some hue names. Given an array of greenish Munsell chips, and the instruction 'pick the chip that to you presents a unitary green—not at all yellowish and not at all bluish', an individual will manage to pick one chip, and will reliably pick that same chip again on repeated occasions. But different individuals will (reliably) pick different chips (see Hurvich 1981: 223). The differences are readily perceptible to anyone involved in the test: Otto's chip may look noticeably bluish to Sally, while hers will seem yellowish to Otto.

These differences present a problem. Given the variations in normal colour vision, we cannot identify colour kinds with classes of stimuli. The class that presents aqua to Otto may not do for Sally. What then generalizes? Why do we even think that two different people are sometimes presented with the same aqua quale? What we find is that while the exact stimuli needed to make a match vary somewhat from person to person, there is an identifiable pattern of relations that generalizes from person to person. The structure of relations generalizes, even if the particular relata do not.

[2] The 'Nagel anomaloscope' presents observers with a split field, in half of which is a stimulus of 589 nm. In the other half is a mixture of 535 nm and 670 nm light, whose relative intensities are adjusted by the observer until they match the 589 nm stimulus. That ratio is commonly used to assess various colour vision anomalies. But even in the range of ratios considered 'normal', some observers will completely reject matches made by others. See Hurvich (1981: 227, 230).

The point can be put in simple logical terms. Consider just those with 'normal' colour vision: the non-anomalous trichromats. One might expect to be able to generalize across those people at least with propositions of the form:

> There is a stimulus *y* perceived by every person *z* as a green that is not at all yellowish and not at all bluish.

This hope is dashed, even among those with normal discriminations. Instead we need to say something of the form:

> For every person *z* there is a stimulus *y* perceived by *z* as a green that is not at all yellowish and not at all bluish.

Once we identify such 'unitary' hues for an individual, we can state a great deal about the colour perceptions of *that* individual in terms of those hues. For example, suppose we find that to Otto wavelength mixture *x* presents unitary blue, and mixture *y* presents unitary yellow. Then for Otto *x* and *y* are complements: some combination of them will yield achromatic white. But that same pair may not be complements for Sally (even though she too has 'normal' colour vision). To Sally *x* may appear to be a slightly greenish blue, and *y* a slightly greenish yellow. Their combination will always appear somewhat greenish to Sally, no matter how she adjusts their relative intensities. So it is false to say

> There are wavelength spectra *x* and *y* such that, for any person *z*, *x* combined with *y* yields an achromatic white for *z*

but true that

> For any person *z* there are wavelength mixtures *x* and *y* such that *x* combined with *y* yields an achromatic white for *z*.

The structure generalizes, but the particular stimuli occupying a particular place in the structure vary somewhat from person to person.

Even if colour discriminations were identical across a population, one could not define a particular colour term by the stimuli that happen to present it. Suppose we say that a thing looks red if it presents the same colour as geranium petals. One problem is paradigm existence: if this is to be true by definition, it implies that geranium petals must exist if anything in the universe looks red. But geraniums might not have evolved, even though we still paint our fire trucks red. Secondly, treating the identification as an analysis would imply that geranium petals must be red. But that too seems a contingent matter: they might have evolved to be yellow instead.

A simple conclusion follows: if we are to define a term for a qualitative property (a 'qualitative term'), we cannot mention any stimuli. We can mention only the structural properties that give the quale its place in the quality space. 'Orange' cannot be defined as 'the colour of ripe oranges' or in any similar way, no matter how sophisticated. It can only be defined as something like 'the colour midway between red and yellow, and more similar to either than to turquoise'. The terms 'red', 'yellow', and 'turquoise' would all receive similar analyses.

For any particular observer, the qualitative property 'orange' corresponds to a place in the colour quality space, which can be indicated, picked out, or tagged by a sample of stimuli. But as soon as we change observers (or the illumination, or the state of adaptation of the observer, or any other member of the eight-term matching relation mentioned in Section 1.1.1), those stimulus pointers may point to the wrong quality. We must somehow identify *corresponding* places in the colour quality spaces of other individuals, but this cannot be done by listing stimuli. Only a structure description will do. The problem is analogous to identifying, in the skeleton or anatomy of different people, which parts are the femur or the nose. These parts are rarely shaped identically in any two people, so one could not identify the femur or the nose with any specific configuration of bone or tissue. Fortunately, the pattern of relations that they bear to other parts of the person generalizes in a rough (homologous) fashion. We pick the corresponding place in a homologous (not identical) pattern of relations.

The implication is that we cannot treat stimulus characterization as any part of the *analysis* of qualitative terms. If we try to define a particular qualitative term, the only sorts of facts we can mention are facts about the similarities of the given quality to others. And all the essential properties that we might think of are encompassed in that structure. So, for example, no particular stimulus is such that it must look orange, but anything that does look orange presents an appearance that must be somewhat reddish and somewhat yellowish. Anything that presents that particular quality presents a quality that must stand in particular relations of qualitative similarity to red and to yellow. Otherwise it is not orange. Orange could not more closely resemble turquoise than it does red.

One can think of these structural facts as 'geometrically intrinsic' properties of the colour quality space. Intrinsic properties are properties that a geometer confined to its surface could discover; they are all facts of structure. If your goal is to define a particular qualitative term, then all the essential properties you might need (and more besides) will be found in this list. Extrinsic facts require a coordinate scheme defined independently of the surface itself. The fact that a particular stimulus happens to present a particular quality is extrinsic to the geometry of the space itself. So, as implied by the analogy, it has nothing essential to do with that quality.

Previously I described the contrast between intrinsic and extrinsic properties in terms of the difference between 'distal' and 'proximal' functional roles (Clark 1993*b*: 201–2), but the latter terminology now seems wrong. As just seen, the causes of sensations of orange add nothing essential to and cannot be mentioned in any analysis of the qualitative property that those sensations instantiate. We are also unlikely to find any set of effects that is unique to having had a sensation of orange, and in terms of which the qualitative property might be defined. What makes orange orange is not a fact about the consequences of seeing something as orange. If we bundle those causal relations together as a 'functional role'—a job that signals of orange serve within the psychological economy of the creature—then there does not seem to be any particular functional role that sensations of orange invariably and distinctively perform. Whatever orange signals could just as well have been signalled by turquoise. If defined by their jobs, these employees are truly interchangeable.

Talk of 'distal' and 'proximal' functional roles was based on Ned Block's distinction between 'long-arm' and 'short-arm' roles. Block defined 'long-arm' functional roles as ones that 'reach out into the world of things' and 'short-arm' functional roles as ones that are 'purely internal' (Block 1990: 70). He asked:

> Why can't the functionalist identify intentional contents with long-arm functional states and qualitative content with short-arm functional states? The result would be a kind of 'dualist' or 'two factor' version of functionalism. My response: perhaps such a two factor theory is workable, but the burden of proof is on the functionalist to tell us what the short-arm functional states might be. (Block 1990: 70)

One might think of the short-arm roles as all and only those that define the place of a qualitative property in its quality space, but for the fact that the latter are not functional roles at all. Just as there is no particular stimulus unique to orange, so there is no particular causal consequence that is unique to sensing orange. The box in the flow chart to be labelled 'sensations of orange' cannot be identified by its causal relations to other boxes. Instead those contents are to be identified by the relations of qualitative similarity that define the intrinsic spacing of qualities.

