

Against Unifying Accounts of Attention.

Henry Taylor

This is the penultimate version of a paper that appeared in *Erkenntnis*

2015. 80(1): 39-54.

Correspondence: jht30@cam.ac.uk

**Abstract.**

There have recently been a number of attempts to put forth a philosophical account of the nature of attention. Many such theories aim at giving necessary and sufficient conditions for something to be attention. In this paper I will argue that any such theory must meet two criteria. Then I shall examine four prominent accounts of attention in some detail, and argue that all of them face problems meeting one or the other of the criteria. I propose an alternative view, which involves taking seriously a pluralistic approach to attention. If the position I advocate is correct, then much of the philosophical work currently carried out on attention is fundamentally misguided, as most of the prominent theories of attention currently available are based upon assumptions which should be rejected.

**1: Attention Essentialism.**

The four theories that I will examine fall within a general overarching view of attention that I call 'attention essentialism'. This position can be characterised thus:

**Attention essentialism:** There exists a set of properties or conditions,  $\alpha$  such that all and only those entities that possess  $\alpha$  are instances of 'attention'. Instantiating the set  $\alpha$  is necessary and sufficient for something to count as attention.

Two clarifications: in ascribing attention essentialism to a certain collection of thinkers, I do not mean to imply that they hold any one particular ontological account of essence. Also, the term is not intended to apply only to those thinkers that believe that we can give necessary and sufficient conditions for attention at the *neural* level.

The views that I will examine in the most detail are those of Jesse Prinz, Wayne Wu, Christopher Mole and Sebastian Watzl, all of whom attempt to give necessary and sufficient conditions for attention in a non-circular way. It is important to emphasise that, though these positions differ greatly from each other, what they all share is a commitment to attention essentialism, either implicitly or explicitly. I examine these accounts because they are particularly prominent in the literature, though they are not the *only* accounts available.

I shall present two criteria that must be fulfilled by any account of attention which is in line with attention essentialism (section 1). I shall argue that all four accounts in question have difficulty fulfilling one or the other of the criteria (sections 2-5). Then I shall argue that we should reject attention essentialism itself (section 6).

The two criteria that any attention essentialist account should meet are as follows:

- 1) *The extensional adequacy criterion*: The account should not be so lenient as to include any cases which clearly are *not* cases of attention, or be so parsimonious as to exclude any cases which clearly *are* cases of attention.
- 2) *The non-circularity criterion*: The account should not be circular, either obviously or non-obviously.

Criterion (1) is required in order to prevent the account from being false. Attention essentialism tries to give necessary and sufficient conditions for attention, so if the account fails by criterion (1) then it will have failed to do this. The extensional adequacy criterion is in place to ensure that we do not twist our concept of attention in a Procrustean way in order to fit a particular theory.

It is important to note that the word 'attention', like many concepts, may be vague in certain ways. It may not be obvious whether certain cases are cases of attention, just as it is not obvious whether dried yeast is alive or whether a lift counts as a vehicle, due to vagueness in the concepts 'alive' and 'vehicle'. We may be willing to accept that a theory wedded to attention essentialism should be allowed to dictate whether certain cases 'really are' cases of attention when it comes to these fuzzy instances. This strikes me as reasonable. However, criterion (1) can allow that an adequate theory be allowed to dictate whether fuzzy cases count as attention or not, the only restriction that it places on the theory is that it not include cases which *clearly* are not instances of attention, and that it not exclude cases which *clearly are* instances of attention.

It may be said that criterion (1) is too much to ask for. In support of this complaint, it may be said that many theories will have some exceptions, and that we should not expect one theory to encompass all and only instances of attention. However, criterion (1) is not too much to ask of a theorist who attempts to give necessary and sufficient conditions for attention *tout court*, as attention essentialists do. It may be that we will have to give up on hoping to fulfil criterion (1), but if we do, we will have given up on attention essentialism.

Criterion (2) is a constraint on any version of attention essentialism that aims to be informative and non-trivial. An extreme example of a trivial version of attention essentialism would be to say that 'attention' applies to all and only those processes that have the property of 'being attention'. Whilst that would be true, it would be a useless account. Of course, all of the thinkers that I will examine see their theories as being non-circular, and so all must obey criterion (2).

Importantly, criterion (2) states that any account must not be circular, either obviously or non-obviously. Sometimes analyses can be circular in a way that is not initially clear. One classic example is Davidson's (1969, 1970) criterion for the individuation of events, which stated that event A and event B are identical iff. they have the same causes and the same effects. The problem with this account was that Davidson's theory stated that all causes and effects are themselves events, so deferring to causes and effects as a way of individuating events assumed that these events were themselves well individuated, which rendered his account circular in a non-obvious way.

All four thinkers have developed rich and interesting accounts of attention. My aim in this paper is not to analyse *all* aspects of the theories in question, just those aspects that are relevant to the question of whether or not they fulfil the criteria outlined above.

## **2: Jesse Prinz.**

Prinz (2007a, 2010, 2011 and 2012)<sup>1</sup> has advanced his own account of attention as part of his theory of phenomenal consciousness. Prinz claims that we should look for a brain mechanism that underpins all folk psychological uses of the word 'attention' and then identify attention with it (see Prinz, 2012, pp.90ff.). Prinz argues that there is such a common brain mechanism, he says "there exists a uniform, empirically motivated account of what attention is" (2012, p.95). This brain mechanism is availability to working memory. To Prinz, all and only the information that is available to working memory is attended to (see, e.g. Prinz, 2012, pp.90ff.). Prinz claims that "the folk-psychological insight implicit in the range of phenomena that we call attention can map onto the empirical construct of availability to working memory" (2012, p.95). As Prinz says: "[w]hen a stimulus is attended, it becomes available to working memory, and if it is unattended, it is unavailable" (2012, p.93).<sup>2</sup> Prinz

also says that “this account provides the only common denominator across the range of cases that we regard as examples of attention” (2012, p.95). It is clear that, to Prinz, being available to working memory is necessary and sufficient for something to be attended.

