

Some logical features of feature integration

Inaugural lecture for the Italian Institute for Philosophical Studies, International School of Biophysics study program "From Neuronal Coding to Consciousness", Ischia, Italy, 12-17 October 1998.

In Werner Backhaus, (ed), *Neuronal Coding of Perceptual Systems*. New Jersey: World Scientific, *Series on Biophysics and Biocybernetics*, vol 9, 2001, ISBN 981-02-4164-X, pp 3-20.

Austen Clark
Department of Philosophy U-54
University of Connecticut
Storrs, CT 06269-2054

ABSTRACT

One of the biggest challenges in understanding perception is to understand how the nervous system manages to integrate the multiple codes it uses to represent features in multiple sensory modalities. From different cortical areas, which might separately register the sight of something red and the touch of something smooth, one effortlessly generates the perception of one thing that is both red and smooth. This process has been variously called “feature integration”, “binding”, or “synthesis”. Citing some current models and some historical precursors, this paper makes some simple observations about the logic of feature integration. I suggest that “feature conjunction” is not strictly speaking conjunction at all, but rather joint predication; and that the critical task in “binding” is not simply grouping scattered representations together, or providing them a common label, but rather identifying those that have a common subject matter—those that are *about* the same thing. If this is correct, it follows that the vocabulary of sense includes not only features but something analogous to referring terms.

Initially I thought of providing you with a comprehensive survey of all the conceptual and logical problems, large and small, standing between our understanding of the neuronal codes found in different sensory systems and our goal of explaining perceptual consciousness. We would scan all the bumps in the road, the pot-holes, the washed-out bridges, the detours, and the swamps. But the prospect of that landscape was so dreary, the arguments so bleak, that I could not force

myself to produce such a catalog. Instead I will focus on what I think is the biggest challenge: the biggest bump in the road ahead. It is so big, in fact, that it presents different aspects to practitioners of different disciplines, who look at it from different angles. I hope to convince you that these apparently different aspects are really just different faces of the same bump, and that if we all haul together we can get our buggy over it.

The bump in question is called “feature integration” by psychologists, “binding” by neuroscientists, and “synthesis” or “concretion” by philosophers. Here is a simple example. Suppose at lunch you pick an apple out of the fruit basket. You simultaneously see its colour and shape and feel its smooth skin, temperature, and weight. It emerges that the different sensory systems involved in the episode employ distinctive neuronal codes, about which we all hope to learn more in the coming days. The various sensory features (or, to use the old word, “sensible qualities”) of colour, shape, texture, and so on are coded in different ways by different parts of the nervous system. How then do we manage to perceive the one apple as both red and smooth?

What makes the problem particularly challenging is that even if we had a thorough understanding of the neuronal coding for colour and of the neuronal coding for texture, a solution could still elude us. To get to the perception of one apple as both red and smooth, one must somehow integrate the results from the separate modalities. Neither modality alone can do this. There is a logical gap that the separate stories will not enable us to cross. Schematically: we have in one modality the impression of something that has the sensible quality *F*, and in another modality the impression of something having the quality *G*. From those, how do we generate the perception of one thing that is both *F* and *G*: one apple, that is both red and smooth?

1. Posing the binding problem

Recently this question has become something of a lightning rod attracting considerable interest and excitement in psychology and neuroscience. It is the centerpiece of an influential article by Francis Crick and Christof Koch entitled “Towards a neurobiological theory of consciousness”, first published in 1990 in *Seminars in the Neurosciences*. I will discuss some of the features of that model. But the problem of “feature integration” is quite an old one, posed in a recognizably modern form in Locke’s *Essay Concerning Human Understanding*, published in 1690. Locke noticed that the some “simple ideas” of sensation were modality specific, and depended on

specific sensory organs to be “conveyed” into the mind:

Thus light and colours, as white, red, yellow, blue ... come in only by the eyes. All kinds of noises, sounds, and tones, only by the ears. The several tastes and smells, by the nose and palate. (Locke 1690, II iii 1)

These “uncompounded appearances” or simple ideas of sensible qualities Locke thought of as the raw materials of the mind, the atoms, which it can neither create nor destroy, and out of which it must fashion all the more complex ideas. If he were to use today’s terminology, Locke might call these “simple ideas of sensible qualities” *features*, and define them as follows

In this target article, a *feature* will refer to any elementary property of a distal stimulus that is an element of cognition, an atom of psychological processing. (Schyns, Goldstone, & Thibault 1998, 1)

Allow for three centuries of semantic drift, and this proposition could have come right out of Locke. Just as in Locke, different feature-families are logically distinct and independent of one another, so that, for example, colours and tastes, as properties that can exist independently of one another, neither imply nor contradict one another. Yet when we perceive particular objects, our ideas are compounds of many such qualities:

Though the qualities that affect our senses are, in the things themselves, so united and blended, that there is no separation, no distance between them; yet it is plain, the ideas they produce in the mind enter by the senses simple and unmixed. ... a man sees at once motion and colour; the hand feels softness and warmth in the same piece of wax: yet the simple ideas thus united in the same subject, are as perfectly distinct as those that come in by different senses. The coldness and hardness which a man feels in a piece of ice being as distinct ideas in the mind as the smell and whiteness of a lily; or as the taste of sugar and smell of a rose. (Locke 1690, II, ii, 1)