If we are looking for a 'functional definition' then conventional wisdom presumes that we are looking for a definition in terms of a theory whose root relation is '. . . causes . . .'. One identifies nodes in a causal network, or boxes in a functional flow chart model, and specifies causal relations between the inputs and outputs of the boxes. Such a theory is converted into a 'Ramsey sentence', which provides a higher-order description of a structure of relations. (To make a Ramsey sentence, replace each theoretical term with an existentially quantified variable whose values are specified solely by the relations they bear to the values of other such variables. See the Appendix for details.) In a conventional 'functional definition', the relations providing the structure are causal. Particular clauses of the Ramsey sentence (particular 'Ramsey functional correlates') are then used to define particular theoretical terms. Each is assigned a job.

But facts of qualitative structure cannot be assimilated to this model. The root relations that define this structure are not causal ones; they are relations of qualitative similarity. Rather than accept a theory that requires us to specify some particular 'causal niche' that is always and only occupied by sensations of orange, we should abandon the theory (see Levine 1995). There is no 'causal niche' filled characteristically and uniquely by sensations of orange. But a relational structure can be built up using relations other than 'causes'. I suggest that if our goal is to describe qualitative character, the root relations will be those of qualitative similarity.

We should think of the definiens for a qualitative term as a *qualitative niche*. The niche for orange is provided by the structural facts that place it between red and yellow. That is where orange belongs; that is its proper home. When we attempt to define a qualitative term we should confine ourselves to those structural properties that locate the given quality in its quality space. Describe the relations of qualitative similarity among the occupants, but mention no stimuli. So 'orange' could at best be defined as something like 'a reddish yellow, equally similar to red as to yellow, complementary to turquoise, more similar to red or to yellow than it is to turquoise', and so on. Each of the other qualitative terms would receive an equally enigmatic treatment. Clearly this regimen can succeed only if it is utterly methodical: one must form the Ramsey sentence for the entire structure of qualitative relations, and associate each particular term with a particular Ramsey correlate.

Of course it is unlikely that any term in a natural language has as part of its language-meaning a structure description of this sort sufficiently rich to determine its extension. But there are much simpler ways the term might latch onto its extension. Any creature that could use the term natively to report its sensory episodes is endowed with the sense of qualitative similarity that such structure descriptions *describe*. So all that creature needs is a few presentations of paradigm and foil, and a reference-fixing ceremony of the form 'orange is any colour more similar to this (paradigm) than to those (foils)' will succeed admirably. We do not need to describe the relations of qualitative similarity satisfied by the quality, since we can *sense* their instances.

1.3 INTRINSIC VERSUS RELATIONAL

A consequence of this account is that qualitative character is a relational affair. Qualitative properties seem to be intrinsic properties, but they are not.

When one sees a patch of orange, the experience seems to involve an intrinsic monadic quale: the quale orange. But this experience is an illusion. The facts in virtue of which that experience is an experience of orange, and not of some other quality, are all relational facts.

Of course when one has an experience as of something orange, one need not have an experience as of something related in such-and-such a way to other things red and yellow. Qualitative content may be relational even though it is not invariably experienced as relational. Analogously, weight is a relational property even though it seems intrinsic to the loads one might carry. It seems outlandish to suggest that one could change the weight of a sack of groceries without disturbing a single atom of its contents. Nevertheless, the facts in virtue of which the sack weighs what it does are relational facts about mass and distance. The sack is somewhat lighter at high tide, and if aliens stole the core of our planet they would make all of our groceries lighter. Similarly, the facts in virtue of which a sensation has the particular qualitative character it has are all relational facts.

This suggestion is a 'psycho-functional' hypothesis and not an 'analytic' definition (see Block 1980). It is not derived from analysis of our ordinary concepts. A psycho-functionalist believes that many of the putative necessary truths that one derives from analysis of those concepts are, in the end, neither necessary nor true. Instead empirical inquiry is to be given the task of discovering the 'real essence' of our psychological states. The theory describing that nature may contradict smaller or larger portions of our ordinary conceptual scheme yet still, on the whole, be the best theory going. Perhaps it is part of our ordinary conceptual framework that qualia are intrinsic. And perhaps that part is a portion of our folk inheritance that we must renounce. Empirical inquiry suggests that the facts of qualitative character have at root a relational form.

The most dramatic illustration uses the blandest qualia imaginable: those had when looking at a 'grey scale', or array of non-chromatic stimuli ranging from white through the various greys to black. Such non-chromatic colour qualia turn out to be just as mysterious as the chromatic varieties. Suppose that one perceives a patch somewhere in the middle of the series as a middling grey. In virtue of what are two sensations both sensations of this particular phenomenal property—the appearance of a middling grey?

In 1948 Hans Wallach performed a simple experiment that suggested a surprising answer to this question. He projected a circle of light onto a screen, and set up a second projector that could surround that circle with an annulus, whose luminance could be controlled independently. Wallach found that the apparent brightness of the centre disc could be changed without touching the projector that sent light into it. He had only to change the brightness of the surround; the brighter it was, the darker the centre appeared. In other words, the same fixed luminance reflecting off the centre circle on the screen could be made to appear as any one of many different shades of grey. He performed a second experiment in which the observer was confronted with two such centre–surround displays at different luminance levels. The observer could adjust the intensity of light in the disc in one of the displays, and was told to adjust it until the centre discs in both displays looked to be the same grey. One centre spot might be reflecting up to ten times as much light as the other. Nevertheless, it could be made to appear the same grey as the other if it bore the same ratio to the luminance of its surround. So the facts that make something appear grey are relational in form: that its luminance stands in a certain ratio to the luminance of its surround. This ratio principle holds over a large range of intensities.[3]

What appears to be an intrinsic property of the centre patch in fact registers a relation between luminance levels of the centre and the surround. That central grey certainly looks as if it is a monadic property, ensconced in the centre, independent of its surroundings. Just like the weight of the sack of groceries, the grey of the patch seems to be intrinsic to the patch, but it is not. It depends on relations among luminance levels. Note that the latter in turn may or may not be relational in form; the claim should not be taken to imply that *all* of the facts on which the greyness of the patch depends are relational in form. Greyness depends on relations of luminance levels, but perhaps physicists and metaphysicians will ultimately tell us that luminance levels are intrinsic properties. Analogously, while weight depends on facts about the relations of mass and distance, the latter facts may or may not ultimately have a relational character. Perhaps the mass of the cheese one purchased is an intrinsic property of that cheese. Its ultimate metaphysical status can be left for others to resolve. Whatever that resolution, it is clear that mass and weight are different properties: even if the sack becomes weightless, it still has mass. It is also clear that the sack weighs what it does in virtue of relational facts about mass. Similarly, luminance and lightness are distinct properties, and lightness depends on relations among luminance levels.