Working memory (as Prinz understands it) is a particular memory system where information is encoded and stored. Stimuli encoded in working memory become available for reasoning systems in the brain to access, and can also be used in certain kinds of action control and are available for verbal report.<sup>3</sup> I think that Prinz’s account of attention fails to fulfil criterion (1). Specifically, Prinz’s account seems to encompass *too much*. That is, it seems to count many things that are certainly not attended to as attended to.

One of the central concepts in this account, that of ‘availability’ to working memory is a dispositional concept. For something to be ‘available’ to working memory is for it to be *able* to be encoded in working memory *if* it is required for such encoding. For something to be *available* to working memory, it need not actually be encoded, just as money in the bank can be available even if it is not withdrawn. The distinction between availability to working memory and encoding in working memory is of central importance in Prinz’s theory of consciousness (see esp. 2012, pp.99-106).

The problem is that there seem to be a lot of mental states that are *available* in this way to working memory, but which are clearly not attended to. I am thinking of what Ned Block (1995) calls ‘quiescent beliefs’, which are beliefs that we have, but of which we are not thinking at this particular time. One example is the belief that Canberra is the capital of Australia. This belief is one that you will already have had, and you have likely had it for several years. This belief is certainly available to working memory. You can report the content of the belief, and you can use the belief in reasoning, for example, you could infer

from the fact that John is going to Canberra, and that Canberra is the capital of Australia to the conclusion that John is going to Australia. The belief that Canberra is the capital of Australia is available to working memory, and it has been available to working memory for the whole time that you have been reading this, because it *could* have been encoded in working memory *if* it was required.

By Prinz's definition of attention then, this belief must have been attended to all along (because it was available to working memory all along). However, this seems wrong. It is clear that the belief was not *attended to* at all (at least, not until you read this paragraph). You are very unlikely to have been attending to the fact that Canberra is the capital of Australia until I brought it up just now, but all along it was available to working memory for you. So here we seem to have an example of something which is available to working memory, but which is not attended to. What this objection highlights is the fact that mere *availability* to working memory is too broad a notion to be an accurate account of attention.

It seems absurd to deny that this belief was available to working memory all along, and it seems equally absurd to claim that it was attended to all along. So here we have an example where attention and availability to working memory dissociate. For this reason, we have good *prima facie* reason to reject Prinz's account of attention, as it fails by criterion (1).

### **Replies.**

Prinz may argue that quiescent beliefs are not really available to working memory. How might he do this? One likely route will be to invoke his account of the neural *realisers* of availability to working memory. In his (2012, ch.4), Prinz claims that availability to working memory is realised by 'gamma synchrony'. Prinz could then say that (if we assume that availability to working memory is realised by gamma synchrony) it will be an empirical

question whether quiescent beliefs are available to working memory because it is an empirical question whether they involve gamma synchrony.

One obvious issue with this response is that it will not so much show that Prinz's account is right, as it will leave it an open empirical question whether it is right or not. However, a deeper problem with this response is that it relies on identifying availability to working memory with gamma synchrony, and to rely on this identification in response to my concern is to put the cart before the horse.

To see this, let's examine how the identification between availability to working memory and gamma synchrony is motivated in the first place. The argument draws on Lewis' (1970) method for identifying mental faculties with brain states.

Prinz's argument has the following structure:

- 1) Define 'attention' functionally (as availability to working memory).
- 2) Look at those areas of the brain that fulfil this functional role (i.e. look at those areas of the brain that do support availability to working memory).
- 3) Find the properties of those brain areas in virtue of which they fulfil this functional role.
- 4) Claim that these properties (in this case gamma synchrony) realise the role of availability to working memory (this delivers a neural definition of the functional role specified in step (2)).

Notice that the definition of 'availability to working memory' used in step (2) cannot be a *neural* definition, because the point of this argument is to help us find such a neural definition. The problem is this. *Before* we can make the identification between gamma

synchrony and availability to working memory, we must decide which areas of the brain fulfil the functional role in question, defined non-neurally (this is step (2)). Only once we have already decided which areas of the brain *do* in fact fulfil the functional role in question can we then go about making the kind of theoretical identification between this functional role (availability to working memory) and a certain neural property (gamma synchrony). But when we look at the areas of the brain that realise the role of 'availability to working memory' (defined non-neurally as in step (2)) we find that quiescent beliefs fulfil this role perfectly (before we even look to the neural details).

So, when we perform step (2) in the above argument, we conclude that quiescent beliefs *do* fulfil the role of availability to working memory. It would be a mistake at this point to attempt to claim that they do not count as available to working memory because they do not involve gamma synchrony, because that assumes that we have good reason to identify availability to working memory with gamma synchrony in the first place, which is the question at issue. If it does turn out that quiescent beliefs are not correlated with gamma synchrony, then we would effectively have undercut the motivation to accept the identification of availability to working memory with gamma synchrony in the first place. This is because *before* we make a neural identification of this kind, we must already be convinced that the neural property in question (gamma synchrony) is found in those areas of the brain that fulfils the role of being 'available to working memory' (defined without reference to the neural details, in step (2)), and it does seem as though quiescent beliefs fulfil this role.

My argument may seem abstract, but really the problem falls out of the motivation for identifying gamma synchrony with availability to working memory in the first place.

