Attention and the Problem of Unity

Abstract

Before we see a tree or hear a lion, early sensory processing divides the features of those sensory objects. How do these features come back together so that we can perceive the tree and the lion? This is the classic problem of unity at the level of perceptual objects, solved by Treisman's Feature-Integration Theory thirty years ago. This paper tackles a different problem of unity. Within every perceptual experience we find that more important, more meaningful percepts stand out in the foreground, whereas less important, less meaningful percepts are relegated to the background. But importance and meaningfulness change as our tasks and interests change. How does our perceptual system organize early sensory processing so that it develops layers of meaning? While tackling the problem of unity at the level of perceptual experience, this paper defends Treisman's claim that early sensory processing divides stimuli into separable features that are only later re-combined into perceptual objects. It also criticizes Bayne and Chalmers' account of conscious unity and presents in its stead a form of unity that it claims to be the most universal and fundamental form of perceptual unity. This form of unity is then claimed to be brought about only through the work of voluntary attention. This paper thus claims that attention is necessary for conscious perception insofar as it provides for the most fundamental and universal form of perceptual unity. This argument can be contrasted with those that focus on the role of attention as a quantitative selection; those debates are concerned with whether any particular percept is attended to in the case that it is consciously perceived. This paper instead looks at the role of attention as a qualitative selection: perceptual experience requires a transformation of early sensory processing into a unified structure, and it is attention that provides for this transformation.

1 Introduction

This paper aims to explain the unity of conscious perception. The reader may be familiar with the concept of experiential unity from other work in philosophy and cognitive science. In philosophy, Tim Bayne's work is responsible for considerable discussion on the unity of consciousness (Bayne

2010). In cognitive science, Anne Treisman's work is responsible for at least as much discussion on the unity of perceptual objects (Treisman 1998). The unity of conscious perception is a form of unity in between these: it is the unity of the perceptual field, discoverable through phenomenological reflection.

This paper shows how attention can solve the so-called "problem of unity" for conscious perception. The problem of unity is that of explaining the origins of perceptual unity,[1] given its absence at the level of early sensory processing. This problem is based on the presumption that conscious perception shares some structural correspondence with its physical underpinnings and that each instance of conscious perception depends for its existence on the early sensory processing of sensations. Once one accepts these two claims, the problem of unity that motivates this paper can take hold.

In what follows, I will argue that attention is necessary to supply the most fundamental and universal form of unity that we find in conscious perception. Before I get to that argument, I will first motivate the problem of unity, starting with the work of Treisman on the unity of perceptual objects and then discussing Bayne's work on the unity of consciousness.

2 The Unity of Perceptual Objects

Treisman put forward the Feature-Integration Theory (FIT) with Garry Gelade in order to explain the unity of perceptual objects. This explanation was deemed necessary because of their finding that certain visual features (e.g. color and orientation) can be detected with little to no interference from other features and/or objects. Since this is thought to reveal the early sensory processing of features, before the processing of multi-feature objects, it then became a problem to explain how the features are perceived together as individual objects (the so-called "binding problem"). That is, it became a problem to explain how the color and orientation of the trunk of an ash tree in a long row of ash trees become bound to that particular trunk, given the supposition that color and orientation are first processed separately. In FIT, so-called "focal attention" is claimed to bind the features of each object through spatial coding. The idea is that focal attention brings together just those features that have

[1] "Perceptual unity" and "the unity of conscious perception" are used interchangeably.

same spatial location code as the attended spatial location. When I focally attend to the location of the trunk of an ash tree, for example, only the color that has the same location code as that particular trunk will be bound to it. Although FIT has been widely criticized in philosophy and cognitive science, I think some of its claims can withstand those criticisms. Specifically, I want to uphold the claim that certain visual features are first processed with little to no interference and only later grouped into objects. This claim plays a key role in justifying the problem of unity, which is the starting point of this paper. I thus present and defend it in this section before turning to the evidence for perceptual unity in the next.

2.1 The Feature-Integration Theory, or FIT

In 1980, Treisman published "A feature-integration theory of attention" with Garry Gelade, changing the landscape of writings on attention. Previous milestones in attention research looked at the role of attention in filtering sensory processing (primarily for auditory processing) (Broadbent 1958; Deutsch & Deutsch 1963). Treisman and Gelade's work shows that attention also transforms the character of sensory processing (primarily for visual processing), such that it is now largely taken for granted that there are two types of sensory processing—parallel and serial. As Treisman and Gelade put it:

> In our model, which we call the feature-integration theory of attention, features are registered early, automatically, and in parallel across the visual field, while objects are identified separately and only at a later stage, which requires focused attention...focal attention provides the 'glue' which integrates the initially separable features into unitary objects. (Treisman & Gelade 1980, 98)

In other words, according to Treisman and Gelade, "focal attention" transforms the parallel processing of sensory features into the serial processing of objects, effectively filtering out the features unrelated to those objects.

The evidence for FIT comes largely from visual search tasks: Treisman and Gelade found a distinction between two types of search tasks that later became the paradigm examples of parallel and serial processing. Namely, if a participant is given only one object in a visual display, whether they are asked to report on the presence of a single feature (e.g. something red) or a feature conjunction (e.g. a red circle), the response time from stimulus onset to button press is the same—around 400 ms. However, if the number of objects in the display is increased, such that the participant has to search for the feature or feature conjunction, a difference emerges: single-feature search takes around 400 ms *no matter how many objects are added to the display*, but response times in feature-conjunction search *increase linearly for each object added to the display.*

For example, a red circle in a field of green circles and squares takes no more time to detect than a red circle on its own, but a red circle in a field of green circles, green squares, and red squares takes more time to detect for each distractor that is added (i.e. each non red-circle). Treisman and Gelade reason from this that when search is based on single features the required processing takes place in parallel to the processing of other features, allowing for short reaction times, but when search is based on feature conjunctions the required processing must combine the results of the parallel processing streams, increasing the reaction time according to the number of required combinations (the total number of objects to be searched). They postulate that focal attention enables feature-conjunction search by binding the features together into multiple-feature objects, using the spatial locations of those features to do so.