This finding did not surprise those who had absorbed the principles of the opponent process theory of vision; all the surprises are already contained within that theory. Ewald Hering's model for an opponent process was a group of opposing muscles; he thought the chromatic quality presented by a stimulus was essentially the *resultant* of opposing forces within the visual

[3] Wallach's experiment was careful enough that he detected slight deviations from the ratio principle: sometimes the setting for the disc in the less intense pair was somewhat lower than predicted by an exact ratio, sometimes higher. And high intensities could make the centre disc appear luminous (Wallach 1948: 317). These deviations from perfect brightness constancy later became targets for independent investigation.

system. Green pushes one way, and red pulls the other. One could cancel green by adding red. This is a very odd and counter-intuitive theory, and over a century passed before experimenters took it seriously. Now the evidence is quite strong in its favour; it seems to be the truth about how we perceive colours.

Hering argued in 1878 that the quale black is in fact a relational property. The 'resting-point' for the visual system is a middling grey, neither black, nor white, nor of any hue. To nudge the system out of its neutral resting-point, it must be pushed or pulled in one or another direction. Physiologically it has three axes, and it can be activated or inhibited along each such axis. But it followed, said Hering, that one could not see black unless one added inhibition somewhere else. That is, you cannot see black unless you can also see white. In fact one must be able to see white somewhere else at the same time. Otherwise, no black.

Wallach confirmed this prediction. The only way to get the centre circle to look black is to turn up the intensity of the surround until that surround looks white. If the surround is some middling grey, then so is the centre. You need the bright white surround, creating lots of inhibition, to get the centre to look black. Some ordinary experiences can confirm this. In the complete absence of light, what sensory quality is apparent when you open your eyes? The interior of even the darkest room does not look as black as the interior of a filled inkwell seen in daylight. Dark indeed it is—we use the term partly to designate luminance levels—but the quality appearing before one's eyes does not match the blackness of the inkwell when one turns on the lights. To get black, you must add luminance, create inhibition, to push the opponent process in the appropriate direction. Hering called the dark grey that you see in the complete absence of light 'brain grey': it corresponds to the resting-point of the opponent process.

A simpler example is described by C. L. Hardin (1988: 24). When a television set is off, the screen always appears to be some shade of grey. But with the television turned on, one can see on the screen surfaces that appear to be black: the black hat of the villain, or black oil gushing from the Oklahoma oil well. The appearance is created by *adding* luminance to other portions of the screen. Even high technology cannot produce a raster gun that subtracts photons. So how is it that we get the appearance of black by adding light? It makes sense only if black is not just the absence of light, but a quality produced when an opponent system is inhibited, or pushed below its baseline.

So the appearance of black is a relational affair, even though it appears to be intrinsic. The same moral holds for hues. Orange is a resultant of activation in two opponent processes. It has a red component and a yellow component in the same way that a north-easterly heading has a northerly component and an easterly one. Just as one cannot head north-easterly unless one can also travel easterly, so one cannot see orange unless one can also see red. The hypothesis is, admittedly, outlandish, implausible, perhaps even conceptually incoherent. Analysis of our ordinary concepts would never lead to such a bizarre position. Its only virtue is that it seems to be true. As Joseph Levine has quite reasonably written, ‘it certainly seemed possible that someone could experience a sensation with a reddish quality even if they were incapable of experiencing sensations of other chromatic types’ (Levine 1995: 286). I agree. It does seem possible. Our intuitions yield not the slightest murmur of complaint at the prospect. But alas, our intuitions have once again led us astray. Even though it seems possible, it is not. A person incapable of seeing green could not see red. The best available theory implies that it cannot happen. Levine goes on to say: ‘I think we could take this to an extreme and imagine someone whose entire visual experience involved just one hue, and that one was qualitatively similar to the one involved in my experiences of type Q_R. Why shouldn’t this be (at least conceptually) possible?’ (Levine 1995: 286). Indeed, it is conceptually possible: nothing in our ordinary concepts forbids it. Nevertheless, it cannot happen. That one could not see red if one lacked the capacity also to see green by no means follows from analysis of our ordinary concepts. It follows from a theoretical account—based on opponent process theory—of colour vision. Such theories always yield an indissoluble mix of new claims and new concepts, and this one will indeed force us to revise some of our pre-theoretic notions.

To the complaint that ‘this method does not explain the *intrinsic* nature of a colour experience’ (Chalmers 1996: 235), my response is ‘Guilty as charged, Your Honour’. There is no such nature to be explained. Not only is there no intrinsic nature to colour experience, but given what we know about vertebrate sensory physiology, it is difficult to see how intrinsic properties could account for the qualitative character of any sensory experience.

The difficulty derives from the opening gambit made by every vertebrate sensory system so far studied: transduction. At the outermost afferent ends of any such system we find transducers, which are just specialized cells devoted to converting energy of one sort into another sort. In vertebrates the other sort is always a difference in electric potential energy: differences in ion concentration across a neural membrane. Many varieties of stimulus energy are rendered into such differences. Sensory systems convert electromagnetic energy, compression waves in the air, thermal energy, mechanical distortion, etc., into differences in electric charge across the membrane of neurons. In a human about 70 per cent of all transducers are visual, which may explain why visual examples so dominate the literature.

Now for each transducer of a given kind there is some stimulus that is the optimal stimulator for that transducer. Of all the stimuli out there it is the sort that is most likely to make the transducer shriek with joy. For example, the short-wavelength cones in your retina are most likely to respond when bombarded with photons whose wavelength is roughly 430 nanometres (nm). But although this is the optimal stimulator, others will do. If we bombard the same cones with light of 480 nm, they will respond, but less readily, less happily, at perhaps half the rate. Simply ramp up the intensity of the 480 nm stimulus, double the number of photons, and the two stimuli will become indistinguishable. The cones respond identically to 430 nm light or to more intense stimulation at 480. We cannot treat them as 'wavelength detectors'. They conflate wavelength and intensity.

The problem is that transducers are 'broadly tuned' or 'broad-band'. So how can one use a cell that conflates wavelength and intensity to discriminate among wavelengths? The answer is that one cannot use a cell to do this, but if one had two such cells, whose optimal wavelengths differed somewhat, and one could compare their outputs, one could begin to separate variations in wavelength from variations in intensity. After transduction one of the next moves in a sensory system is invariably a comparison of results. As Russell and Karen De Valois said, 'This mode of operation, with broadly tuned receptors feeding into neural circuits that compare and contrast the outputs of different receptors to extract specific information, seems to be general to all sensory systems' (De Valois and De Valois 1975: 121). They go on to say 'it is the comparison of the excitatory and inhibitory inputs at each synaptic level which forms the very basis of the information processing that takes place in sensory systems' (De Valois and De Valois 1975: 125).