Perhaps a simpler way of putting it is this: the only reason we had to identify gamma synchrony with availability to working memory was that we find gamma synchrony in those areas of the brain that are available to working memory (defined functionally and non-neurally). But quiescent beliefs *do* fulfil this functional role perfectly well. So if it does turn out that gamma synchrony is *not* found in the areas of the brain that correlate with quiescent beliefs, then all that will happen is that we will have lost the motivation to identify availability to working memory with gamma synchrony.<sup>4</sup> The issues here are complex, but at the very least we would need to be told more about how Prinz intends to rule out quiescent beliefs from counting as ‘attended to’ on his theory.<sup>5</sup>

### **3: Extending the criticism to other accounts: Wayne Wu.**

My focus so far has been on Jesse Prinz’s theory of attention, but I think that quiescent beliefs also cause trouble for other attention essentialist accounts, notably that of Wayne Wu (2008, 2011a, 2011b and 2011c).

At the core of Wu’s account is *action*. Wu claims that attention is involved in solving the ‘Many-Many problem’ which (roughly) is the problem faced by a system which has many potential inputs, and many potential actions it could perform, and needs to choose between them. One of Wu’s examples (2011b, p.53) is an agent looking at a tool bench. There are many potential objects that the agent could act upon (the different tools) and many different actions that she could perform with each tool. This generates what Wu (2011a, p.100) calls a ‘behavioral space’: there are many potential ‘inputs’ (in our case, perceptual information about the tools), and many potential ‘outputs’ (actions that could be performed, such as picking up a hammer) so selection must occur which links a particular input to a

particular output. Attention is just such a generation of an appropriate 'linkage', it is a selection of a certain input which guides a specific output.<sup>6</sup>

Wu uses these ideas to provide a set of necessary and sufficient conditions for attention, thus:

"S's attention to X at t is S's selection of X so as to solve the Many-Many Problem present to S at t—namely, selection of X inherent in S's traversing a specific path in the available behavioural space at t." (2011a, p.109).

Or, more informally:

"If there is to be action, a specific link must be selected, one that is constituted by an input-output connection *where the former guides the latter*... Once the structure of behavioural space is thus characterised, a solution to the Many-Many Problem... plausibly identifies a form of attention—namely, the subject's selection of information (input) that guides or otherwise informs his or her response" (2011a, p.101).

Wu develops his account to include exogenous (or involuntary) attention as well, but my focus will be on the endogenous, voluntary forms that we find in the 'hammer' example. As we can see from the above quotation, the notion of 'guidance' is crucial for Wu's account, the input must *guide* the output. Wu also cashes out this idea by saying that the input must *inform* the output (2011a, p.93) and by saying that the subject must be *attuned* to certain inputs in order to respond to them (2011a, p.111). Wu intends these notions to rule out the familiar problem of deviant causal chains, as when an intentional state causes a certain bodily response apparently without the agent's participation.

The issue arises when we take seriously the idea that attention selects inputs which *guide* and *inform* a response. Return to the example of selecting a hammer to hit a nail, say. It is true that the perceptual experience of the hammer will be one appropriate input to guide the response, but on its own this is certainly not sufficient. A great many other inputs must also be guiding the response. For example, the agent's beliefs that the hammer is an appropriate tool for the task at hand, and also her belief that the hammer can be manipulated in certain ways by the agent's hand. This general feature of a great many actions has been emphasized by Burge (1997).<sup>7</sup> When I go to the shop to buy soup, my action is guided and informed (in part) by beliefs such as that the shop will sell me soup in exchange for money, that soup is nourishing, that I have the means to open the can when I get home, and so on. Without these beliefs (or some beliefs like them) my action of buying the soup would not make sense.

Given this plausible feature of much purposive intentional action, it would seem that a great many quiescent beliefs are being selected to inform and guide a particular action, and that they are being used to guide the agent in traversing a particular behavioural space at a particular time. They are being selected to build appropriate linkages between input and output in a way conducive to creating purposive action. So, by Wu's account, these quiescent beliefs will presumably count as attended to. However (as was the case with Prinz) this seems to be the wrong result: these beliefs are surely *not* attended to when we perform these actions, rather they are entirely *ignored*. When I reach for a can of soup, I am not paying attention to the can-opener I have at home, in any way at all. For this reason, quiescent beliefs cause trouble for *both* Prinz and Wu, and make their accounts stumble by criterion (1).<sup>8</sup>

#### **4: Christopher Mole.**

Mole (2011a, 2011b) has recently put forward an admirably thorough, metaphysically-oriented theory of attention. Mole is certainly an attention essentialist as I have defined the term, he explicitly rejects the view that attention is a “family resemblance” term, or is “hopelessly folksy” or “ambiguous” and then goes on to give an analysis of “attention in all its forms” (2011a, vi-vii). I will lay out some crucial aspects of Mole’s theory below and then assess it in section 4.2. I will also consider a possible modification of Mole’s theory in section 4.3.

##### **4.1: Mole’s theory.**

Mole very clearly offers necessary and sufficient conditions for attention thus:

“Let  $\alpha$  be an agent, let  $\tau$  be some task that the agent is performing, and call the set of cognitive resources that  $\alpha$  can, with understanding, bring to bear in the service of  $\tau$ ,  $\tau$ ’s ‘background set’.  $\alpha$ ’s performance of  $\tau$  displays cognitive unison if and only if the resources in  $\tau$ ’s background set are not occupied with activity that does not serve  $\tau$ ...  $\alpha$  performs  $\tau$  attentively if and only if  $\alpha$ ’s performance of  $\tau$  displays cognitive unison.” (Mole, 2011a, p.51).

Let us see how Mole defines a ‘task’:

“A subject’s ‘tasks’... are the things that the subject is in the business of doing and that she is active with. To specify the tasks in which an agent is engaged, we adopt the agent’s point of view on her own activities. Normal human tasks are such things as making a cup of tea, following a conversation, or looking for car keys” (2011a, p.52).