The above is the standard description of FIT. However, this account is complicated by two considerations. First, to account for known evidence of parallel feature binding, Treisman and Gelade reason that parallel binding can occur for "integral" features (that act as a single feature), but not for "separable" features (Treisman and Gelade 1980, 101). Integral features include, for example, hue and saturation. Separable features, such as color and orientation, are thought to be the more fundamental perceptual features that make up perceptual objects. Thus, the general point about focal attention being necessary for object unity is not overturned by this admission.

Second, Treisman and Gelade postulate the existence of inaccurate conjunctions to account for the fact that normal perceivers have experiences of bound objects beyond the focus of attention: "Our claim is that attention is necessary for the *correct* perception of conjunctions, although unattended features are also conjoined prior to conscious perception" (Treisman and Gelade 1980, 98). These conjunctions outside of attention can result from any combination of expectation or chance, but will always be less reliable than those formed through focal attention. Importantly, the divide between accurate and inaccurate conjunctions is not meant to be sharp. The spatial spread of attention is not fixed, such that attention can be spread widely for less conjunction accuracy or narrowly for more conjunction accuracy: “Attention can be spread over a large or a small area; the narrower the focus, the more precisely located and accurately conjoined the features in that location will be” (Treisman 1988, 203). Moreover, familiar conjunctions can be formed with some reliability outside of attention, such that the inaccuracy of conjunctions formed outside of attention may only be revealed in an unfamiliar setting, where expectation is not much of an aid.

Thus, in this account, narrowly focused attention is neither necessary for conjunction nor for accuracy in conjunction, even for separable features. According to FIT, attention is necessary for new conjunctions of separable features, but not for conjunction in general.

In the next few sections I will examine some influential criticisms of FIT. Although some of these criticisms forced key changes in FIT, not one of them touches the core finding that sensory features are processed first in near parallel, before the processing of multiple-feature objects. It is this claim that is crucial to setting up the problem of unity.

2.2 The Guided Search Model

Jeremy Wolfe, Kyle Cave, and Susan Franzel expose two glitches in the early accounts of FIT. They first show that ruling out the presence of a target is more dependent on the number of distractors than detecting a target, which is more efficient (Wolfe et al. 1989, 422). In other words, the reaction times of a participant asked whether a red circle is absent from the screen are more sensitive to the number of distractors than the reaction times of a participant asked whether the red circle is present. They also show that increasing feature contrast makes feature-conjunction search nearly as efficient as single-feature search (Wolfe et al. 1989, 426). Using larger shapes, for example, can make the search for the red circle nearly immune to distractors. These findings show that the distinction between serial and parallel search is not sharp, leading Wolfe, Cave, and Franzel to hypothesize that the relatively parallel processing of features is not wholly separate from but guides the relatively serial processing of objects (Wolfe et al. 1989, 428). In other words, the more visual information about the available features, the more efficient the process, which explains both the inefficiency of search for absent features in the first case and the efficiency of search for salient features in the second. They call their hypothesis the "Guided Search Model."

Wolfe, Cave, and Franzel test their hypothesis with triple-conjunction search. FIT might be thought to predict that three-feature searches will have slower response times than two-feature searches. The Guided Search Model, on the other hand, predicts that the extra information about the target yielded by the third feature will allow for faster search. In fact, they find that "overall, triple conjunction responses are faster than standard conjunction responses...and practically independent of the number of distractors," supporting their hypothesis (Wolfe et al. 1989, 430).

Wolfe, Cave, and Franzel present a significant challenge to FIT, because FIT held that feature-conjunction search depends on focal attention. In contrast, the Guided Search Model holds that the mechanism of feature-conjunction search can be controlled by other forces, because of the evidence described above. Treisman's later work incorporates this and other evidence to suggest a mechanism of guidance between the processing of features and the processing of multiple-feature objects (Treisman 1988). She introduces the concepts of "feature maps" and "master maps": separable features are automatically organized into feature maps, which register features and their location codes, whereas a master map keeps track of all of the location codes. In keeping with the earlier account, focal attention sometimes binds the features of its objects through their spatial location, using the master map (Treisman 1988, 203). Treisman uses the imagery of a needle sewing through layers of fabric, where focal attention is the needle that sews together features from different layers of fabric, using the location codes of those features to pick out a complete object in each stitch. But feature binding can also be controlled by the features themselves:

> In previous accounts, I have left open the question how attention is controlled…The "calling" of attention by a salient feature suggests the possibility that locations in the master-map might also be selectively activated or inhibited through links downward from particular feature maps…The resulting stronger activation of the location of the unique feature in the master-map might be indistinguishable from the activation that would be produced by focused attention to that location. If so, it would produce the same consequence: namely that the

features linked to the active master-map location in all the other feature maps would be automatically accessed and conjoined. (Treisman 1988, 226)

Treisman's later account allows for the possibility that relative saliency is processed along with separable features, and that this saliency information is a partial driver of feature binding. Fitting this adjustment, recent work has found that the early visual cortex likely processes saliency information along with feature information: "Whereas the cells' 'identities' (the labeled lines to higher visual centers) code the features and locations of the underlying stimuli, according to our proposal, the cells' firing rates report the stimuli's saliences regardless of the actual features represented by the cells" (Li 2002, 10).[2] Thus, to account for the evidence put forward by Wolfe, Cave, and Franzel, the feature-binding needle must be directed, at least in part, by the relative saliency of features.

2.3 The Synchrony Hypothesis

A more radical criticism of FIT emerges from a competing theory of feature binding that attributes binding to neuronal firing synchrony. Based on Francis Crick and Christof Koch's stipulation that the synchrony of neuronal firing oscillations binds the features represented by those oscillations (Crick and Koch 1990, 263), Wolf Singer and Charles Gray hypothesize that binding *wholly* depends on neuronal firing synchrony, which I call the "synchrony hypothesis." The synchrony hypothesis is a rejection of theories like FIT that impute the source of binding (and synchrony) to top-down mechanisms, such as focal attention. In particular, it rejects FIT's commitment to spatial location codes as a necessary component of the binding process. Singer and Gray claim that location codes are overly resource-expensive in that they require neurons to keep track of not only the feature that they represent but where in space (and, perhaps, time) that feature is. They reason that dropping this requirement allows for faster shifts and more flexible integration between different brain areas, which they take to be evident (Singer & Gray 1995, 558).