If that is the basis, then we are limited in the choice of structures we can build on top of it. Even if we imagine that some receptor is stimulated by packets of intrinsic red, any such receptor could be stimulated to respond by other, unprivileged packets as well. Within a synapse or two, to sort out the jumble, the nervous system must resort to comparing the responses of its different constituents. At that point we must abandon the pretence that it is detecting or registering intrinsic properties; instead we find the beginnings of a relentless process of comparison. Thereafter the sensory system transacts its business in relational terms alone.

Perhaps this is the reason such systems seem appallingly insensitive to absolute intensities, and tend instead to favour ratios and relations. While Wallach could vary the intensity of illumination falling on his centre disc by a factor of roughly one to a thousand, normal photopic vision copes with a gamut of intensities a thousand times wider, from one to a million (Schiffman 1982: 179). Yet Wallach's centre disc still appeared the same shade of grey, as long as it stood in roughly the same ratio of luminance to its surround. So in nervous system terms, that apparently intrinsic shade of grey does not register a monadic property or a particular level of luminance intensity. Instead it registers a relation between the luminance of the patch and the luminance of its surround: a relation that might obtain anywhere within a rather large gamut of absolute intensities. Opponent process theory suggests that the same moral should be applied to the chromatic colours: what seems to be bright monadic red can be cancelled by green. Like a positive charge, a heavy weight, or a northerly heading, the hue quality is at root relational.

The claim that qualitative character is intrinsic also runs foul of the facts of qualitative structure. If these were intrinsic properties, then no necessity would attach to any of the relations of qualitative similarity. If orange were a monadic property, then it need not be composed of red and yellow. And the resemblances among colours would be contingent consequences of the monads in question. Monadic properties as distinct existences cannot imply or necessitate any of their relations. So, on this line, orange is the property that it is; that it appears to resemble red and yellow is a contingent by-product of the funny way our vision works. If our vision worked in a different way, perhaps orange—that very same monadic property—could appear most closely to resemble blue or green. If you find this consequence jarring, then you feel the strong tug of resemblance among the hue qualities. The relations among them are not mere accidents; what appear to be monadic qualities derive from a relational root.

1.4 FOUR REFUTATIONS

Any account of qualia immediately faces a barrage of a priori objections. While any stratagem for deflecting such projectiles is to be commended, it is particularly valuable to find an account that provides a coherent defence against a large number of them simultaneously. In this section I pick four of the most famous objections and show how this account has the resources to answer all four.

1.4.1 Spectrum Inversion

To imagine spectrum inversion is to imagine a new mapping from stimuli to the colour qualities they present. Suppose Otto and Sally are spectrum-inverted relative to one another. If inversion is to preserve 'functional isomorphism', then all the judgements of matching and relative similarity made by Otto must be made as well by Sally. This implies a rather strong and surprising condition on colour quality space: to be invertible, it must manifest some variety of symmetry. There must be at least two distinct assignments of stimuli to colour qualities that preserve all the relations of qualitative similarity among the colours that those stimuli present. Otherwise in some region or other Otto will find two classes of stimuli to be relatively similar that Sally does not, and the game is up.

Since it is a biological product, it would be remarkable if the colour quality space had the structure of any of the Platonic solids. That is to say, it seems overwhelmingly likely that the colour quality space is asymmetrical. It is a lumpy, anisotropic, asymmetrical ovoid. These asymmetries will also give the clues needed to hook the structure up with neural hardware.

Human colour quality space is not isotropic; in some parts of the spectrum it is much easier to discriminate wavelength differences than in other parts (see Boynton 1979: 256, 281; Hurvich 1981: 296). So using physical coordinates, the space will vary in 'density'. We add other constraints. In moonlight we are all monochromats, but as rosy-fingered dawn approaches, reds and greens become visible first, followed by yellows and blues.[4] The red–green process has a lower threshold, and wakes up before the yellow–blue process (see Hurvich 1981: 72). Hence at the dark end of the colour ovoid, hues start bumping out in the red and green directions before much happens in the way of yellow or blue. We can note other asymmetries. Some hues can become more saturated—less similar to white—than others. The most saturated yellow still seems more similar to white than does the most saturated red. (If saturation is a radius, the most saturated red will be further away from the achromatic centre of the hue 'circle' than will be the most saturated yellow.) The 'colour solid' is not a Platonic solid. It is a lumpy product of our biology.[5] So even though it is conceivable that there be a systematic inversion, in fact any such inversion would be detectable.

Nevertheless, the mere thought experiment causes considerable consternation in the camp of analytic functionalists, who attempt to define the meanings of our ordinary qualitative terms by appeal to causal or functional roles.

[4] This is the 'Bezold–Brücke' phenomenon. The apparent brightness of hues changes as illumination changes. Although red may be the first visible hue, yellow soon makes an appearance, and can eventually surpass red in apparent brightness. The yellow–blue process requires more illumination to get going in the morning, but then accelerates more quickly. See Hurvich (1981: 72–4).

[5] Two other asymmetries are worth noting. The first is that hues presented by a given package of wavelengths typically shift as luminance increases, but there are exceptions, called 'invariant hues'. An interesting identity statement (that *can* be explained) is: the invariant hues = the unitary hues (Hurvich 1981: 73). The combination of curved and straight hue contours at different illuminations provides an asymmetry. Secondly, red–green and yellow–blue processes are not equally or symmetrically distributed across the retina. Yellow–blue discriminations fail for very small objects, even though objects of that size can be seen as red or green. This is called 'small field tritanopia' (Hurvich 1981: 21, 162).

According to the analytic functionalist, two instances of the same qualitative character are instances of that same character because they share a functional role. They have the same job; they perform the same causal duties. But it is conceivable that what looks red to you looks green to me, so that the functional role served by sensations of red in you is played by sensations of green in me. Hence what makes your sensation of red a sensation of red is not its functional role, and functional accounts fail as an analysis of qualitative character.

The account proposed above agrees that the facts in virtue of which a sensation has a particular qualitative character are all relational in form. But the relation at work is not '. . . causes . . .', but rather those relations of qualitative similarity that give the particular quale its place in the spacing of qualities. If your aim is to *define* a particular qualitative term, the only facts you could mention are such facts of qualitative structure. One must acknowledge that the results are meagre. Try it. Lay out the surgical implements, roll up your sleeves, then perform conceptual analysis on the word 'red'. There is scant stuffing to unpack; the attempt terminates with shocking abruptness. As William Lycan (1996: 80) has noted, analysis yields slim pickings in the domain of colour perception, and indeed in any sensory domain. Perhaps this is because much of the semantics of these terms is implicit in the operation of sensory mechanisms. We cannot *describe* the extension of the term; instead we point to a few samples and hope that our interlocutor shares our sense of what matches what. If analysis yields anything at all, it will yield only facts intrinsic to the geometry of the quality space—those that a geometer confined to its surface could discover. These are facts about the structure of relationships which the qualities bear to one another. Such analyses do not and cannot mention any stimuli.