So how do we perceive both softness and warmth in the same piece of wax, coldness and hardness in the ice, and smell and whiteness in a lily? Locke’s answer: these perceptions, as “complex ideas”, must be created by mental operations of compounding or uniting simple ideas of sense. All the bigger mental molecules, per hypothesis, are created by repeating, comparing, and compounding simple ideas (Locke 1690, II, ii, 2). To put it as precisely as possible, the operation Locke had in mind was the compounding of two ideas into one, generating from impressions of two distinct sensible qualities the idea of one object that has both. As the citation also shows, Locke was well aware that such compounding operations applied not only across modalities, but also across feature families found within a given modality, such as texture

and temperature, both sensed by touch.

Experimental psychologists today call Locke’s compounding operation “feature integration”. According to Anne Treisman’s “feature integration theory of attention”, features in the various modalities are extracted in bottom-up, automatic, and parallel fashion. The idea is substantiated by data on search times in identification tasks. We specify a target—find a green T, for example—flash a screen full of differently coloured letters in front of the subject, and time how long it takes the subject to find the target. If a target can be identified by its unique value in some feature family—if for example it is the only green thing in a multicoloured field—it “pops out” from distractors, and search times are basically constant. They do not increase significantly as more distractors are added. This is one reason why Treisman and Gelade (1980) described feature extraction as automatic and parallel, proceeding for many objects simultaneously.

What happens if we require a compounding operation across different features? For example, we set up our experiment so that a target can only be identified by its unique combination of different features—it is the one green T in an array which includes brown T’s, green X’s, and brown X’s. It cannot be identified by colour alone (since there are other green letters) or by shape alone (since there are other T’s) but only by their conjunction. In these tasks, search times increase linearly with the number of distractors. It is as if some serial process must examine each possible target in turn so as to accept or reject it. Treisman’s hypothesis: that process is precisely focal attention. “Objects characterized by conjunctions of separable features are correctly perceived only through serial focusing of attention on each item in turn” (Treisman 1988, 210). Attention focused on a location allows the “integration” or “conjunction” of the various features that characterize that location. It makes it possible to integrate the colour with the shape.

One strong piece of evidence for the reality of feature integration is the intriguing and otherwise unexpected phenomenon of “illusory conjunction”. The compounding operations sometimes go awry. Under time pressure the processes needed to identify conjunctions of features produce false positives. In a display of green T’s, blue X’s, and red O’s, for example, subjects sometimes report seeing a blue T or a green O. With brief displays and particular task demands, this can happen in as many as one out of every three trials:

The subjects made these conjunction errors much more often than they reported a color or shape that was not present in the display, which suggests that the errors reflect genuine exchanges of properties rather

than simply misperceptions of a single object. Many of these errors appear to be real illusions, so convincing that subjects demand to see the display again to convince themselves that the errors were indeed mistakes. (Treisman 1986, 117).

One or another feature is literally misplaced: blue and T are seen as characterizing the same place, when in fact no place in the array has both. Notice that this illusion is distinct from misperceiving the character of any feature in the scene (see Prinzmetal 1995). There are indeed blue things and T's; the features present in the scene may all be correctly perceived. But some are mislocated or misplaced; there is nothing that is both blue and T.

In neuroscience Locke's problem has come to be known as the "binding" problem. What gives the problem its own special flavour within neuroscience is the discovery that the neural mechanisms for distinct sensory modalities are found in distinct regions of the central nervous system. Vision and touch depend on different bits. The fractionation proceeds even further. Various of the different features one can detect by sight—colour, motion, local contour, shape, texture, and so on—have been found similarly to be processed independently of one another, in multiple, distinct "feature maps" localized in different regions of cortex. So the same neuroanatomical problem that arises for vision and touch arises as well for the perception of colour and of shape. As Crick and Koch put it

seeing any one object often involves neurons in many different visual areas. The problem of how these neurons temporarily become active as a unit is often described as "the binding problem". As an object seen is often also heard, smelled, or felt, this binding must also occur across different modalities. (Crick & Koch 1997, 284).

Suppose activity in one cortical region represents the apple as red, another as round, a third as smooth. How do we achieve the perception of one apple, which is red, round, and smooth? One unrealistic solution is to posit "convergence": some place downstream from our three processing regions, where neural processes from them converge, and generate a unified representation. It is widely agreed that such convergence is neuroanatomically implausible, and that registering combinations of features in this fashion would tax even the large number of neurons that we possess.

The alternative is a "distributed" representation, one scattered over discrete regions of cortex. How is it that activity in the various disjoint regions manages to represent one thing as both *F* and *G*? This is perhaps the simplest and clearest kind of binding problem. It is sometimes called "property binding" (see Treisman 1996, 171); I shall

focus exclusively on it.