Cognitive resources are the available mental processes which can, if required, be dedicated to a certain task. The set of cognitive resources that can be dedicated to  $\tau$  are what Mole calls  $\tau$ 's 'background set.'<sup>9</sup> To Mole, attention can only be deployed when it is in the service of a 'task'. As we have seen, Mole also says that attention can only be deployed when no cognitive resources that can be devoted to a task are devoted to another different task.

From this, it follows if there are two tasks,  $x$  and  $y$ , and the cognitive resources that could be deployed in order to perform  $x$  are the same as those that could be deployed to perform  $y$  (and  $x \neq y$ ) then one cannot pay attention to  $x$  and  $y$  simultaneously, on Mole's theory. This is because, if we deploy any cognitive resources to  $x$  whilst we are also deploying some to  $y$ , then it will follow that some cognitive resources that *could* be dedicated to  $x$  are in fact being dedicated to  $y$ , and vice versa. Upon Mole's account of attention, neither would be attended to. I will return to this point later.

Relatedly, Mole is eager to allow for the fact that attention can be 'partial' (2011a, p.83ff.). So let us say that there is some task, say  $A$ , and that we have a set of cognitive resources (call it  $\beta$ ) that we *could* potentially dedicate to performing  $A$ . So if we are at a state of full attention in performing  $A$ , then we will be using 100% of  $\beta$ . Now consider a case of partial attention. Suppose that we are using 60% of the cognitive resources in  $\beta$  in order to perform  $A$ , when we could be using 100%, does this count as a case of attention? Mole's answer is that it depends upon what the other 40% is doing. If the other 40% is not being used *for anything at all* then it will be true that there are no cognitive resources that could be dedicated to the service of  $A$ , but which are dedicated to a task different from  $A$  so in this case (on Mole's account) the subject will count as paying attention to  $A$ . However, if some or all of the other 40% of  $\beta$  is being used in the service of a task other than  $A$ , say  $B$ , then it will

be true that some of the cognitive resources that we *could* bring to bear on A will be in the service of another task, B. So in this case we will not count as performing A attentively, on Mole's account.

#### **4.2: Problems with Mole's account.**

Having explained this aspect of Mole's theory, I shall now argue that it leads to incorrect consequences. Specifically, it seems to count some cases that are instances of attention as *not* instances of attention, and thus fails by criterion (1).

Suppose there are two tasks, P and Q (and that  $P \neq Q$ ). Suppose that the set of cognitive resources that one could use in order to perform P is the same as those that could be deployed in order to perform Q (call this set of cognitive resources  $\delta$ ). So if we performed P with full attention, we would be using 100% of  $\delta$  in the service of P. Alternatively, if we were performing Q with full attention, we would be using 100% of  $\delta$  in the service of Q.

If a subject performs P, and dedicates only 10% of the cognitive resources in  $\delta$  to the service of P, but *does not do anything at all* with the remaining 90% of  $\delta$ , then by Mole's theory she will count as paying attention to P. Even though very few cognitive resources are dedicated to P (and she is presumably performing the task extremely haphazardly), she will still count as paying attention to P, because there are no available cognitive resources that could be used to perform P, but which are dedicated to anything that is not the performance of P (because the remaining 90% of  $\delta$  are doing nothing at all). Call this case 1.

Now consider another case, case 2. In case 2 the subject is dedicating 95% of the cognitive resources in  $\delta$  to the service of P, and the remaining 5% of  $\delta$  is dedicated to the performance of Q. According to Mole's theory, in case 2 the subject will *not* count as paying attention to P at all, even though the amount of cognitive resources that are being dedicated

to P in the second case is vastly more than in case 1. This is because in case 2, some of the cognitive resources that *can* be dedicated to P are in fact being used in the service of Q. So the upshot is that in case 1, whilst the subject is dedicating only 10% of  $\delta$  to P, the subject counts as paying attention to performing P, but in case 2, she is dedicating 95% of  $\delta$  to P, but Mole's theory counts her as not paying attention to P at all.

A concrete example will help. Imagine that in case 1, a subject is driving her car very haphazardly, only dedicating 10% of her cognitive resources to driving the car, but that she is not doing anything with the remaining 90% of her cognitive resources. Normally we would describe this as a case of inattention (after all, she is not really concentrating on driving the car). Contrary to this, Mole's theory counts her as paying attention to the driving. In case 2, she is dedicating 95% of her available cognitive resources to driving, and is as a result, driving far more carefully and safely than in the first case, but she is also using some of her remaining cognitive resources that could serve the task of driving in the performance of another task (for example, she may be vaguely thinking about her dinner). By the above reasoning, Mole's theory rules that she *is* paying attention to driving in the case 1, even though she is driving so shoddily, but she is *not* paying attention *at all* to the driving in case 2, even though she is driving far more carefully and with a greater supply of cognitive resources than in case 1.

Now, this clearly seems to be the wrong result. It seems extremely odd that in case 1 we can dedicate only a few cognitive resources to a task, and perform the task extremely shoddily and haphazardly, and yet still count as paying attention to the task, and then in another case dedicate vastly more resources to the task, and perform it much better and with greater accuracy and so on, and yet in the second case we will not count as paying attention

to this task at all (by Mole's theory, case 2 will not even count as a case of partial attention). It seems that the correct result is that the subject should count as paying much *less* attention (if any) to the driving in the first case, and paying much *more* in the second case, so it looks as though Mole's theory fails by criterion (1), as it delivers what seems clearly to be the wrong result on this occasion.

### **Replies.**

Perhaps Mole's best response is to claim that in case 2, the subject is not paying attention to P or to Q at all, but she is actually paying attention to some 'wider' task, which is the conjunction of the two tasks: (P&Q). Mole says some things along these lines (2011a, pp.81-2.).