If that is our concept of red, we can explain why spectrum inversion is conceptually possible. No conceptual rule is violated if we suppose that what looks green to you looks red to me. Inversion leaves unaltered the structure of relationships in which the various colour qualities stand. Red is still the complement of green, orange is still between red and yellow, and so on. What changes is merely that new stimuli come to present those qualities. The structural facts are unperturbed by an inversion. So contemplation of it does not cause any squawk of protest from our concepts.

It follows that this account does not deny the conceptual possibility that the stimuli presenting a given quale might all change. It is conceivable that the causal consequences of having a sensation with a particular qualitative character might all change as well. Since it is not a causal fact that places orange between red and yellow, the account remains indifferent to the job changes a quale may undergo in the course of its career. It is still the same old orange, but now activated by different events out at the periphery, and having different effects within. The same old orange moves into a new office and is given a new job. What makes it the same old orange is that it still stands in just the same place in the structure of qualitative similarities.

The only sort of inversion that the structural account need fear would be one in which (allegedly) the same quale is moved to a different place in the structure of relations to other qualities. For example, if the quale orange could come most closely to resemble blue or green, or if it could cease to have a reddish component and a yellowish component, even while remaining the same quale, then the facts of qualitative structure would not suffice to fix qualitative character. But this thought experiment is a very different beast from the one loosed upon the analytic functionalists, and it is much more difficult to manage successfully. Frankly I find it impossible to conceive of the envisioned state of affairs. If the quale is still orange, it could not lack a reddish component, and it could not more closely resemble blue than it does yellow. To move the quality into a new, greenish-blue neighbourhood, we must either change its character—so that it flies new colours—or change the constitution of the relations of qualitative similarity and difference that it bears to its neighbours. One or the other must give, yet this thought experiment requires both to be fixed even as the move is consummated.

The relations that generate the structure are not causal but qualitative. Even if the stimuli presenting the quale orange were to change, and the job or functional role of sensations of orange were also to change, we could still apply the same techniques of analysis to the discriminations of our hapless observer, determine that the same structure of qualitative relations obtains, and identify within that structure the same old orange, now with new stimulants and a brilliant new career. We could congratulate our old friend—good old orange!—on its successful move to a new office. The structure description associated with a given qualitative term describes a qualitative niche, not a causal one.

1.4.2 What it is Like to be a Bat

Although the analysis of discriminations can yield a structural definite description for a given qualitative character, it is implausible to think that such a description is part of the meaning of any word of a natural language. There is no synonymy between 'orange' and 'a reddish yellow, equally similar to red as to . . .', even though the two may have the same extension. The words we have for sensory qualities cannot readily be associated with conditions true of all and only the stimuli that present them. As noted, 'conceptual analysis' of such words disgorges little. Perhaps we get some paltry description of orange (e.g. that it is a colour, that it is a reddish yellow, etc.), but such descriptions fail to determine the extension of the term.

Fortunately our sensory systems themselves can furnish the wherewithal to connect a term to its extension. The structural definite description attempts to describe the relations of qualitative similarity that obtain among the qualities. If one is endowed with the modality in question, there is no need to describe those similarities, since one can *sense* them. Instead of trying to cobble together a description using the relatively immature faculties of linguistic representation, one can achieve the same end by activating the ancient and well-honed machinery of sentience itself.

Perhaps the semantics of qualitative terms is in large part implicit in the operations of mechanisms of sensory discrimination. A search for semantic regularities takes us in short order out of the civilized realm of rules and conventions and into the jungle of pre-linguistic psychological capacities. Perhaps the best way to understand what the average English-speaking human means by ‘orange’ is to possess the sensory capacities of the average human, and then look at enough samples so that one can discern the gamut that falls within ‘orange’, and the gamut that falls without. A language-learner has nothing more than this to work with, so it must be enough. More formally, one needs a shared sense of qualitative similarity—those capacities of discrimination, matching, and relative similarity that order the qualities in a given sensory modality—and presentation of a sufficient number of paradigms and foils (samples within and without the extension of the term) so that one learns the correct use of the term. Any term in any natural language for any sensory quality could be learned in this fashion.

If we knock out one or another of these two conditions, we get one or the other of two big thought experiments. The first is Thomas Nagel’s ‘What is it like to be a bat?’ (1979). The basic difficulty posed by the echolocating bat, according to Nagel, is that echolocation presents the bat with experiences that have a particular qualitative character, and we humans cannot form concepts adequate to describe that character. Why can’t humans form such concepts? This sensory modality is radically unlike any of our own: ‘bat sonar, though clearly a form of perception, is not similar in its operation to any sense that we possess, and there is no reason to suppose that it is subjectively like anything we can experience or imagine’ (Nagel 1979: 168). The bat has an alien structure of qualitative similarities—of what resembles what. Not only can the bat detect stimuli that we cannot, but it will discriminate stimuli that to us are indiscriminable (such as two minutely different locations of a fluttering moth), and it will fail to discriminate stimuli that to us are readily discriminable (such as the colour of the moths). This yields a structure of qualitative similarities and differences for which we can find no analogue.

To understand what it is like to be a bat chasing a moth is to understand the structure of qualitative similarities operative in the bat. Put another way, we need to understand what classes of experiences would be qualitatively identical *for the bat*. The problem is precisely that bat echolocation has a qualitative structure that is alien to us. In it stimuli *x* and *y* resemble one another, where for us they do not. So mentioning *x* does not help us at all to understand what it is like to experience *y*. We cannot imagine what it is like, in the sense that our paradigms and our sense of qualitative similarity cannot get us to the point where for us (like the bat) *x* and *y* are qualitatively similar. We would need to acquire the bat's quality space.

Nagel goes on to define 'subjective facts' as facts that embody a particular point of view. I suggest that this talk of 'points of view' is another way of asking whose sense of similarity is germane. To experience things from the bat's point of view is to experience the resemblances that the *bat* experiences; to have *its* quality space. This is why, to adopt that point of view, we must imagine radical changes in our own mental structure; the entire constellation of similarities and differences must change.

By hypothesis the bat operates with an alien structure of qualitative similarities. We *cannot* sense the matches that it does. Allow me to mangle Wittgenstein. If the bat had words for its qualia, we could not learn them. More precisely, we could not learn successfully to apply the words in direct, first-person fashion. To be sure, with measuring instruments and calculators of sufficient speed and capacity, we could simulate the sensory system of the bat, and laboriously calculate the parameters for the application of a given term. Reading the meter, we could then apply the word. But *our* internal meters—our sensory systems—are just not built that way, and the calculations do not correspond to anything going on within our nervous systems, so without those instruments and calculators we would be at a loss. In that sense we *could not* learn the words for bat qualia.

Now if 'to know what it is like' to be a bat is not just a matter of know-how, or of knowledge by acquaintance, then it requires that one make some judgement. That in turn requires that one deploy concepts (or at least terms). But by the preceding, humans could not learn to use the needed terms. It follows that there is a sense in which we cannot know what it is like to be a bat. We could not natively deploy the concepts that a bat might use to describe its qualia. We do not share the structure of qualitative similarities on which those concepts rely.