Unfortunately these days the term "binding" is applied to almost any kind of perceptual grouping process, and it is important not to confuse the various formulations. Here is one description from Crick and Koch:

Our experience of perceptual unity thus suggests that the brain in some way binds together, in a mutually coherent way, all those neurons actively responding to different aspects of a perceived object. In other words, if you are currently paying attention to a friend discussing some point with you, neurons in area MT that respond to the motion of his face, neurons in V4 that respond to its hue, neurons in auditory cortex that respond to the words coming from his face and possibly the memory traces associated with recognition all have to be "bound" together, to carry a common label identifying them as neurons that jointly generate the perception of that specific face. (Crick & Koch 1997, 284)

They make other references to the "unity" and "coherence" of the percept. "Binding" is defined as what generates that unity or coherence. How is it, they ask, that we "seem to have a single coherent visual picture of the scene before us?" (Crick & Koch 1997, 282). Since spatial convergence will not work, the proposed mechanism for achieving coherence is temporal: neurons in the implicated regions come to fire in rough synchrony, at frequencies of around 40 hertz. As they say, these oscillations "join together some of the existing information into a coherent percept" (Crick & Koch 1997, 288). The hypothesis is summarized as follows:

The information about a single object is distributed about the brain. There has, therefore, to be a way of imposing a temporary unity on the activities of all the neurons that are relevant at the moment. (Incidentally we see no reason at all why this global unity should require fancy quantum effects.) The achievement of this unity may be assisted by a fast attentional mechanism ... The required unity takes the form of the relevant neurons firing together. (Crick & Koch 1997, 290)

The proposed mechanism explains the "unity" of perceptual processes.

But what, pray tell, is a "unified percept"? (Notice that this sounds like a philosophical question!) It cannot mean simply having *one* representation, since the problem starts from the fact that we have multiple representations of the object, distributed in different parts of sensory cortex. If we suppose that there must be one representation that somehow gathers in the content of all those separate feature registrations, we have again endorsed convergence. Nor can we rest the definition of "unified percept" on some notion of perceiving "just one" object. Criteria for what one might count as "one object" are notoriously elastic. Strolling across the piazza, one sees exactly one

flock of pigeons, clustered on the ground; but on the other hand one sees a multitude of individual pigeons, a greater multitude of beaks, wings, eyes, feathers, and claws; and an even greater multitude of coloured, iridescent, and textured surfaces in motion. If “binding” is understood as that which generates a unified and coherent percept, and a unified and coherent percept is understood as any perception of “one” object, then the binding problem is as ill defined as the notion of “one” object. If the question we start with is that murky, the answers will be worse!

It is also risky to define “unity” or “coherence” in terms of what is *experienced* to be unified or coherent. This maneuver shifts the burden onto some definition of what it means to be *experienced as* coherent; and any such definition is likely to be even more elastic than an account of what it is to perceive “one” object. Perhaps that one flock presents a blooming buzzing confusion of disunified and incoherent percepts. Certainly when the pigeons take flight one may be greeted with the explosive impression of bits and pieces flying every which way. Nevertheless some rapidly moving shapes are perceived as coloured, and some flapping gray spots have feathery textures. The simplest kind of property binding can proceed even if one’s percepts are less than fully coherent or unified.

2. The logic of feature integration

My goal in what remains is to make some simple and I hope uncontroversial observations about the logical character of feature integration. If the mind worked as described by the models I have mentioned, then certain consequences follow. They have certain interesting logical implications about what is going on in feature integration. Notice that my claim is just a conditional, an if-then claim: if you accept these models, then certain things follow. I am confident those things *do* follow; the if-then claim is (I think) incontrovertible. But as a philosopher I am not competent to tell you whether or not you should actually *accept* either model. The current state of play in the experimental literature is beyond my ken. So perhaps the conditional, incontrovertible though it may be, is uninteresting (though I doubt it). Indeed, both models have become vastly more complicated than my over-simplified sketch has revealed. There are interesting new wrinkles about what does or does not “pop out”, on what is or is not a feature, on feature hierarchies, and on the many different *kinds* of binding problems (see Treisman 1988, 1993, 1996). But since my goal is just to derive certain logical implications from the models, it is best to focus

on the simplest instance of the mechanism in question. If I can show that these implications follow from the central and simplest kind of binding, the subsequently added wrinkles are not too worrisome.

Consider, then, the example of the apple plucked out of the fruit basket. One has an impression as of something red, and an impression as of something smooth. The challenge is to “integrate” these features, so as to yield an idea of one thing that is both red and smooth. The impressions to be integrated are found in distinct “feature maps”, in different portions of the nervous system. Think of binding, as Locke did, as the compounding of ideas. One can then ask about the logical operations required to secure such compounding. What ideas go into the compounding, what ideas come out of it, and what operations are needed in between to secure the transition? If we list the contributions of those feature maps as distinct premises, and the required output as a conclusion, the logical challenge we face can be schematized as follows:

- 1 Something is red
- 2 Something is smooth

- 3 Something is both red and smooth

As any philosopher will quickly tell you, this inference is not valid. To make the inference truth-preserving, we must insure that the red thing is the smooth thing, or at least that there is some overlap between red and smooth. Using ‘*a*’ and ‘*b*’ as names for whatever is sensed as red or smooth respectively, we need to add a third premise:

- 1 *a* is red
- 2 *b* is smooth
- 3 *a* = *b*

- 4 Something is both red and smooth

Now we have an inference that is truth preserving—an inference of this form won’t ever take us from true premises to a false conclusion—but to get it some variant of the third premises is essential. Unless we can identify the subject matter of the first premise with that of the second, we cannot logically secure the conclusion.