Lynn Robertson defends FIT's commitment to spatial location codes against the synchrony hypothesis by looking at what happens when spatial location codes are no longer available (Robertson 2003). Namely, assuming that the parietal cortex provides the spatial location codes necessary for

[2] Li does, however, reject the notion of feature maps in her paper.

accuracy in feature binding, if FIT is right, damage to the parietal cortex should prevent accurate feature binding (Robertson 2003, 96). Supporting this conjecture, patients with damage to the parietal cortex exhibit increased rates of binding errors. As Robertson puts it:

> Direct links between these neurons are not sufficient to explain surface feature binding, because damage to the parietal lobe disrupts such binding. An external source that synchronizes neural responses between cortical feature maps seems to be necessary, and this source seems to be spatial attention mediated by the parietal lobe. (Robertson 2003, 97)

In other words, FIT will be in a better position to explain the key role of spatial location codes that is exhibited by patients with parietal damage than theories that defend the synchrony hypothesis. Because of these patients, it seems likely that even if neural synchrony aids the mechanism of binding, it is not a sufficient source of binding, contra the synchrony hypothesis.

2.4 Binding in Unilateral Neglect

Another criticism of FIT emerges from a couple of papers in a recently published collection on attention. Sebastian Watzl writes in a footnote: "I argue that [Treisman's] view is, among other things, undermined by recent findings of feature binding in the absence of attention (e.g. in Ro and Rafal (1996))" (Watzl 2011). The paper by Tony Ro and Robert Rafal that Watzl refers to describes a patient with unilateral neglect who perceives the Judd Illusion on the neglected side. Unilateral neglect is a syndrome resulting from parietal damage that manifests itself as a failure to perceive and react to stimuli on the "neglected" side. The Judd Illusion, on the other hand, is the illusion that the midpoint on a line between two unidirectional arrowheads should be further away from the open arrowhead. As the Judd Illusion relies on a particular distribution of arrowheads to generate the illusory midpoint, Ro and Rafal conclude that the patient perceives feature binding (Ro and Rafal 1996, 977). Moreover, because the patient suffers from unilateral neglect, Ro and Rafal conclude that this binding occurs without attention. Christopher Mole concurs: "neglect patients are unable to *attend* to the neglected side of space. It is this that explains their neglect. The feature binding processes that occur in neglected patients must, therefore, be feature-binding processes that occur without constituting attention" (Mole 2011, 121).

I find two reasons to set this criticism aside. First, since Treisman is not committed to the view that attention supplies all binding but only that it supplies accuracy in the binding of separable features, a counterexample to her theory would have to show that *accurate* binding can take place without attention. The purported successful binding of Ro and Rafal's single patient must be weighed against the body of evidence reviewed by Robertson that shows unilateral neglect patients suffering from higher than average binding errors for separable features.

Second, and more importantly, an influential and oft-cited account of unilateral neglect by Marcel Kinsbourne in 1987 differs from Mole's in counting attention still present in neglect: "it is no explanation...of unilateral neglect of space that it has damaged a center for 'attention'...Unilateral neglect results not from attentional deficit but from an attentional bias: imbalance in an opponent system that controls lateral orientation and action" (Kinsbourne 1987, 69). This account makes better sense of the nearly universal view that attention operates *through* the parietal cortex, rather than being based in the parietal cortex. Moreover, this stance by Kinsbourne and others—that neglect results from attentional bias rather than from attentional absence—would help to explain how it is that unilateral neglect patients can sometimes repair attentional pathologies on the neglected side with training (Robertson 1993). That is, biases, but not absences, can be re-set.

Thus, feature binding in unilateral neglect is neither compelling evidence against FIT nor a clear instance of feature-binding in the absence of attention. Ro and Rafal's finding that feature binding can take place in patients with damage to the parietal cortex is consistent with the evidence discussed in favor of FIT and in no way undermines it.

2.5 Summary and Implications

Despite the above criticisms, FIT is well equipped to explain the binding of features into objects, answering the problem of unity at the level of perceptual objects. According to this account, sensory input is automatically divided into separable features and focal attention serially binds these features into objects using spatial location codes, making this sort of binding more accurate than binding that does not make use of location codes. What is essential for setting up the problem of unity is the finding that perception begins with the division of the perceptual field into separable features, which are only later re-grouped into multiple-feature objects. Without this evidence one could claim that the unity found in conscious perception is simply a feature of the perceived world, and not in need of explanation. Supporting Treisman and Gelade's behavioral evidence for this point are the numerous articles that demonstrate early neural specificity for separable features, such as color (e.g. Hadjikhani et al. 1998) and orientation (e.g. Haynes and Rees 2005), recently reviewed by Kalanit Grill-Spector and Rafael Malach (Grill-Spector and Malach 2004). In other words, the evidence showing that neural processing is first organized in terms of features rather than objects supports Treisman and Gelade's behavioral evidence that the detection of certain features is faster and more efficient than the detection of multiple-feature objects. This fact of early sensory processing calls for an explanation for how these features are combined into objects, which is provided by FIT.

Although the discussion so far has been centered around the problem of unity as it pertains to perceptual objects, I said early on in the paper that my target is the unity of perceptual experience, in general. The unity of perceptual objects is one type of perceptual unity, but I want to claim that there is a more fundamental and universal form of perceptual unity, which is the form I aim to supply an explanation of here. As it happens, my explanation for the existence of the more fundamental and universal form of perceptual unity also relies on attention, albeit a different form of attention than that discussed by Treisman and Gelade. Treisman and Gelade find *focal attention* to be necessary for new conjunctions of separable features, whereas I find *voluntary attention* to be necessary for bringing about the unifying structure of the perceptual field. One crucial difference is that focal attention can be mimicked by purely bottom-up processes, as was admitted by Treisman in the quote above (2.2), whereas voluntary attention essentially depends on top-down processes, which I will say more about later.