So we can agree with many of the premisses, and assign an agreeable interpretation to many of Nagel's claims about our abilities to form concepts of alien minds. For example, he says: 'We are forced, I think, to conclude that all these creatures have specific experiences which cannot be represented by any mental concepts of which we could have first-person understanding' (Nagel 1986: 24). If by 'have first-person understanding' Nagel means 'acquire the ability natively to deploy' or 'have the capacity for direct and unstudied observational use', then I agree. But this does not show that there is some alien *concept* whose content is of a different order from objective concepts. The reason we could not have 'first-person understanding' of the concepts of phenomenal experience of an alien mind is simply that we do not have the same structure of qualitative similarities as the alien mind, the same quality space. As Nagel says, 'our structure does not permit us to operate with concepts of the requisite type' (Nagel 1979: 171). Since use of the terms relies on native abilities to sense similarities among stimuli, if we lack such abilities, we cannot pick out the extension of the term without our instruments.

But I think this admission is a harmless one; it does not entail that there is some metaphysical fact which is perpetually beyond our comprehension. Nagel goes on to argue that there are 'subjective' facts accessible *only* from that subject's point of view (Nagel 1979: 172). Only a bat or something sufficiently similar to a bat could form the concepts necessary to understand what it is like to be a bat. This last inference is the problematic one.[6] Granted, we cannot be bats, nor could we learn first-person terms for bat qualia. But we can still study the similarities that structure the bat's quality space. The subject-matter of sensory resemblance is not one that is accessible *only* from one point of view; it can be studied in the various halls of psychology, physiology, and neuroscience (see Section 3.6). In studying those similarities we are studying the very same facts as are accessible to the bat from the bat's point of view.

1.4.3 The Knowledge Argument

Frank Jackson's 'knowledge argument' is a variant of Nagel's worries. It again adopts the perspective of someone studying a sensory modality without enjoying its use. But instead of concentrating on an alien modality—one structured by resemblances that we do not share—here the modality is one that the student possesses, but has never exercised. Mary, the brilliant neurophysiologist, knows every physical fact there is to know about human colour vision, about the reflectances of surfaces, about the wavelength composition of light, and so on. But she has never seen any colours. One day she is let out

[6] The critical move is from the claim that some facts 'embody' a point of view to the claim that some facts are accessible *only* from that point of view. This in turn relies on the suggestion that 'experience does not have, in addition to its subjective character, an objective character that can be apprehended from many different points of view' (Nagel 1979: 173). This latter premiss is unsupported. For a powerful reply to 'subjective facts', see esp. Lycan (1996: 45–68).

of her room, and sees red for the first time. She learns *what it is like* to see red, and what it has been like all along for all the people living their lives outside the room. But by hypothesis she already knew all the physical facts there were to know. Hence knowing what it is like to see red is knowing something other than a physical fact.

There have been many different responses to this argument (see Van Gulick 1993). The general principle is to admit that Mary would learn something, but then urge that because the word 'know' is ambiguous, accepting the premiss that Mary learns something does not entail that Mary learns some non-physical fact. For example, Lawrence Nemirow and David Lewis have argued that what she learns is know-how rather than knowledge that. She learns a new ability to place herself, at will, in a state representative of the experience (see D. Lewis 1990). This involves skills of imagination, memory, and recognition. Learning a new skill is consistent with her prior knowledge of all the facts. So even though she already knew everything (in one sense), she can still learn something.

One worry about the Nemirow–Lewis line is that what happens when Mary is let out of the room does not seem akin to what happens when you learn a new skill like juggling or bird-watching. Granted, thereafter she has various new abilities, but they all seem to be consequent upon a judgement, of the form '*that's* red', whose content cannot be identified with a cluster of skills. And the transmission seems remarkably abrupt. Does what she learns take practice? Not all the knowledge gained seems to be know-how.

I have argued that qualitative terms have an essential indexical component. To engage the mechanisms allowing unstudied direct observational use of the term, one must have some actual historical episode of a successful demonstrative identification. Otherwise even a well-honed sense of qualitative similarity will not enable you to identify any of its terms—to pick out any instances of a quality. When Mary steps out of the room, she makes her first successful demonstrative identification of a colour. The content of this identification is expressed, naturally enough, with a demonstrative: something to the effect of 'so *that's* red!'.

Why can't Mary in her room learn what she learns when she steps outside it? One explanation is that what she learns is a non-physical fact. But there is a simple alternative. In her achromatic chamber Mary cannot demonstratively identify *any* colour. This is not because colours are non-physical properties, but simply because no actual samples are present. So no matter what text or other transmissions are sent into the room, we could not make it possible for her to identify a colour demonstratively.

Now even within her room Mary would be able to calculate the verdict of a particular observer outside the room who is presented with a particular colour patch and queried whether or not it is red. Suppose our observer is Otto. By hypothesis she knows the reflectance properties of the patch, the character of the ambient illumination, the densities of the various classes of cones in Otto's eyes, and their state of adaptation. To make it interesting, perhaps the illumination is provided by energy-saving fluorescent lights with some prominent spectral lines (which give them a skewed colour-rendering), and Otto has been sitting in the sun reading, while wearing pink sunglasses. Nevertheless, Mary could calculate the product of ambient and reflectance spectra, obtain the resulting absorptions in Otto's three classes of cones, adjust for his adaptation state, and finish with something like a colorimetric coordinate for that stimulus occasion. Then since she knows Otto's past track record of verdicts *re* 'red?' for stimuli with just those coordinates, she could calculate the probability that Otto will assent. She could even calculate its distance in quality space from those occasions that Otto took to be paradigms of red. But even though she could complete such inferences, Mary would still be unable to say, just by looking at them herself, which objects are red. When she ascribes 'red', she must always proceed by inference and calculation. For her 'red' is not an observation term; its reference is always assisted, never direct.[7]

Suppose we let Mary out of the room, confront her with a novel stimulus, ask her whether it looks red, and set the clock ticking. She starts madly scribbling away to calculate reflectance efficiencies, wavelength spectra, and the effects on her retina, but before she is done her time runs out. Then we tell her: 'that's *red'*. She learns: 'ah, so *that's* red'. Perhaps she could not have made this judgement prior to her release, not because red is a non-physical property, but simply because demonstrative identification requires an actual sample.

Second trial. We confront her with a stimulus which is a metamer to the previous one; it reflects a very different spectrum, but that spectrum has the same effect on Mary's cones. We set the clock ticking. She glances at it and says immediately, 'that's red too'.