There are various problems with this reply. Firstly, it appears *ad hoc*. Secondly, the reply does not really address the worry that we originally had with Mole's theory. The problem is that Mole's theory must still count the subject as paying attention to P in the first case, and not paying attention to P at all in the second case, even though in the second case she is using far more cognitive resources in the service of P. This worry would not really have been addressed simply by introducing another possible candidate for what the subject may be paying attention to. The problem is that when asked whether the subject is 'paying attention to P' in the second case, Mole must always answer 'no', and this seems implausible.

A third worry with this reply is that it seems very odd to claim that the subject is paying attention to (P&Q) in case 2, but also to *deny* that she is actually paying attention to P or to Q individually at all. This response would force Mole to say that in case 2, the subject was paying attention to the conjunctive task (driving & thinking about dinner) but that she

was not actually paying attention to driving, or paying attention to thinking about dinner *at all*. This is not an outright logical inconsistency, but it is certainly an extremely peculiar result.

### 4.3: Cognitive Unison Redux?

Perhaps there might be a way to modify the cognitive unison theory in order to avoid my criticisms. In particular, rather than say that agents can only perform a task attentively if the cognitive resources for that task are not taken up by any other task, could Mole not say that an agent performs a task attentively iff. at least *some* of the cognitive resources for the task in question are in unison with respect to performing the task, *whether or not* the rest of the background set is doing anything.<sup>10</sup> On this theory, in case 2, the subject would count as paying attention to *both* tasks, because *at least some* cognitive resources are being dedicated to each task.

My main concern is to rebut Mole's theory, rather than similar alternatives, so I cannot offer a full analysis of this suggestion here. However, I will say this: if we accept that all it takes for someone to perform a task attentively is that they dedicate *some* cognitive resources to that task, then a great many things that we would normally see as *inattention* will count as instances of attention. Specifically, any tasks at all that are performed using any cognitive resources would count as attended to, on this view. For example, the task of 'walking' is something that we perform, and dedicate at least some cognitive resources to a lot of the time in our waking lives, but we would not wish to say that whenever we are performing that task we are constantly *paying attention* to it, in any way at all. Indeed, among the most interesting features about walking is that it can go on in the *entire absence* of

attention to it. On the revamped version of the cognitive unison theory, walking must would count as attended to.<sup>11</sup>

Consider also experimental paradigms which attempt to study *inattention* by presenting certain stimuli in conditions of distraction. Subjects are asked to *ignore* these stimuli, whilst searching for something else. Typically, of course, the agent is taken to *not* be paying attention to such stimuli. However, it is well known that processing of such (supposedly unattended) stimuli goes on at a high level in the occipito-temporal cortex (Kouider et al. 2007). This seems to show that *at least some* cognitive resources are being dedicated to processing of these stimuli, and as such they must count as attended to, on the current version of the cognitive unison theory. This runs counter to the consensus in experimental psychology, that these are cases of *inattention*. Of course, it is possible that experimental psychology has got it wrong here, but it is at least a very serious problem for a theory if it runs counter to the general view of the empirical sciences.

In defence of this version of the cognitive unison view, one could say that walking and the 'ignored' stimuli in the experimental paradigms are subject to a *little* bit of attention all of the time. We are paying attention to these things, if only a little. I find this implausible, as it would clearly be a warping of the way that attention is normally thought of, both in normal discourse and in empirical psychology. Both commonsense and empirical psychology recognise the possibility of inattention of the sort outlined above, and if these are to count for anything in our investigation of attention, it will not do to simply ignore the divisions between attention and inattention that these practices routinely make.

One core worry here is that it is a general aim of a theory of attention that it explains the difference between attentive and inattentive action and the revised cognitive unison

view is in danger of simply eliminating the latter of these. The more general, and very deep worry here is that we should be careful about twisting our concept of attention in order to suit a particular theory, rather than explaining what we normally take attention to be in folk psychology and empirical psychology.

I have criticised the theories of Prinz, Wu and Mole. I will now press on to considering the final theory under scrutiny, which is Watzl's.

### **5: Sebastian Watzl.**

Watzl (2010, 2011a, 2011b and 2011c) has put forth a phenomenologically based theory of attention, which attempts to give necessary and sufficient conditions for attention.<sup>12</sup> Watzl's view is that attention 'structures' the stream of consciousness:

'consciously attending to something consists in the conscious mental process of structuring one's stream of consciousness so that some parts of it are more central than others' (2011a, p.158).

This 'centrality' relation is complemented by the 'peripherality' relation, thus:

'The relevant structure has as its primitive the phenomenal *peripherality* relation "x is peripheral to y" ... Consider the case where you are focussing your attention only on the sound of the piano. In the corresponding attentional structure, all other parts of your experience are peripheral to your experience of that sound. It is helpful to also have a name for the converse of this relation-that x is central to y, just in case y is peripheral to x.' (2011a, p.160).

We can summarise Watzl's view thus:  $x$  is attended to over  $y$  iff.  $x$  is more central than  $y$  in consciousness. And  $x$  is more central than  $y$  iff.  $y$  is peripheral to  $x$  in consciousness. The peripherality relation is itself taken as primitive.

The worry is that attention is explained in terms of 'centrality' which is itself explained in terms of 'peripherality', which is left as primitive. But it is hard to see what something being 'central' in one's consciousness really *could* mean other than that it is attended to over something else. Equally, it is hard to really understand what it might mean for something to be 'peripheral' other than to say that other things are attended to over it. That is to say, when we ask what 'peripherality' means, we are in danger of having no clear answer other than to say that  $y$  is peripheral to  $x$  iff.  $x$  is attended to over  $y$ . Clearly such an answer will not do, as it would render the account circular, thus failing by criterion (2).