Such identification may be partial. That is, we don’t need the red portion and the smooth portion of the scene to be perfectly coincident, or identical, as long as there is some overlap. If ‘*a*’ names the red portion, and ‘*b*’ the smooth, then it suffices that some part of *a* = a part of *b*. This variant of the third premise suffices to get to the conclusion that something is both red and smooth. But notice that there is still an underlying identity required. We need something red such that the very

same thing is smooth.

What makes it possible to bind colour and texture together? According to both feature integration and the “property binding” model, it is that they are both perceived to be features of *the same place*. As Treisman and Gelade put it initially:

We assume that the visual scene is initially coded along a number of separable dimensions, such as color, orientation, spatial frequency, brightness, direction of movement. In order to recombine these separate representations and to ensure the correct synthesis of features for each object in a complex display, stimulus locations are processed serially with focal attention. Any features which are present in the same central “fixation” of attention are combined to form a single object. Thus focal attention provides the “glue” which integrates the initially separable features into unitary objects. (Treisman and Gelade 1980, 98)

Or, in a more recent statement, Treisman (1996, 171) says property binding is mediated by a serial scan of spatial areas, “conjoining the features that each contains and excluding features from adjacent areas.” Features are bound together because they characterize the same location: what gets “bound” to the colour feature are whatever other features are sensed to characterize the same place. The system must in one way or another detect that coincidence of locations in order for binding to proceed successfully. Treisman’s model includes a “master map” of locations, whose job is to ensure that such coincidence can be detected readily. When it goes wrong, we get “false conjunctions”.

Crick and Koch also assume that binding applies to just those features that are sensed to characterize what they call the “attended location”. Attention is directed at a distal location, a place in front of the eyes, where the stimulus is located. We must distinguish that attended location from the various locations of the feature maps that represent it. There is one place, a distal place, on which attention is focused; but focusing attention on that location somehow serves to activate the particular portions of the various feature maps that have information *about* it. The process sounds daunting, but Crick and Koch explain how the proposed thalamo-cortical oscillations could do the job.

With this we can make our schema a bit more realistic. A colour map registers the locations of hue features, and similarly for a texture map. The two premises might be rendered

- 1 Here it’s red.
- 2 There it’s smooth.

where the “here” and “there” are stand-ins for whatever characteristics

the respective feature maps employ to identify the places that are red or smooth respectively. To achieve the binding of colour and texture, to get to the perception of something that is both red and smooth, the same form of suppressed premise is necessary, namely

- 3 Here = there

or at least a partial overlap of regions, yielding

- 3 Some place here = a place there.

Without one or another identity, feature integration fails. The premises would fail to show that something is both red and smooth.

So we identify the place of one feature with the place of another. The reason this simple finding is significant to philosophers is that identifications (or the requisite identity statements) are logically interesting animals. They are, for example, an order of magnitude more complex than simple conjunctions. By “an order of magnitude more complex” I mean “cannot possibly be expressed by”. Conjunction is a truth function, true if and only if both of the statements conjoined together are true. But identity statements are of a different species altogether. Features provide us with what philosophers would call “general terms”, terms that can be predicated of multiple things. To get to identity statements we need to add a new *kind* of term, with a distinct function. These are singular terms, names or terms like names, that are used to identify things. So if feature integration works as these models propose, then within sentience itself we find capacities that fill two distinct logical functions. One is predicative: the capacity to sense red (or any other feature) both here and there. The other is referential: the capacity to identify the place that is red as the place that is smooth.

Why is this important? Logically, once we have two kinds of terms, we can formulate contents that *cannot* be expressed using just general terms and conjunction. If feature integration requires contents of this order of complexity, then it is not simply conjunction. The “vocabulary of sense”, to use an old phrase, must encompass more than features. Here is one way to show this. In his book *Perception*, Frank Jackson (1977, 65) discussed what he called the “Many Properties” problem: the problem of discriminating between scenes that both contain all the same features, but differently arranged. This is a variant of what neuroscientists later came to call the binding problem. Consider the problem of distinguishing between the following two fruit basket scenes:

- Scene 1: smooth red next to hirsute green
- Scene 2: hirsute red next to smooth green

A creature equipped merely with the capacity to discriminate smooth from hirsute and red from green will fail the test, since both scenes contain all four. Instead the creature must divvy up the features appropriately, and perceive that scene one contains something that is both red and smooth, while scene two does not. In this way Jackson's problem is a version of the binding problem.

The fact that we can distinguish between the two scenes shows that our visual system can solve the Many Properties problem. Both scenes involve two textures and two colours, and so the simple capacities to discriminate textures or colours separately would not render the scenes discriminable from one another. Instead one must have the capacity to detect the coincidence or co-instantiation of features: that the first scene contains something that is both red and smooth and something *else* that is hirsute and green. The texture and colour features are "integrated", are perceived as features of one thing. Only the particular combinations (or bindings) of features differentiate the two scenes from one another.