The demonstrative episode has given Mary new abilities. She can now rely on her built-in sense of qualitative similarity, instead of calculating the eventual verdict. We have fired up the ancient engines; for the first time, the juices start flowing through Mary's chromatic systems. 'Red' soon becomes an observation term; she can apply it in an unstudied, first-person fashion. The

[7] Mary, before she leaves her black-and-white room, can *refer* to red, and can deduce the conditions under which the term applies. But her reference is assisted, laborious, and inferential in ways that the normal use is not. Her use of the word 'red' before she steps out of the room is essentially that of a theoretical term. It is also analogous to the use she might make of a term she has coined for the qualia of a *bat*.

baptism gave Mary her bearings in quality space; now she has a touchstone for application of the term. Like the Nemirow–Lewis line, this account agrees that Mary learns something new when she leaves the room, and that she gains know-how. But on this account those abilities derive from a successful demonstrative identification. So, parting company with Nemirow and Lewis, and agreeing with Lycan (1996) and Van Gulick (1993), Mary does learn something in addition to a skill. In her a sensory system of representation gets fired up for the first time. The sample activates what for Mary is a new way of representing states of affairs in front of her sense organs, and this new way of representing gives her new skills.

Perhaps an analogy would be helpful. We need some domain in which one could learn everything propositional there is to learn, yet still be totally at sea until a successful demonstrative identification has been made. Fortunately, there is a delightful analogy to be had: the Ozma problem (see Gardner 1991). This is roughly the problem of defining ‘left’ and ‘right’. These terms are part of a family (including clockwise and counter-clockwise, east and west, north pole and south pole), any one of which can be defined in terms of the others, but all of which seem to rely ultimately on some successful demonstrative identification. In particular, suppose we begin receiving transmissions from a planet on the far side of the galaxy (so far away that no stars are mutually observable, or at least we cannot tell from the descriptions that they are mutually observable). The aliens have terms ‘lana’ and ‘rana’, which we know mean left and right, but we don’t know which is which. Similarly they also have rotational terms kana-wise and counter-kana wise, directions eana and wana, planetary poles nana and sana, and we know that these stand for one or the other of our cognate notions, but we don’t know which is which.

The Ozma problem is: can you think of any possible transmission that would allow an unambiguous translation of these terms? We might learn that if on their planet you face the sun at sunrise, you are looking towards eana, but unfortunately their planet might be rotating in the reverse direction from ours, so ‘eana’ is west, not east. We do not know whether ‘kana-wise’ means clockwise or counter-clockwise. We learn that if you curl the fingers of your lana hand, they are curled kana-wise, but ‘lana’ might be the right hand, so ‘kana-wise’ would be counter-clockwise. They could transmit pictures, but we don’t know if they normally scan them from left to right or from right to left (and of course there is no way for them to tell us). So perhaps we’re printing all their negatives backwards; or, to use the correct printer’s term, *flopped*. Call their planet ‘Flopped Earth’. (By the way, it is called the ‘Ozma’ problem because Oz was flopped: the planet (and clocks) rotated the wrong way; the sun rose in the west, which was called ‘east’, and so on.)

You learn the latitude and longitude of the visitor’s centre on Flopped

Earth, and boldly volunteer to fly there under suspended animation. You awake many centuries later after your NASA rocket malfunctions and crash-lands. You are about to step out of your rocket. Which way should you start walking to head to the visitor's centre?

I submit that your position is precisely analogous to that of Mary when she is about to step out of her black-and-white room. You have a mass of propositional knowledge but an inability to employ the terms demonstratively. But all it would take is meeting a native who would say, '*This* is your lana hand. *That* is your rana hand.' Or even more simply, the native need simply point in the correct direction ('The visitor's centre is *that* way'). From your knowledge of the latitude and longitude you could then deduce, 'Ah, so "nana" is the south pole, "eana" is east, and "kana" is clockwise.' All the terms would lock in place and you would know how to orient yourself.

You have learned something, but you have not learned some funny new fact. In a sense you already knew all the facts—anything which could be conveyed linguistically was already in that packet of transmissions. But those transmissions did not suffice to fix the reference of 'lana' and 'rana'. A successful demonstrative identification finally does the job. You learn something to the effect of 'the lana hand is *this* hand'. Such learning is not learning some spooky new kind of fact.

Analogously, Mary does learn something when she steps out of the room, but she does not learn a new non-physical fact. She learns a second, demonstrative, route to the identification of properties with which she is already professionally acquainted. She can then *use* her inborn faculties of perceptual resemblance, rather than merely mention them in her calculations. This gives her the abilities to use colour terms demonstratively and as observation terms. But the indexical element is essential to that function of the terms. It is therefore not surprising that Mary must actually confront a sample and proceed through the baptism of demonstrative identification before she can use the terms in the normal way.

Like Mary, given enough time once you had arrived on the planet, you could probably *deduce* the correct application of the various terms from what you know. You face the sun at sunrise and then can yourself complete the demonstrative identification 'eana is *that* way'. Similarly, if Mary knew all the physical facts, she could presumably deduce on her own that the first coloured object she encounters is red and not green. Deduction of the appropriate labels to apply is not the issue. Unstudied first-person use of the terms requires actual confrontation with a sample.

Use of our qualitative terms relies essentially on a shared sense of qualitative similarity, and on some successful demonstrative identifications. Nagel's worries about 'what is it like to be a bat?;' are illustrative of what happens if we do not share a sense of qualitative similarity. Jackson's worries about Mary in the achromatic chamber show the need for successful demonstrative identifications. We could not successfully deploy first-person terms for bat qualia because their sense of resemblance is so radically different from our own. Mary could not make first-person observational use of colour terms because she has never made a successful demonstrative identification of any colour. Both are necessary.

1.4.4 The Explanatory Gap

Finally we come to an objection that can be formulated without use of modal notions, counterfactual conditionals, odd locutions, begged questions, or bluster. This least objectionable and most difficult of objections is Joseph Levine's 'explanatory gap'. It is difficult in part because answering it requires that one provide a successful explanation, and all parties admit that such explanations are nowhere in sight. So here, in something of a role-reversal, the respondent is forced to take off on a flight of fancy.

Levine's objection does not deny the truth of identity statements that, for example, identify a sensation of red with a particular state of the brain. The problem, Levine says, is not that these identities are false, but that they are inexplicable. In many other domains, physical science both establishes identity claims (heat is the motion of molecules, water is H_2O) and yields full and satisfying explanations of the identities. The explanations provide what Levine calls a 'bottom-up necessitation' for the identity statements. Such necessitation is a matter of ruling out alternatives; it does not require any necessary truths to be lodged within the explanans, but instead a derivation of the identity from the contingent physics and chemistry that we know. One shows that denying the identity is inconsistent with the physical story conjoined with an ordinary understanding of the words 'water' or 'heat'. But it is difficult to see how such an explanation could be given for identities of the form 'sensing redly = brain process squiggle-squiggle', where 'squiggle-squiggle' is the technical name for the process in question. The explanatory gap is encountered in attempts to explain identities of this form.