In the next section I will review different forms of perceptual unity and present the case for thinking that one of these forms is the most fundamental and universal form of unity within conscious perception. To do so I will start by discussing the work of Tim Bayne and David Chalmers on reviewing candidate forms of unity that might stand in for the unity of consciousness, which is a form of unity that I take to be more general than the unity of conscious perception.

3 The Unity of Conscious Perception

Recall that the problem of unity is that of explaining the unity of conscious perception given observed disunity at the level of early sensory processing. It is worth keeping in mind that I do not mean to conflate the neural and experiential levels, but only to assume that sensation and conscious perception have manifestations in each level, and that the defining features of these mental processes will be shared by both manifestations. Thus, as I will explain below, it is the fact that the early sensory processing of sensations, at the neural level, does not manifest the essential property of subject unity that we find in conscious perception, at the experiential level, that motivates the problem of unity for this paper.

As I stated in the introduction, the problem of unity is "based on the presumption that conscious perception shares some structural correspondence with its physical underpinnings and that each instance of conscious perception depends for its existence on the early sensory processing of sensations." The necessary components of the structural correspondence that I alluded to there will be fleshed out in this section, starting with a fuller description of perceptual unity. So far we have seen that for Treisman the problem is to account for the unity of perceptual objects given the fact that the features of those objects are first processed in near parallel. That is, to perceive an object is to first process its features, separately and with little to no interference. The fact of this separation calls for a re-unification of the object, which Treisman claims is due to either the binding of focal attention, the expectation of familiarity, or chance. In this paper, the problem is not just to account for the unity of perceptual objects, but for the unity of conscious perception, of which the unity of objects is only a part. Nonetheless, the problem of unity, at the level of conscious perception, is also based on the finding that the unity of conscious perception is not supported by the early stages of sensory processing, the details of which will become clear.

So what is the unity of conscious perception? The unity of consciousness, in general, is discussed at length by philosopher Tim Bayne, both in articles and in a monograph. Although the unity of conscious perception is a form of unity less general than that of the unity of consciousness, some of Bayne's findings are relevant to understanding the unity of conscious perception. I will thus start by looking at the candidates for the unity of consciousness reviewed by Bayne before presenting the case for one of these as the most fundamental and universal form of unity in conscious perception. Once I present what I take to be the most fundamental and universal form of perceptual unity, the reader will be in a position to understand the problem of unity and how voluntary attention might solve it.

3.1 Bayne and the Unity of Consciousness

Bayne describes the unity of consciousness as an intuitive feature of consciousness that begs further characterization and explanation. In his first paper on the topic, co-authored with David Chalmers, Bayne and Chalmers write, "At any given time, a subject has a multiplicity of conscious experiences…But at the same time, the experiences seem to be tied together in a deep way. They seem to be *unified,* by being aspects of a single encompassing state of consciousness" (Bayne & Chalmers 2003, 1). In trying to determine the nature of this unity, Bayne and Chalmers specify several candidates: objectual unity, spatial unity, subject unity, and subsumptive unity. They soon set aside objectual and spatial unity as failing to characterize the type of unity they are after. Having determined subject unity to be an empty concept ("true by definition"), they discuss the notion of subsumptive unity for the rest of the paper. I will review their reasons for rejecting objectual, spatial, and subject unity before arguing that subject unity is passed over too quickly as a candidate for the unity of consciousness.

First, I echo Bayne and Chalmers in finding that objectual and spatial unity cannot fully account for the unity of conscious perception (much less consciousness, in general). This is clearer in the case of objectual unity, which is the type of unity explained by FIT: the unity of conscious perception is not just the unity of objects, but something that transcends the unity of any individual object or objects. Take, for instance, the perceptual experience of a cat on a chair or a monkey in a tree: we can perceive the relationship between cat and chair, monkey and tree without perceiving new objects that contain these objects. In general, we perceive relations between objects without necessarily perceiving the relation as a new object or part of a new, larger object. So the unity of conscious perception will require more than the unity of objects.

A candidate for the unity of conscious perception that could provide for relations between objects is spatial unity—a common space-time framework that unifies its objects. To set aside spatial unity as a candidate for the unity of consciousness, Bayne and Chalmers appeal to those conscious experiences that do not make use of space, such as emotional experiences and abstract thought (Bayne & Chalmers 4). Unlike Bayne and Chalmers, I cannot make use of such examples since I do not know of a clear instance of perceptual experience that occurs without space. Instead, I claim that spatial unity is not a candidate for the most fundamental and universal form of perceptual because changes in spatial structure cannot account for certain structural changes in conscious perception, even given an unchanging stimulus. My presumption is that whatever supplies the most fundamental and universal form of perceptual unity should be accountable for all major structural changes in conscious perception, relative to the stimulus. Take, for example, the changes in conscious perception that occur while viewing multistable percepts. A Necker cube, for instance, may be seen as a 3-D cube with a forward face in the lower right-hand corner or with a forward face in the upper left-hand corner (or a number of other things). One perceptual experience undergoes a transformation or re-organization to bring about the next. These changes within perceptual experience come with changes in 3-D spatial structure: one percept appears closer on the bottom-right, whereas the other appears closer at the top-left. However, these changes in spatial structure do not seem to drive the changes in conscious perception, nor to explain them. Instead, *we* seem to drive the change in some cases, the percept seems to drive it in others. The changes in spatial structure are experienced as a *result* of the perceptual change, not as its cause.

Beyond this experiential evidence, there is neural justification for the claim that these changes within perceptual experience are not driven by changes in spatial structure: a recent article finds equivalent fluctuations in the parietal cortex during internally and externally-induced changes in binocular rivalry (Knapen et al. 2011). That is, the parietal fluctuations that come with changes from one version of a multistable percept to the other are matched when the experimenters simply show the participant non-rivalrous percepts for the same durations. They conclude that neural activity in the parietal cortex is a result of the change, rather than its cause. That is, something other than fluctuations in the parietal cortex will be required to explain the change from one stable percept to another. Since the parietal cortex is thought to be responsible for processing spatial information, something other than changes in spatial structure will be required to explain these changes in conscious perception.