Now suppose we have only general terms and conjunction with which to discriminate the two scenes. Instead of spatial identifiers "here" and "there", we must recast such putative names as spatial qualitative features: something like hitherness and thitherness. These are simply conjoined to the other features found in the scene. So our first fruit basket presents

(red & smooth & hither & green & hirsute & thither)

while the second presents

(red & hirsute & hither & green & smooth & thither)

Unfortunately, the two conjunctions are precisely equivalent. If we treat the spatial character of experience in this way, we lose the capacity to distinguish between the two fruit baskets.

As Quine puts it (1992, 29), "conjunction is too loose". We must somehow focus the attribution of qualities: we need smooth and red in one place, and green and hirsute in another, distinct place. If places are reduced to mere features that get added to the list, this focusing becomes impossible, and feature integration disintegrates. Furthermore, although our conclusion describes a conjunction of features, the logic that gets us there is not conjunction. It is predication. The critical finding is that both features are features *of* the same place. Consider the difference between pronouncing both "Lo, a pebble" and "Lo, blue", and saying simply "Lo, a blue pebble". Quine says

The conjunction is fulfilled so long as the stimulation shows each of the component observation sentences to be fulfilled somewhere in the scene—thus a white pebble here, a blue flower over there. On the other hand the predication focuses the two fulfillments, requiring them to coincide or amply overlap. The blue must encompass the pebble. It may also extend beyond; the construction is not symmetric. (Quine 1992, 4)

"Pebble" is not exactly a feature, but the logic is impeccable. To get something that is both blue and a pebble, we must identify some blue place as the same place occupied by the pebble. This cannot be done with general terms and conjunction. We require some capacity to identify and discriminate the subject matters of the general terms: whether or not this one is the same as that one.

Binding can be characterized as a grouping process: its result somehow is to associate, focus, bind, or group together a collection of features (see Sajda and Finkel 1995, 268). Faced with a conjunction like

(red & hirsute & hither & green & smooth & thither)

the overwhelming temptation is to leap into the mind directly, and stick in some parentheses to indicate grouping: that (red & hirsute & hither) go together, as do (green & smooth & thither). Or we might invent nesting operations among our features, so that some can modify others. Perhaps we have (redly (hirsute)) and (smoothly (green)). While grouping is indeed the goal, it cannot be secured with such devices. The commutative and associative properties of conjunction assure that such parentheses have no semantic significance. To get the desired grouping, a logical operation of a different order is required: not conjunction or nesting, but predication. What constitutes the needed "group" is a common subject matter: that those features are all features *of* one thing. The groups are groups of predicates true *of* the same subject; boundaries between the groups are boundaries between the distinct things *to which* the features are attributed.

If the reasoning thus far is correct, we have stumbled upon a profound logical difference between features and bindings. While features are general terms, which can have multiple instances, binding requires identification, reference, the picking out of places. The work of feature integration or binding is the work of identifying the subject matters of the various feature maps. This map and that map map the same territory. If such identification relies on spatial discrimination, it is not a process of attributing additional features, or of adding more qualifications, qualities, or descriptive clauses to the already fulsome list. Even if we postulate a "master map" of locations, that map clearly

is not just another feature map like all the others. Instead its role is to help identify the subject matter of one feature map with that of another. Some place occupied by feature F = a place occupied by feature G . Such identities drive the bindings. So these models endow sentience with capacities of two distinct logical kinds: capacities not only to discriminate features or sensible qualities, but also to identify that which the qualities qualify.

3. Some implications

Now for some mollifying ways to put the conclusion (see also Clark, forthcoming). While the logical schemata are useful, I certainly do not mean to suggest that there are literally little names running around inside sensory systems, or that those systems indulge in the *sotto voce* enunciation of propositions containing both predicates and referring terms. These systems after all are non-linguistic and (in some sense) non-conceptual. Any account of what goes on inside them needs to be interpreted in terms of primitives that we would be willing to grant to any creature that can sense something. We are willing to grant sensory features and feature-dimensions to such creatures; they can sense values along ranges of contraries such as hot-warm-cold, dark-dusky-bright, red-gray-green. My argument is that if feature integration works as these models propose, then we must also endow such creatures with capacities to identify the places that are characterized by such features. Otherwise, no binding. And this second capacity to discriminate and identify that which the qualities qualify is as distinct from discrimination among features as names are from predicates. Capacities to identify are analogous to those granted by the use of referring terms, but they do not require the literal use of such terms.

In effect this conclusion puts more structure, more smarts, into sensory systems than has been classically allowed. In an ancient but unfortunate philosophical picture of how sentience proceeds, a mind that stops at mere sensation is thought to be nothing more a flux of simple and uninterpretable sensory qualities. These “raw feels” or “qualia” are unanalyzable units, atoms of pure sensation, which per hypothesis are meaningless. They gain significance only insofar as they signal or correlate with other events that are significant. A mental life of pure sensation would be nothing but a stream, a flux, a flow of such stuff. But this picture, ancient as it is, radically underestimates the sophistication needed by even the simplest animal. An animal whose mental life is a pure flux of qualities could not solve the Many Properties problem. It could not distinguish smooth red next to hirsute

green from smooth green next to hirsute red. Nor could such an animal experience two distinct simultaneous instances of the same quality.