How do we explain identities in other domains? Levine suggests that we have some pre-theoretic understanding of what stuff water is, or of what heat does. In a 'quasi-analytic' stage of the proceedings, one analyses the essential features of these notions, and identifies a particular causal niche for each. Water is the substance that quenches our thirst; flows as the predominant ingredient in all inland springs, streams, and rivers; freezes into ice; condenses into clouds and falls from the sky as rain and snow; and so on. Heat is the common property of most things that can burn the skin, turn water into steam, cook our food, and so on. We identify each by identifying a causal role; water or heat is whatever occupies that role. As Levine puts it: 'our very concept of water is of a substance that plays such and such a causal role' (Levine 1993: 131).

But then when we turn to physical science and the world as it happens to be constituted, we find that there is just one thing that fills that role. It is just H_20 that quenches our thirst, flows in inland streams and rivers, freezes into ice, and so on; and it is the kinetic energy of molecules that burns the skin, boils water, cooks our food. We identify the physical occupant of the given role. From such an identification we can *derive* the identity statement. It is implied by the conjunction of the quasi-analytic description and the physical story. Suppose that there are various superficial macro properties by which we identify something as water. While it is perfectly conceivable that something other than H_20 manifest those properties, Levine suggests that if we take the physical story as a premiss—if we 'keep our chemistry constant'—then there is a sense in which it is inconceivable that H_20 fail to manifest those properties. 'There is an apparent necessity that flows from the reduction of water to H_20, a kind of necessity that is missing from the reduction of' sensations of red to brain process squiggle-squiggle (Levine 1993: 128). If the identity can be derived from the physical story and our quasi-analytic description of a causal niche, then indeed we can rule out any other alternatives, and so we achieve a 'bottom-up necessitation' even though none of the sentences employed are necessary truths.

According to Levine this approach will fail to explain why a sensation with a particular qualitative character ('sensing redly', for example) is identical to a particular brain state. There is nothing about sensing redly that necessitates its identity with that particular brain state rather than some other one. Or, given the brain state in question, it seems equally conceivable that that very brain state be a state of sensing greenly, and not redly. The identities seem to be arbitrary and inexplicable. As John Locke put it: 'the *Ideas* of sensible secondary Qualities, which we have in our Minds, can, by us, be no way deduced from Bodily Causes, nor any correspondence or connexion be found between them and those primary Qualities which (Experience shews us) produce them in us' (Locke 1975: iv. iii. 28). To understand the connection, Locke says, 'we are fain to quit our Reason, go beyond our Ideas, and attribute it wholly to the good Pleasure of our Maker' (iv. iii. 6); the particular connections are attributed to 'the arbitrary Determination of that All-wise Agent, who has made them to be, and to operate as they do, in a way wholly above our weak Understandings to conceive' (iv. iii. 13).

The objection is formidable; many of its premisses must be granted. First we should acknowledge that current science has yet to confirm any precise mind–brain identities of the sort envisioned. Even in colour science such identifications have yet to be secured. Secondly, it follows that currently we lack any explanations for such identities, and speculation about their form is just that. So all sides agree that an explanatory gap exists today. Thirdly, analysis of the meanings of ordinary language terms for red, green, and so on seems unlikely to provide the wherewithal for any variety of 'bottom-up necessitation'. In this domain there is little grist for the conceptual analysis mill; returns are paltry. For 'red' we cannot describe a uniquely identifying causal role, as we might with 'water' or 'heat'. It is unlikely that any analysans we produce for the natural language term 'red' even manages uniquely to identify its extension. From what was said above, this is only to be expected, since that semantics is largely embodied instead in the reliable operation of mechanisms of sensory discrimination. (This is why it follows that nothing in those ordinary concepts forbids spectrum inversion; the latter is 'conceivable'.) Fourthly, if we must close the gap using just these resources, it is very hard to imagine how it might be closed. Nothing in our ordinary concept of 'red' or 'green' seems sufficient to pin sensations of red or of green to particular brain states. So, in those terms, the association between sensations of red and brain process squiggle-squiggle seems indeed an arbitrary determination, wholly above our weak understandings to conceive.

If we have to close the gap in just the way that Levine proposes, it seems unlikely that it could ever be closed. But perhaps there is another way. Instead of analysing the words, we could embark on an empirical analysis of the qualities themselves. That is, analyse the mechanisms of discrimination whose reliable operation underwrites any successful application of those words. The construction and analysis of quality space *replaces* the analysis of concepts. It can serve a similar role in the closing of the gap as does conceptual analysis, but in this domain it will not provide us with a 'causal niche' occupied uniquely by a particular qualitative property. Instead we identify a 'qualitative niche': a place in the structure of qualities. The structure derives not from relations of cause and effect, but from relations of relative similarity, matching, and discriminability. We should also deny that the resulting structural definite description, whatever it is, is any analysis of the meaning of the words 'red' or 'green'. It identifies the 'real essence' of those qualities, but the structural definite description is not something that anyone who knows the language knows.

To close the explanatory gap, the world must be such that there is one best way that the spacing of qualities could be implemented in the nervous systems with which we are endowed. But here the asymmetries in the human colour quality space, noted in the reply to spectrum inversion, come to our rescue.[8] Once we have discerned the lumpy, asymmetric, anisotropic structure of that space, and we have detailed the home-grown capacities of our sensory nervous systems, we will see that there is just one best way to fit the two together.

With these different resources employed in this different way, it is at least conceivable that the explanatory gap can be closed. We could logically derive the identities in question from the conjunction of completed neuroscience and—not completed analysis, but—completed quality space. So I claim that we can conceive of a conceptual scheme and of empirical details under which it would be inconceivable that sensing redly be anything other than brain process squiggle-squiggle. Here is one such story. Sensing redly is conceptualized as a state with a particular qualitative character. That qualitative character is a particular place in a quality space: a qualitative niche, identified by facts of qualitative structure. Only that quality could combine with yellow to give orange, could cancel green, and so on. We form a structural definite description, identifying this niche. Future neuroscience confirms Paul Churchland's bold conjecture (1986, 1989: 102, 1995: 21–7) that a quality space is a vector space of activation patterns in particular populations of neurons. Such-and-such an activation pattern in that vector space fills just the qualitative niche needed for sensations of red. That activation pattern is brain process squiggle-squiggle. Ergo, sensing redly = brain process squiggle-squiggle. How could it be anything else?

Formal details for this flight of fancy are presented in the Appendix. Of course it is still just a schema; actually producing such an explanation requires the scheme to be fleshed out empirically on both sides of the identity sign. Undoubtedly it will not happen in this way. But just seeing how it *might* be done relieves some of the distressing vacancy of the explanatory gap.

[8] A logically analogous asymmetry—a failure in parity in the decay of cobalt 60 atoms—plays a similar role in solving the Ozma problem. Gardner (1991: 94) suggests the following transmission: 'Cool the atoms of cobalt-60 to near absolute zero. Line up their nuclear axes with a powerful magnetic field. Count the number of electrons flung out by the two ends of the axes. The end that flings out the most electrons is the end that we call "south".' Just as with the colour solid, this asymmetry would allow us to fix the reference of the word 'left', but it is not part of the *meaning* of the word. Analysis would not reveal it.