Consider someone who did not know what 'attention' was. Could one explain to such a person what Watzl's notions of 'centrality' and 'peripherality' are? This seems unlikely, it is hard to get a handle on these concepts without assuming some notion of attention. The core issue is that 'attention' is explained in terms of 'centrality' and 'peripherality' which are themselves just as obscure and ephemeral as the concept of attention itself.

Watzl does give various *examples* of the peripherality and centrality relations (in the quotations given above he uses the example of listening to a piano piece). However, this is not very helpful in telling us what the peripherality relation actually *is*. It is hard to see how this does any more good than presenting examples of something that is attended to over other things and then leaving it at that.

Is there some way that we can make the notions of centrality and peripherality more clear? It does not seem likely that these concepts could refer to *spatial* relations, because obviously we can attend to something that is not in the centre of our visual field, so if the concepts in question are intended this way, then Watzl's account would be clearly false.

It has been suggested that attention may have an influence in making phenomenal consciousness represent items in the world with a greater level of *determinacy* (see Nanay, 2010, Cohen and Dennett, 2011 and Stazicker, 2011). Perhaps, then, for something to be 'more central' than something else is for it to be represented with more determinacy than other things. However, if this is Watzl's view, then it seems open to the obvious objection that this cannot be an exhaustive account of attention, because sometimes we attend to objects which are nonetheless represented with less determinacy than those that are not the focus of our attention (see Wu, 2011c for an extended discussion of this).

What Watzl's theory appears to deliver is less an *account* of attention but rather a set of equivalence relations between various concepts which are not themselves explained, and which are left at the intuitive level. Of course, we could always substitute synonyms for the word 'attention' and then use them in the theory. We may talk of 'saliency' or of 'clarity' or 'focus' or something like that, but once again the problem will be that using those terms just is giving new labels to substitute for the word 'attention'.

### **Replies.**

In order to avoid the charge of circularity, Watzl could claim that the notions of 'centrality' and 'peripherality' are grasped on the basis of introspection, but that *we do not grasp them as attentional* per se, but as some kind of primitive structure in the phenomenal field, which we then use in our analysis of attention.<sup>13</sup>

Of course, this reply on behalf of Watzl will only work if it is plausible that we can grasp the peripherality and centrality notions in a non-attentional way, using concepts that do not presuppose the notion of attention, or some synonym. The first problem with this suggestion is that I must confess that I find it difficult to see how this could be so. I simply can't see how one might understand centrality and peripherality in a way that does not already involve attention. Recall that Watzl's method for identifying these relations is to give examples of instances of attention, and then use these as ways of isolating the relations in question. So, the concept of 'attention' is built into the way that we isolate the notions of centrality and peripherality in the first place. Indeed, far from being able to grasp these notions in a non-attentional way, it is not even clear how these notions are *different* from the concept of attention itself, rather than just being synonyms for it. Given this, I find it hard to see how we could understand the central notions in a non-attentional manner.

A second point against this response on behalf of Watzl is that even if we do take certain notions to be primitive, we can typically still say *something* about them. We can chart the relations that our primitives have to other entities, and describe them, even if we cannot give a full analysis of them. If we can grasp the notions of peripherality and centrality in a non-attentional way, then we would expect to be able to say something substantive about them, even if we cannot give a full analysis of them. But the only clear thing that we really can say about these relations is that they are phenomenal relations that emerge as a result of attending to something.

A similar response on behalf of Watzl may be to insist that I am being unfair to his theory. He may say that every theory is allowed its primitives, and his just happens to be the peripherality relation. To say that a theory has primitives is no real objection at all.

In response to this, I will say that the problem is *not* that Watzl's theory has a primitive, but rather that the primitive that it has robs the theory of explanatory force. The trouble is that the theory packs everything that was supposed to be explained into the concept 'peripheral', and then this word is left opaque. For these reasons, I find it hard to see how Watzl's theory fulfils criterion (2).

### **6: Another view.**

I have argued that four prominent versions of attention essentialism face difficulty. But enough of the bad news, it is time for me to say something positive. One option of course would be to offer another version of attention essentialism. However, I shall propose an alternative to attention essentialism itself. Attention essentialism aims to give necessary and sufficient conditions for attention. I suggest that a different plausible view would involve rejecting the hunt for such conditions. This would not only involve rejecting the individual accounts on the table, but a wholesale rejection of the background assumption that all of these philosophical theories of attention rest upon.

The approach which would emerge would thus be very different from the approach of attention essentialism. Such an approach would see the predicate 'attention' as applying to a collection of *different* entities, which need not share any core set of properties, in virtue of which all of them can be called 'attention'. The claim would be that the entities referred to by 'attention' share a family resemblance, and it is in virtue of *this* resemblance that the predicate 'attention' applies to them all.<sup>14</sup>

We would then be left with a very different picture of what attention may be like. The view would be that the word 'attention' refers to a variety of different faculties in virtue of their sharing similar properties, just as the predicate 'game' refers to a variety of different activities which are similar in various ways. What we should resist is the claim that this 'similarity' must manifest itself in a set of privileged necessary and sufficient conditions for something to count as 'attention'. This should not embarrass us, and it should not lead us to say that the predicate 'attention' is somehow useless or deficient. We should also not think that this move threatens realism about attention. We can retain realism by claiming that the truthmakers for claims about 'attention' are the entities that are bound by this family resemblance, and to which the predicate 'attention' applies. If we accept this, then we can accept that claims about attention can be literally true, but resist the pressure to identify a certain core set of properties or conditions that all and only instances of 'attention' must instantiate. We can resist attention essentialism, and still retain realism about attention. Such a view would mark an important and under-represented view in the philosophical literature on attention.<sup>15</sup>

Because this position involves a rejection of particularly prominent assumptions in the literature, it can appear quite radical. However, on the contrary, I think that such a result should not surprise us. It is quite a frequent occurrence that our predicates map on to a collection of different though similar entities in the world, which are significantly different in certain important ways. Two plausible (though controversial) examples of this occurring before are the example of 'species' in philosophy of biology (Ereshefsky, 1992 and Brigandt, 2003) and the concept 'concept' in empirical psychology (Machery, 2005 and Weiskopf, 2009).