To pass even these simple tests it must be endowed with more than just “raw feels”, more than a pure flux of qualities; it needs some capacity to discriminate their distribution in space, to identify the nasty places, and to use such identifications to help it wiggle towards the better ones. Qualia arrayed in such identifiable spatial distributions, serving as goads and guides, are no longer quite so simple, “raw”, or uninterpretable. They certainly are not so to the animal in question. One can even begin to explain why it might be advantageous for an animal to acquire some new sensory capacity. If that sensory system participates in property binding of the sort I have discussed, then it provides for the discrimination of a spatial distribution of features, so that the host can identify places that those features characterize. At the very least this allows more efficient approach and avoidance of contingencies good and bad. It thereby affords more intelligent movement through a space structured by those contingencies. Whereas on the old picture adding a new modality simply clutters the mind with another modality of junk: qualities whose only significance lies in their correlations to other stuff that already was significant.

Here is another implication. If we think of these various modalities as proceeding with more or less distinct and differently organized neuronal codes, it becomes very difficult to imagine how, from such a babel of codes, one might construct a unified or coherent percept. Such unity or coherence would seem to require a super-code, into which all the others can be translated, or in which their results could be expressed. Creation of such a super-code would require processes not found in any one of the particular modalities. So if the binding problem is cast in terms of forming a unified or coherent percept, both the end result and the processes by which it is achieved become deeply mysterious.

But suppose we allow that a sensory system includes not only capacities to discriminate among the different sensory features, but also a capacity to discriminate and identify the places characterized by those features. We can then *demystify* the binding problem. It is nothing more or less than establishing that various of one’s sensory “neuronal codes” are codes *about* the same things. The problem is to determine whether or not this neuronal code and that neuronal code have the same subject matter. To bind A and B is to establish that A and B are codes (perhaps in different modalities) that are both about the same subject; it is to identify the subject matter of A with that of B . In the core case of property binding, which I have discussed, they characterize the same

place; the two features are coinstantiated.

The problem of property binding is thus what philosophers call a problem of *identification*: of determining whether this representation and that one are both about the same subject matter. This problem is neither trivial nor confused. How the nervous system manages to identify that this portion of this feature map and that portion of that feature map are maps *of* the same territory is a fascinating and difficult empirical problem. The tender mercies of philosophers will neither solve nor dissolve this problem. In fact if the reasoning about logical features of feature integration is correct, then “binding problems” are inevitable once we invoke the terminology of neuronal codes. Binding is not some additional mysterious problem that we face only when we try to get to consciousness, but instead is intrinsic to any model in which different codes are used to represent different features of the same place. The only “unity” or “coherence” required of those scattered and sundry representations is that they are all *about* the same subject matter. We do not need to collect all those representations into one place. We do not need to form one unified super-code that includes all their contributions. We do not need mysterious processes creating some coherent percept into which they can all be neatly slotted. The only unity required is the unity of that to which they refer: that they all have the same subject matter.

4. Binding and Bishop Berkeley

It is interesting to speculate on what happens if we lack the machinery of binding that has just been described: if we lack the capacity to attribute two distinct sensible qualities to one thing, and find our abilities downgraded and reduced from predication to mere conjunction. Fortunately these speculations were completed for us, in 1713, by George Berkeley, and we can simply view the sad results.

Berkeley accepted many of the tenets of the “way of ideas”, which he had learned from Locke, but he famously took issue with the notion that those ideas refer to some mind-independent “material substance”, in which the various sensible qualities inhere. Locke’s handling of the notion was, admittedly, somewhat obscure: “material substance” is that to which the various sensible qualities are attributed, and in which they inhere, but as a substance it exists independently of all our ideas about it, and it is something distinct from all the qualities we perceive it to have. What is it other than those qualities? All Locke can say is that it is something, I know not what (Locke 1690, II, xxiii, 3).

Berkeley argued that not only is there is no such thing as matter, but

that there can be no such thing; that the very notion of “material substance” is self contradictory. He heaves it out of his ontology, leaving just spirits and ideas, or minds and sensible qualities. The result is Berkeleyan immaterialism. The only substances—the only things that can exist independently—are spirits. Ideas are accidents, modifications of spirits: states of mind. They are mind-dependent; their essence is to be perceived.

For my purposes it is not important to examine the rather interesting arguments that led Berkeley to these conclusions. But it is interesting to consider their consequences. What would happen if a mind actually worked this way? First, Berkeley is obliged to provide some replacement for Locke’s account of perceiving a material object. If a die is white and hard, then according to Locke there is some self same something, I know not what, that has both the colour and the texture; and that something is distinct from either the colour or the texture. The very same thing that is white is also hard. Berkeley thought this account was incoherent (Berkeley 1710, §49). A thing, on Berkeley’s account, can be nothing more than the collection of its sensible qualities. In a famous passage he says

I see this cherry, I feel it, I taste it: and I am sure *nothing* cannot be seen, or felt, or tasted: it is therefore red. Take away the sensations of softness, moisture, redness, tartness, and you take away the cherry, since it is not a being distinct from sensations. A cherry, I say, is nothing but a congeries of sensible impressions, or ideas perceived by various senses: which ideas are united into one thing (or have one name given them) by the mind, because they are observed to attend each other. (Berkeley, 1979 [1713], 81)

A cherry is just a collection or bundle of ideas. The cherry is “not a being distinct from sensations”, says Berkeley, so oddly enough our ideas of its various qualities do not refer to anything other than themselves. There is no extra-mental “object” of sensation, nothing outside the mind to which sensory ideas refer; the only such objects are bundles or collections of the sensible qualities themselves (see Berkeley 1710, §99).