As stated above, Bayne and Chalmers move quickly from objectual and spatial unity to subsumptive unity ("two conscious states are *subsumptively unified* when they are both subsumed by a single state of consciousness"), setting aside subject unity as an empty concept: "Let us say that two conscious states are *subject unified* when they are had by the same subject at the same time...The trouble with this version of the unity thesis is that it is *trivial*. It is true by definition and tells us nothing substantive about consciousness" (Bayne & Chalmers 2003, 5). In a later paper, Bayne distinguishes between "co-consciousness" (subsumptive unity) and "co-ownership" (subject unity):

> Co-consciousness is a phenomenal relation—it is an experiential relation between experiences—but although co-ownership is a relation between phenomenal states, it isn't itself a phenomenal relation. Whether the subject of experience is a Cartesian ego, an embodied animal, or something in between, being had by the same subject of experience (at the same time) doesn't itself enter into the character of what it's like to have a set of experiences. (Bayne 2004, 225)

This is where we disagree. If a set of physical objects are "had by the same subject at the same time," then the "subject unity" of those physical objects may be trivial, and only true by definition. One could seamlessly transfer ownership of one or more of these objects to someone else, such a new person "has" them, showing that the having of these objects by a subject does little to nothing to the objects, making their "subject unity" an empty fact. Think, for instance, of a house on closing day—the mere transfer of ownership need make no difference to the house. But for a set of conscious experiences to be "had" by a subject is not relevantly similar to a set of physical objects being "had" by that subject. The physical objects, unlike conscious experiences, have a life outside of the subject, and can be "had" by other subjects. The experiences that are had by a subject cannot be had by others, at least without undergoing significant changes; ownership makes a substantial, perhaps irrevocable, difference to experience. My experience of seeing a close friend, for example, may not be replicated by your seeing my close friend, nor by your seeing your own close friend. To put it another way, the having of experience is intimately connected to the experience in a way that the having of physical objects does not make a difference to those objects.[3] This is one reason to believe that a set of experiences that are had by a subject may be connected in a way that a set of physical objects that are had by a subject may not.

Because of these considerations, I do not find subject unity, understood as the having of conscious experiences by the subject, to be a trivial sense of unity. On the contrary, I find that subject unity is the most fundamental and universal unity of conscious perception, and perhaps consciousness, in general (although I will not argue for this latter point). In the next section I will provide my reasons for finding subject unity to be the most most fundamental and universal form of perceptual unity, but first I want to say a few words on why I dismiss Bayne and Chalmers' subsumptive unity as eligible for this role.

Bayne and Chalmers' invoke subsumptive unity because of their finding that subject unity is an empty concept, a notion that I rejected above. They write, "We can say that two conscious states are *subsumptively phenomenally unified* (or simply phenomenally unified) if there is something it is like for a subject to be in both states simultaneously" (Bayne & Chalmers 2003, 11). They claim that the substance of this unity comes from its opposition to the possible scenario of a subject that has two conscious states at the same time without there being anything it is like for the subject to have these two conscious states at the same time (Bayne & Chalmers 2003, 14)`. That is, its substance is derived from its opposition to empty subject unity. As above, I would argue that this is to presuppose a mistaken account of subject unity based on principles of physical ownership or "having" that do not seem to apply to the case of experience: for a particular subject to have an experience is for that experience to have some value relative to that particular subject, such that there will be a substantive form of unity between any two experiences that are had by a subject. If this is right, then there is no difference between saying that a subject has two experiences at once and that a subject has two experiences at once that are phenomenally unified, since they are already so unified in being had by

[3] This is true even if, as Bayne points out, it is "conceptually possible to have a phenomenal state without experiencing oneself as the subject of that phenomenal state" (Bayne 2004, 235). That is, the subjectivity of experience can require that one experiences the influence of the subject without requiring that one experiences it *as* the subject's influence.

the subject (although they may be more or less unified). Thus, a complete account of subject unity turns the charge of emptiness against subsumptive unity, instead. Subsumptive unity thereby loses its appeal as a candidate for the most fundamental and universal form of unity, since it may well be derivative of subject unity.

3.2 Subject Unity

In the section above I gave reasons to think that Bayne and Chalmers were wrong to dismiss subject unity, but I did not offer much by way of positive reason to think that subject unity is the most fundamental and universal form of perceptual unity. I offer those here. It would help to first talk a little about the nature of conscious perception and perceptual content.

Conscious perception can be usefully contrasted with sensation. In philosophy, “sensation” normally refers to raw feels, such as *pain* and *redness* (Zemach 1966, 317). In cognitive science, “sensation” normally refers to the steps in sensory processing that occur prior to those of perception—the more basic sensory processing (Wolfe et al. 2006, *Introduction*). These views can be reconciled if one assumes that perceptual processing organizes sensations within a structure, such that sensations become the structural elements of conscious perception. That these structural elements are associated with the very same early processing that we previously attributed to “sensation,” now imbued with structural characteristics, allows the “rawness” of sensation to be retained in conscious perception. This fits the tendency to consider sensations raw, brute, and unstructured, but also fits a normal way of talking about perception, wherein perception essentially involves structure. Take Tyler Burge’s description of perception: “a perception—a representational perceptual state instance, or the content of a perceptual state instance—must always involve the context-dependent singular application of (general) perceptual attributives” (Burge 2010, 381). In other words, perception is essentially organized in terms of particularities of generalities: a visual percept must have a particular color on the color spectrum and a particular position in space, whereas an auditory percept must have a particular pitch and timbre of a spectrum of possible pitches and timbres. In the language I have offered so far, the “general perceptual attributives” bring structure to sensations, whereas the “singular application” makes the sensation an element of that structure.

These thoughts are in keeping with Treisman and Gelade's concept of separable features: separable features span the perceptual field, but are nonetheless assigned to particular objects. The visual field, for instance, is made up of colors, luminances, and orientations that belong to particular visual objects, whereas the auditory field is made up of pitches, timbres, and rhythms that belong to particular auditory objects.[4] As shown by FIT, one can account for our ability to tie these features to objects through a combination of focal attention, expectation, and chance: as I walk down a familiar street, it requires my focal attention to accurately tie the color of a passerby's outfit to that outfit, such that the colors of other outfits may be left to little more than chance, but I can match blueness to the sky through the expectation of familiarity.