This view naturally invites a shift towards pluralism about attention. If we accept this pluralism then we would *expect* attention essentialism to fail, because any attention essentialist theory would be attempting to locate necessary and sufficient conditions for something which is simply not suited to an analysis of that kind.

It may initially appear as though this view is simply a wholesale rejection of all of the views that I have discussed in this paper. However, if we embrace such a pluralism then in a sense we can retain the theories of Prinz, Wu, Mole and Watzl, but reassess their explanatory power in a more modest way. We can say that they all elucidate some features which plausibly can be thought to characterise *some* instances of attention, but which need not extend to attention *tout court*. So it is plausible that some instances of attention will involve making information available to working memory, some instances of attention will have important links to action control, some will involve cognitive unison, and some will doubtless involve phenomenological shifts of various kinds. So we need not reject the theories out of hand, but rather just change their explanatory target, by rejecting the attention essentialist assumption that they are based on.

Nonetheless, it is important to note the differences between the different approaches being examined in this paper. Prinz, Wu, Watzl and Mole are all attention essentialists, but their background views on the metaphysics of attention are significantly different. Prinz's view is a 'process' view, meaning that he identifies attention with a certain brain process. By contrast, Wu and Mole identify attention with a certain *way* of doing something which can be realised by different physical systems (which need not be systems of the brain). This difference is particularly emphasised by Mole (2011b), who characterises his approach as the 'adverbial' approach. Watzl's theory is obviously phenomenological, he takes attention to be

best explained at the phenomenal level. My own suggestion is different again: I suggest that we resist the temptation to fit attention into any *one* of these categories, but rather remain open to the possibility that the predicate 'attention' can apply to many different processes, faculties and phenomenological relations. The disagreement between Prinz, Wu, Mole, Watzl and myself concerns how we should even go about approaching the question 'what is attention?'

I should emphasise that I do *not* claim to have refuted the attention essentialist programme. I have not examined all of the attention essentialist accounts on the table,<sup>16</sup> and it is of course still *possible* that a new account may arise which fulfils criteria (1-2) in a satisfactory way. Rather, we can see attention essentialism as one suggestion, to which my own pluralist suggestion is a plausible rival.

## **7: Conclusion.**

Most of the philosophical theories of attention have been along essentialist lines. Against this essentialist route, I intend my arguments as a challenge: if there is good reason to accept attention essentialism, then we must be told what it is. Indeed, unless we can be shown good reason to accept attention essentialism itself, I suggest that the problems faced by the individual attention essentialist accounts give us good reason to reject attention essentialism itself, and embrace a radically different view of attention, in line with my own pluralistic suggestions.

## **Acknowledgements**

Thanks to Peter Vickers and an anonymous referee for comments on earlier drafts of the paper. Thanks to Ned Block and Bob Kentridge for useful conversation while the paper was written.

---

<sup>1</sup> See also De Brigard and Prinz (2010).

<sup>2</sup> This quotation seems to be a very clear statement of attention essentialism. I should note that Prinz sometimes uses the word 'available' and sometimes the word 'accessible' (e.g. Prinz, 2007b) but he seems to understand these as synonymous.

<sup>3</sup>See Prinz, (2011, pp.184-185). Of course working memory is a more complex phenomenon than I have presented it here (e.g. Baddeley, 2003), but this will not matter for present purposes.

<sup>4</sup> In Taylor (2013) I also discuss Prinz's identification of availability to working memory with gamma synchrony.

<sup>5</sup> Another possible response on Prinz's behalf is that much recollection of memories from long term memory is a *constructive* process. Many memories are not explicitly encoded in long term memory but are (in some sense) constructed when memory recall is instigated. Of course, this suggestion will not by itself save Prinz's theory (the long term memories will count as available to working memory whether or not the process of retrieving them is constructive or not) but Prinz could claim that attention is somehow involved in this process of *construction* when long term memories are retrieved. This is of course, an empirical claim, which will have to await further evidence, but even if it is a plausible suggestion, I do not think that it will save Prinz's theory because notice that in order for my criticism of Prinz to go through there need only be *at least one* belief that is quiescent but explicitly encoded, even if not in long term memory. It would be an extremely extravagant claim that *no* such memories exist!

<sup>6</sup> The links to philosophy of action are worked out in greater detail by Wu, especially in his (2008) and (2011b)

<sup>7</sup> Smithies (2011) discusses this issue in relation to his own account of attention.

<sup>8</sup> As they stand, the terms 'guidance' and 'inform' are vague enough to count quiescent beliefs as attended to. For this reason, a possible response from Wu may be to attempt to

---

sharpen these terms in order to exclude the quiescent beliefs from counting as attended to. I obviously cannot hope to argue that *no* such response could work, but I will say that these terms are notoriously hard to get a grip on, despite their frequent use in the philosophy of action. For this reason, the task ahead of Wu is extremely large and I hold out little hope for its success. Another response from Wu would be to claim that quiescent beliefs are relevant to the formation of *intention* rather than the deployment of physical *action*. However, this would merely force the problem back a stage: suppose an intention were formed by *deliberation* (and that deliberation counts as a kind of mental action). Then Wu's theory would count *all* the beliefs that went into the creation of this intention as attended to (because they guide the mental action of *deliberation*). However, clearly these need not always be attended to, at least not in all instances of deliberation. Thanks to an anonymous referee for discussion of this issue.