The argument I gave earlier would imply that if this were so, we would lose our capacity to “bind” features across modalities. If we cannot identify that which the various sensible qualities qualify, then we could not establish that this sensory idea and that one are both ideas of the same thing. And sure enough, in Berkeley’s system, such feature integration disappears. Strictly speaking, it is an illegitimate operation. As he put in the *New Theory of Vision*,

we never see and feel one and the same object. That which is seen is one thing, and that which is felt is another. If the visible figure and extension be not the same with the tangible figure and extension, we are not to infer that one and the same thing has divers extensions. The true consequence is that the objects of sight and touch are two distinct things. (Berkeley 1709, §49)

The same moral is applied across different modalities. As Philonous explains to Hylas in the third of the *Three Dialogues*:

Strictly speaking, Hylas, we do not see the same object that we feel; neither is the same object perceived by the microscope which was by the naked eye. ... it follows that when I examine, by my other senses, a thing I have seen, it is not in order to understand better the same object which I had perceived by sight, the object of one sense not being perceived by the other senses. (Berkeley, 1979 [1713], 78).

It also follows that for Berkeley there is no one self-same *place* that two sensory modalities can both characterize. Berkeley denies that space or extension has any existence outside the mind. Our sensory ideas are not ideas of objects that exist outside the mind; and space—extension itself—must be included among those objects. Since, furthermore, ideas of touch and ideas of vision are qualitatively distinct, it follows that in Berkeley’s system it is not possible to see the same place that one touches:

there is no one self-same numerical extension, perceived both by sight and touch ... the particular figures and extensions perceived by sight, however they may be called by the same names, and reputed the same things with those perceived by touch, are nevertheless different, and have an existence very distinct and separate from them. (Berkeley 1709, §121).

So all varieties of binding that proceed by coincidence of spatial location—this visual quality and that tactual one occupying the same place—are forsworn.

If we were to form a language that clearly reflected the underlying reality, according to Berkeley, we could note associations and conjunctions of qualities, but we would have no terms referring to extra-mental objects that those qualities qualify. In a mind run on Berkeleyan principles, identification is not an operation proper to sensory ideas. They do not refer to objects other than themselves. The only such operation is association: collecting them into “congeries” or bundles, to which we then (confusingly) apply just one name. What seems to be predication gets recast as membership in a collection: the tart taste and red colour are both properties of a cherry only in the sense that they are both members of the congeries we call “cherry”. Even the use of just one name for all the members of that bundle is justifiable

only as a matter of convenience: it is to avoid the “endless confusion of names” that would otherwise result. But strictly speaking there is no one thing that has both a tart taste and a red colour. While common sense might insist that we taste and see the same cherry, what this means, according to Berkeley, is that we have a gustatory idea and a visual idea, and both ideas are members of the same bundle. Predication is replaced by membership in a bundle.

While I am no psychologist, I believe that a mind constructed on such principles and set loose would run into some problems. The most dramatic one arises with identifications across modalities. Suppose a wood stove in the kitchen sometimes glows red with heat; but other times some non-red things in the kitchen are hot, and some red things in it are not. Let us say a mind is built on “true Berkeleyan principles” if it is built on only such principles and ideas as Berkeley thinks are strictly speaking true. Could such a mind distinguish between the following two kitchen scenes?

Scene 1: one patch, both red and hot

Scene 2: one red patch; a distinct hot patch

The problem is that for Berkeley tactual qualities have only tactual locations, and visible qualities have only visual ones. Tactual and visual ideas are distinct. Perhaps we interpret a tactual location as a bundle of tactual coordinates: whatever tactual features we use to judge the spatial relations of things we touch. Similarly a visual location would be a bundle of visual qualities. But that is all that location comes to: there is no mind-independent place to which these appearances correspond.

How then do we distinguish a scene in which something is both red and hot from one containing something that is red and something *else* that is hot? Both would present

(red & visual coordinates for red & hot & tactual coordinates for hot).

Ordinarily we would say simply that in one scene there is one place that is both red and hot, while in the other the qualities occupy two distinct places. In one scene the visual coordinates *refer to* the same place as the tactual ones, and in the other scene they do not. But Berkeley cannot avail himself of this solution. The bundles of tactual and visual coordinates are qualitatively distinct ideas. As noted, he denies that they have a common object: ideas of vision and ideas of touch are never ideas of the same thing. So in truth we cannot claim that visual ideas and tactual ideas, though qualitatively distinct, might both be ideas of the same place. Without an object of sensation—without a place, distinct from ideas, which those ideas represent—it is difficult to make

sense of the notion that numerically the same place might be represented by qualitatively distinct ideas.

Does his system provide any resources that could be used to distinguish the two scenes? Somehow Berkeley must in one scene bundle together red and hot, and in the other scene place them in distinct bundles; but the resources he allows us for forming such bundles are skimpy indeed. Spatial coincidence or overlap is barred. As far as I can see, the only principles of association by which we can form such bundles are temporal: what Berkeley calls the “co-existence or succession” of ideas. We form bundles or congeries of sensible qualities not because they characterize the same place—that suggestion, he says, is incoherent—but only because they occur at the same (or successive) times.