That focal attention brings accuracy to the process of binding seems undisputed. But how does focal attention land on just the outfit that we want to better see? Our current task clearly makes a difference, since the search experiments these findings spring from are task-based (e.g. "find the red circle" versus "find the green triangle"). But "bottom-up" salience makes a difference, too, as was admitted by Treisman above (2.2); yellow outfits grab attention more reliably than gray ones, for example. How is this bottom-up saliency information and top-down task information combined in a way that guides focal attention? From what we understand about the neural representation of saliency information, relative salience corresponds with the firing rate of the very same cells that process separable features: "the cells' firing rates report the stimuli's saliences regardless of the actual features represented by the cells" (Li 2002, 10). The bottom-up process of assigning this salience is described in an influential paper by Robert Desimone and John Duncan:

> The first neural mechanisms for resolving competition we consider are those that derive from the intrinsic or learned biases of the perceptual systems towards certain types of stimuli. We describe them here as bottom-up processes, not because they do not involve feedback pathways in visual cortex (they may well do so) but because they appear to be largely

[4] The separable features of other sensory modalities are less clear. Taste, for example, is sometimes described in terms of features like "brightness," but "brightness" is not a feature that every smeller recognizes. I thus stick to the visual and auditory examples while supposing that the difference between sensation and perception also holds for other sensory modalities. For the example of taste, this would mean that taste becomes perceptual just when taste sensations become structural elements.

automatic processes that are not dependent on cognition or task demands. (Desimone and Duncan 1995, 201)

As mentioned above, these bottom-up mechanisms can mimic the effect of focal attention, which effectively increases the relative salience of attended spatial locations, objects, or features. Thus, firing rate can also be used to explain the influence of focal attention in those cases where the firing rate is responsive to top-down feedback. However, as will become important later, once it responds to top-down feedback, the firing rate is no longer a reflection of mere salience, but also reflects value to the subject through the subject's current task. That is, whether more forcefully representing a particular color is more *valuable to the subject* than more forcefully representing a particular orientation will depend on what the representations might be used *for*, and this task-relevant information is applied through top-down feedback. Without that feedback, saliency information cannot reflect value to the subject through the subject's current task. In sum, focal attention can be mimicked by bottom-up processes that determine relative salience, but top-down feedback adds more than salience: it adds information about value to the subject through the subject's current task.

The addition of value for the subject is a reason for saying that top-down feedback concerning a subject's current task *plays a useful role* in the control of focal attention and the unity of perceptual objects, which is something Treisman would likely agree with. But I go much further in this paper. I claim that feedback concerning the subject's current task is *necessary* for separable features to form the most fundamental and universal form of unity that we find in conscious perception, which is subject unity. To say that subject unity is the most fundamental and universal form of perceptual unity is to say that conscious perception is essentially structured in terms of the meaningfulness of its percepts to its subject. This structure is discoverable within conscious perception in the division of a centrally meaningful percept from a less meaningful surround (Gurwitsch 1964). Thus, I claim that because all perceptual experiences are unified with respect to the subject and the subject's current tasks and because early sensory processing does not initially take account of the subject's current tasks, feedback concerning the subject's current tasks is necessary for separable features to form the most fundamental and universal form of unity that we find in conscious perception.

Why claim that subject unity is the most fundamental and universal form of perceptual unity? This claim is based on evidence from phenomenological reflection. Take the case of your perceiving the text of this paper, through either sight or sound. I think that if you reflect on your process of consciously perceiving that you will find it to be very much like the consumption of sensory information, in which there is likely a single principle, task, or interest that guides this consumption. All that enables your understanding of the text, when your main goal is to understand the text, is in the foreground of your experience, whereas all that is irrelevant to your understanding or that might hinder it is in the background of your experience. In fact, central to every perceptual experience is that which is relevant to the task or tasks that we are currently undertaking: while reading a paper we see and hear the text as central to our experience, but while making tea we see and feel the teapot as central to our experience, while performing yogic meditation we feel and hear the breath as central to our experience, and while pouring a relaxing bath we smell and feel the water as central to our experience. The very same things that are central to our experience while performing one of these tasks might be less central or even unperceived in another: the character of my breath as I make tea might be just as absent from that experience as the teapot is during yogic meditation. Normally, our experience is centered around that which is relevant to our current task, whereas all that is unrelated to the task is either perceived as less meaningful or not consciously perceived at all. This need not be a motor task: it could be the task of absorbing the beauty of an artwork or even of de-centering experience and transcending this basic structure. But it seems essential about conscious perception that it offers us information about the world, and to be information *for us*, perceptual content must be structured in relation to our current goals, tasks, interests, or needs. In other words, the essential structure of conscious perception involves more than salience: it carries the markers of meaningfulness, or of value to the subject and the subject's current tasks. I take it that every perceptual experience contains these markers of meaningfulness to the subject, making subject unity the most universal form of perceptual unity. It is because object unity and spatial unity depend on subject unity in order to be part of a perceptual experience, whereas subsumptive unity is derivative of subject unity, that I find subject unity to be the most fundamental form of perceptual unity.

And so to get this most fundamental and universal form of perceptual unity, which I call subject unity, the separable features that we find in conscious perception must be organized according to the subject's current task, and not just according to their proper objects. This is how they are able to present layers of meaningfulness to the subject, rather than a mere relief of salience. That is, early sensory processing lacks unity both in terms of perceptual objects *and* in terms of the subject's current tasks. It lacks unity in terms of perceptual objects because it is organized in terms of separable features. It lacks unity in terms of the subject's current tasks and interests because it is assigned saliency information based on past, stable interests ("the intrinsic or learned biases of the perceptual systems towards certain types of stimuli"), but not on the subject's current interests. For instance, in vision, those stimulations that are brightest, fastest, and largest are the most salient (this is effectively modeled in Itti et al. 1998), which can be explained in terms of our common biological or evolutionary past. But, also, those visual stimulations that look like the subject's name will receive more processing resources than those stimulations that look like other words or names (Mack and Rock 1998). This is not explained through biology or evolution, but through the subject's personal history: the subject's name has more stable value for the subject than other words or names. In both cases, the first-wave organization of saliency information in terms of past, stable interests, like the organization of features, occurs irrespective of the subject's current interests. (This is sometimes described as a "bottom-up attention," in contrast with the type of attention that this paper is concerned with, which is a "top-down" or voluntary attention.) Thus, the fact that conscious perception is organized around the subject's current interests requires explanation: what is it that unifies early sensory processing in accordance with the subject's current interests? This is the problem of unity for conscious perception at its most fundamental and universal.

In this section I have provided my reasons for finding subject unity to be the most fundamental and universal form of perceptual unity and have claimed that this leaves us in need of an explanation of this unity, since early sensory processing is not organized in terms of the subject's current interests. At best, early sensory processing is organized in terms of bottom-up salience, which could be described as a reflection of past, stable interests.

4 From Attention to Unity

Once one accepts the existence of the problem of unity at the level of subject unity, the discovery that voluntary attention is the most likely solution is only a few logical steps away. In this section, I will set out the argument linking voluntary attention to the problem of unity.

First, a lemma: I want to be more specific about what I mean by "voluntary attention." The term "attention" has come to be applied to a wide range of phenomena within cognitive science and is sometimes used just to mean neural "resources" or the distribution of these resources. I avoid this broad sense of the term because it makes the task at hand too easy: it is easy to argue that attention is necessary for perceptual unity when attention just means neural resources, or the distribution of neural resources, because these are necessary for any form of neural processing. (The same is true for those who use attention to mean simply "cognitive resources," or the distribution of cognitive resources, since these are necessary for any form of cognition.) I thus use the more specific "voluntary attention" to mean "the task-dependent direction of cognitive and neural resources by the subject." This does not mean that the ultimate distribution of cognitive and neural resources is completely controlled by the subject, but only that this distribution has been influenced by the subject through top-down feedback.

Second, I claimed above that subject unity is a universal feature of perceptual experience. That is, every perceptual experience has some degree of subject unity. This unity normally manifests itself in a division of the perceptual field into layers of meaningfulness, wherein some parts of the perceptual field are experienced as more meaningful or valuable than others. In many cases, a single task or interest shapes the perceptual field, resulting in a foreground/background structure, introduced above (3.2).[5]

Third, early sensory processing does not have subject unity. This step of the argument was also discussed above (3.2). Namely, early stages of sensory processing organize sensations (as used above, see 3.2) in terms of the subject's past interests, but not the subject's current interests. Since correspondence with the subject's current interests is a minimal requirement for subject unity, the early stages of sensory processing do not have subject unity. I take this step to be uncontroversial.

[5] This feature of experience has been discussed at length by the philosopher Aron Gurwitsch (1964; 1985).

Fourth, some process is required to bring about subject unity for these sensations in order to bring about conscious perception. The idea behind this step is that early sensory processing is the causal precursor to conscious perception. That is, without the early sensory processing of sensations, there would not be conscious perception (at least in the known, natural world). Thus, since early sensory processing does not provide the structure required for conscious perception (i.e. subject unity), but is causally necessary for conscious perception, something is required to bring about this structure to enable conscious perception.[6]

Fifth, for a process to bring about subject unity for sensations is for that process to prioritize those sensations with respect to the current tasks or interests of the subject. Unifying sensations according to the current tasks or interests of the subject is relevantly similar to unifying the voices of a choir around a musical piece. To unify the voices of a choir around a musical piece, one normally calls on the voices just as they are relevant to the piece. Naturally high-pitched voices are called on for the higher pitched parts of the piece, and naturally low-pitched voices are called on for the lower pitched parts of the piece. If one manages to call on just the right voices at just the right times, then the voices will be perfectly unified around the musical piece. Similarly, to bring about subject unity for sensations, those sensations will need to be called on just as they are relevant to the subject's current tasks and interests, just when they are so relevant. But unlike the voices of a trained choir, sensations are active long before they are called on; imagine a collection of voices that are always singing, but never the same song. In that case, one could bring about unity by emphasizing the already active voices just as they are able to contribute to the musical piece. This collection of voices may never get as close to the musical piece as the choir that only sings when called on, but it will nonetheless show signs of unity. For sensations to take on subject unity, they will have to be prioritized in just this way.

Sixth, only voluntary attention can prioritize sensations with respect to the current tasks or interests of the subject. Recall that I defined voluntary attention as "the task-dependent direction of cognitive and neural resources by the subject," where the presence of voluntary attention is

[6] Note that this does not entail Representationalism, since some properties or aspects of sensation may be retained despite this imposition of a unifying structure.

established by showing that the ultimate distribution of cognitive and neural resources has been influenced by the subject through top-down feedback. The task-dependent direction of cognitive and neural resources just is prioritization according to the tasks and interests of the subject. The reason that this has to be done by the subject, in my account, is that only the subject is going to be able to accomplish this prioritization according to its current tasks and interests, because it is the subject that holds these tasks and interests. That is, to achieve *information for* a subject, we need *prioritization by* the subject. This is the reason that I said early on that voluntary attention, unlike focal attention, cannot be mimicked by bottom-up processes (2.5).

Seventh, voluntary attention is required to bring about subject unity for sensations in order to bring about conscious perception. This claim synthesizes claims four, five, and six: some process is required to bring about subject unity for sensations in order to bring about conscious perception, for a process to bring about subject unity for sensations is for that process to prioritize those sensations according to the interests of the subject, and the prioritization of sensations according to the current interests of the subject can only be performed by voluntary attention.

Therefore, the eighth and final claim, attention is necessary for conscious perception. This is based on the stipulation that subject unity is a universal feature of conscious perception that can only be brought about through the application of voluntary attention to sensations. It is not merely a causal claim, since I am not simply claiming that attention causes conscious perception by prioritizing sensations. It is a constitutive claim, since I am also arguing that conscious perception is made up of sensations that are held together in a unity that can only be provided with voluntary attention, which is what grants sensations the quality of being meaningful to the subject. Note that the scope of this necessity is grounded in the second premise; to the extent that subject unity is a universal feature of perceptual experience, voluntary attention is necessary for conscious perception.