But these resources seem clearly inadequate. Purely temporal associations cannot always secure the discrimination between our two scenes. We might do an experiment in which one scene always contains something both red and hot, and the other scene always contains something red and something else that is hot. A mind built on true Berkeleyan principles—which represents only that which is strictly speaking true—would find the two scenes to be indistinguishable. Our minds could discriminate between the two scenes. Ergo our minds are not built on true Berkeleyan principles. At the very least, they contain representations that according to Berkeley are not strictly speaking true.

The idea that ordinary physical things are in reality bundles of ideas—congeries of sensory impressions—is a very old one, and indeed it is part of the classic image of sentience that I mentioned earlier. It faces a fundamental problem. On what principles do we form such bundles? How does the mind conjure up these congeries? I said earlier that the binding problem can be considered a problem of grouping—it could be solved if you could somehow get inside the mind directly, and group together these features and not those. Now we can see that bundling is the same problem, expressed in an older terminology. On what principles does the mind bundle together some ideas and not others? It is a process logically distinct from simply associating one idea with another; it is more powerful than mere conjunction. It is rather a process of joint predication: of identifying both sensible qualities as qualities of one thing. I do not see how we can make sense of it unless, contra Berkeley, we can make sense of the distinction between ideas and the objects that those ideas represent. A mind without the capacity to grasp that two qualitatively distinct ideas are ideas of numerically the same thing could not solve the Many Properties problem. Such a creature would not last long if it had to

wend its way through our kitchen. Bundling requires not only a capacity to discriminate among the fearsome flux of sensible qualities, but also to identify that which they qualify. To bundle two ideas is to identify their subject matters. Binding, I submit, is found in that same bundle.

12 October 1998

Other reprints available at:

<http://www.uconn.edu/~wwwphil/aclark.html>

email: aclark@uconnvm.uconn.edu

Acknowledgments

I thank Don Baxter, Tom Bontly, Claudia Carello, Crawford Elder, Len Krimerman, Keya Maitra, Ruth Millikan, Jerry Shaffer, Bob Shaw, Karl Stocker, John Troyer, Michael Turvey, Sam Wheeler, and Virgil Whitmyer for their comments on an earlier version of this paper. Errors that remain are all my own.

References

- Berkeley, George (1709) *An Essay towards a New Theory of Vision*. Reprinted in Calkins 1957. (References in the text are by paragraph number.)
- Berkeley, George (1710) *A Treatise Concerning the Principles of Understanding*. Reprinted in Calkins 1957. (References in the text are by paragraph number.)
- Berkeley, George (1799 [1713]) *Three Dialogues between Hylas and Philonous*. Edited by Robert M. Adams. Indianapolis, Indiana: Hackett Publishing.
- Block, Ned, Flanagan, Owen, and Güzeldere, Güven (1997) (eds.) *The Nature of Consciousness: Philosophical Debates*. Cambridge, Massachusetts: MIT Press.
- Calkins, Mary (1957) (ed.) *Berkeley: Essay, Principles, Dialogues*. New York: Charles Scribner's Sons.
- Clark, Austen (forthcoming) *A Theory of Sentience*. Oxford: Oxford University Press.
- Crick, Francis and Koch, Christof (1997) Towards a neurobiological theory of consciousness. In Block, Flanagan, and Güzeldere (1997),

277-292. (Originally published in *Seminars in the Neurosciences* (1990) 2: 263-75.)

- Jackson, Frank (1977) *Perception: A Representative Theory*. Cambridge: Cambridge University Press.
- Locke, John (1690) *An Essay Concerning Human Understanding*. Edited by Peter Nidditch. Oxford: Clarendon Press, 1975. (References in the text are by book, chapter, and section number.)
- Prinzmetal, W. (1995) Visual feature integration in a world of objects. *Current Directions in Psychological Science* 4(3): 90-94.
- Quine, W. V. O. (1992) *Pursuit of Truth*. Revised Edition. Cambridge, Mass.: Harvard University Press.
- Sajda, Paul and Finkel, Leif H. (1995) Intermediate-level visual representations and the construction of surface perception. *Journal of Cognitive Neuroscience* 7(2): 267-91.
- Schyns, Philippe G., Goldstone, Robert L., and Thibaut, Jean-Pierre (1998) The development of features in object concepts. *Behavioral and Brain Sciences* 21: 1-54.
- Treisman, Anne and Gelade, Garry (1980) A feature-integration theory of attention. *Cognitive Psychology* 12: 97-136.
- Treisman, Anne (1986) Features and objects in visual processing. *Scientific American*. 255 (5): 114B-125.
- Treisman, Anne (1988) Features and objects: The fourteenth annual Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology* [A] 40: 201-37.
- Treisman, Anne (1993) The perception of features and objects. In *Attention: Selection, Awareness, and Control: A Tribute to Donald Broadbent*. Edited by A. Baddeley and L. Weiskrantz. Oxford: Clarendon Press, 5-35.
- Treisman, Anne (1996) The binding problem. *Current Opinion in Neurobiology* 6: 171-78.