One philosopher and scientist that predicted this conclusion is William James. James connects attention to the unity of conscious perception in his *Principles of Psychology*:

> Millions of items of the outward order are present to my senses which never properly enter into my experience. Why? Because they have no *interest* for me. *My experience is what I agree to attend to*. Only those items which I *notice* shape my mind – without selective

interest, experience is an utter chaos. Interest alone gives accent and emphasis, light and shade, background and foreground – intelligible perspective, in a word. It varies in every creature, but without it the consciousness of every creature would be a gray chaotic indiscriminateness, impossible for us even to conceive. (James 1981, 403; *emphasis original*).

James claims here that it is interest on the part of the subject that determines the content of perception, both in quantity and in quality. His claim rests on evidence of sensory selection ("Millions of items...never properly enter into my experience") together with the idea that such selection must be governed by interests. While I do not share the presumption that quantitative selection relies on voluntary attention, and thus disagree that the fact that we do not perceive all of the "items" surrounding us alone shows us that the interest of attention is the gateway to conscious perception, I share James' intuition that the interest of attention provides for certain essential features of conscious perception. I hope that I have provided enough reason here to find that intuition plausible.

5 Conclusion

In this paper I have developed an account of the unity of conscious perception that ties conscious perception to the act of attention, or the prioritization of sensations according to the subject's current interests. I have put forward the conclusion that such attention is necessary for conscious perception, at least in the known, natural world for creatures like us. This conclusion is based on the finding that subject unity is universal feature of conscious perception, that the causal precursors of conscious perception do not have this unity, and that attention is necessary to bring about such unity.

Further development of the ideas in this paper could include a detailed exploration of the various sensory modalities, to examine whether these general principles (e.g. of separable features) apply to particular cases. An exploration of non-perceptual consciousness could be used to further understand the connection between attention and conscious perception and how it differs from the connection between attention and consciousness, more generally. Finally, the divide between past and current interests, and how the current interests of the subject become assimilated by the subject in a way that they no longer requires the application of voluntary attention, deserves further examination. Put another way, the relationship between saliency information and voluntary attention should be further specified. These and other projects will have to be left for other papers, and perhaps other researchers.

References

Bayne, T. and D. Chalmers. (2003) What is the unity of consciousness? In A. Cleeremans (Ed.) *The Unity of consciousness: binding, integration, and dissociation*: 23—58.

Bayne, T. (2004) Self-consciousnes and the unity of consciousness. *The Monist.* 87(2): 224-241.

Bayne, T. (2010) *The unity of consciousness*. Oxford University Press.

Broadbent, D. E. (1958) *Perception and communication*. Pergamon Press.

Burge, T. (2010) *Origins of objectivity*. Oxford University Press.

Crick, F. and C. Koch. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences*. 2: 263—275.

Desimone, R., & Duncan, J. (1995) Neural mechanisms of selective visual attention. *Annual review of neuroscience*. *18*(1): 193-222.

Deutsch, J. A. and D. Deutsch. (1963) Attention: Some theoretical considerations. *Psychological Review*. 70(1): 80—90.

Grill-Spector, K. and R. Malach. (2004) The human visual cortex. *Annual Review of Neuroscience.* 27: 649-677.

Gurwitsch, A. (1964) *Field of consciousness.* Duquesne University Press.

Gurwitsch, A. (1985) *Marginal consciousness.* Ohio University Press.

Haynes, J. D., and Rees, G. (2005) Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature neuroscience*. *8*(5): 686-691.

Itti, L., Koch, C. and E. Niebur. (1998) A model of saliency-based visual attention for rapid scene analysis. *Pattern Analysis and Machine Intelligence, IEEE.* 20(11): 1254—1259.

James, W. (1981) *The Principles of Psychology*. Harvard University Press.

Jennings, C. D. (2012) The subject of attention. *Synthese*. 189(3): 535-554.

Kinsbourne, M. (1987) Mechanisms of unilateral neglect. In M. Jeannerod (Ed.) *Neurophysiological and neuropsychological aspects of spatial neglect.* Elsevier.

Knapen, T., Brascamp, J., Pearson, J., van Ee, R. and R. Blake. (2011) The role of frontal and parietal brain areas in bistable perception. *The Journal of Neuroscience.* 31(28): 10293—10301.

Li, Zhaoping. (2002) A saliency map in primary visual cortex. *Trends in Cognitive Sciences.* 6(1): 9-16.

Mack, A. and I. Rock. (1998) *Inattentional Blindness.* MIT Press.

Mole, C. (2011) The metaphysics of attention. In Mole, C., Smithies, D. and W. Wu (Eds.) *Attention: Philosophical and Psychological Essays.* Oxford University Press.

Hadjikhani, N., Liu, A. K., Dale, A. M., Cavanagh, P., & Tootell, R. B. (1998) Retinotopy and color sensitivity in human visual cortical area V 8. *Nature neuroscience*. *1*(3), 235-241.

Ro, T. and R. Rafal. (2006) Visual restoration in cortical blindness: Insights from natural and TMS-induced blindsight. *Neuropsychological Rehabilitation.* 16(4): 377-396.

Robertson, L. (2003) Binding, spatial attention and perceptual awareness. *Nature Reviews Neuroscience*. 4(2): 93-102.

Treisman, A. M. and G. Gelade. (1980) A feature-integration theory of attention. *Cognitive Psychology*. 12(1): 97—136.

Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology*. 40(2): 201—237.

Treisman, A. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society B*. 353(1373): 1295—1306.

Watzl, S. (2011) Attention as structuring the stream of consciousness. In Mole, C., Smithies, D. and W. Wu (Eds.) *Attention: Philosophical and Psychological Essays.* Oxford University Press.

Wolfe, J. M., Kluender, K. R., Levi, D. M., Bartoshuk, L. M., Herz, R. S., Klatzky, R. L., and Lederman, S. J. (2006) *Sensation and perception*. Sinauer Associates.

Zeki, S. M. (1978) Functional specialisation in the visual cortex of the rhesus monkey. *Nature.* 274: 423-428.

Zemach, E. M. (1966) Sensations, raw feels, and other minds. *The Review of Metaphysics.* 20(2): 317-340.