<sup>9</sup> See Mole (2011a, pp.57-60).

<sup>10</sup> Thanks to an anonymous referee for suggesting this. Wayne Wu has also suggested this to me in personal correspondence.

<sup>11</sup> Thanks to an anonymous referee for suggesting the walking example.

<sup>12</sup> In his (2011a) Watzl puts forth the theory *only* as an account of *conscious* attention, but elsewhere (2010, 2011c) he puts it forward as a theory of attention *tout court*. In any case, my criticisms will apply to Watzl's theory even if it is only read as a theory of *conscious* attention. Of course, it is relatively uncontroversial that attention can alter phenomenal consciousness in various ways (e.g. Carrasco et al., 2004 and Sergent et al., 2012).

<sup>13</sup> Thanks to an anonymous referee for suggesting this.

<sup>14</sup> This suggestion is not new, and is an idea that we find more in the psychological, rather than philosophical literature on attention, see e.g. Duncan (2006). This family resemblance view is

---

sometimes mentioned in the philosophical literature, but is usually dismissed (Mole, 2011b, vi-vii and Prinz, 2012, pp.90ff). Here it is important to notice that there are various uses of the word 'attention' in empirical psychology which are not accommodated by the attention essentialist theories here under scrutiny. A plausible example of this is 'alerting' or 'arousal', which is a variety of attention marked by increased sensitivity to external stimuli (e.g. Posner and Rothbart, 2007). All of the above theories would not allow this kind of activity to count as a variety of 'attention'. Presumably the theorists in question would be forced to claim that arousal and alerting are not 'really' kinds of attention, but is it plausible for them to claim that empirical psychologists have got it wrong here, simply because the use of the term in empirical psychology sometimes deviates from their own use of the term? Rather than become embroiled in these sticky issues over whether the scientific taxonomy which does count alerting as a kind of attention is superior, my own view simply allows that alerting is *one* variety of attention, and that there may be others. I see it as a virtue of my own view that it fits better with how the term 'attention' is used in empirical psychology than the views I have been critiquing.

<sup>15</sup> Cf. Heil (2003) for more on 'family resemblance realism', though he is not himself discussing the case of 'attention' specifically. Also see Taylor (forthcoming, §5.2).

<sup>16</sup> See e.g. Smithies (2011) and Jennings (2012) for more philosophical attention essentialist accounts.

## References

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature*. 4, 829-839.

Block, N. (1995). On a confusion about a function of consciousness. (Reprinted in Block, N., Gulzeldere, G. and Flanagan, O. (Eds.). *The Nature of Consciousness: Philosophical Debates* (pp.375-415). 1997. USA: MIT Press).

---

Brigandt, I. (2003). Species pluralism does not imply species eliminativism. *Philosophy of Science*. 70, 1305-1316.

Burge, T. (1997). Two kind of consciousness. (In N. Block, O. Flanagan and G. Guzeldere (Eds.) *The nature of consciousness: philosophical debates* (pp.427-434). Cambridge, MA: MIT Press.)

Carrasco, M., Ling, S. and Read, S. (2004). Attention alters appearance. *Nature*. 7, 308-313.

Cohen, M. A. and Dennett, D. C. (2011). Consciousness cannot be separated from function. *Trends In Cognitive Sciences*. Vol.15, 358-364.

Davidson, D. (1969). The individuation of events. (Reprinted in *Essays on Actions and Events* (pp.163-180). 1980. New York: Oxford University Press).

Davidson, Donald. 1970. "Events as Particulars." (Reprinted in *Essays on Actions and Events* (pp.181-187.). 1980. New York: Oxford University Press).

De Brigard, F. and Prinz, J. (2010). Attention and consciousness. *Wiley interdisciplinary reviews: Cognitive science*, 1, 51-59.

Duncan, J. (2006). Brain mechanisms of attention. *The Quarterly Journal of Experimental Psychology*, 59, 2-27.

Ereshefsky, M. (1992). Eliminative pluralism. *Philosophy of Science*. 59, 671-690.

Heil, J. (2003). *From an Ontological Point of View*. (New York: Oxford University Press). Reprinted 2009.

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

Kouider, S., Dehaene, S., Jobert, A. and Le Bihan, D. (2007). Cerebral Bases of Subliminal and Supraliminal Priming During Reading. *Cerebral Cortex*. 17, (9), 2019-2029.

---

Lewis, D. (1970) How to define theoretical terms, *The Journal of Philosophy*, 67 (13), 427-446.

Mole, C. (2011a). *Attention is Cognitive Unison: an essay in philosophical psychology*. (New York: Oxford University Press).

Mole, C. (2011b). The metaphysics of attention. (In C. Mole., W. Wu and D. Smithies (Eds.), *Attention: Philosophical and Psychological Essays*. (pp.60-77) New York: Oxford University Press).

Nanay, B. (2010). Attention and perceptual content. *Analysis*. 70, 263-270.

Posner, M. and Rothbart, M. (2007). "Research on attention networks as a model for the integration of psychological science." *Annual Review of Psychology*. 58, 1-23.

Prinz, J. (2007a). The intermediate level theory of consciousness. (In M. Velmans and S. Schneider (Eds.), *The Blackwell Companion to Consciousness*. (247-260). Singapore: Blackwell).

Prinz, J. (2007b). Accessed, accessible and inaccessible: where to draw the phenomenal line. *Behavioral and Brain Sciences*. 30, 521-522.

Prinz, J. (2010). When is perception conscious? (In B. Nanay (Ed.), *Perceiving the World*. (pp.310-332). New York: Oxford University Press).

Prinz, J. (2011). Is attention necessary and sufficient for consciousness? (In C. Mole., W. Wu and D. Smithies (Eds.), *Attention: Philosophical and Psychological Essays*. (pp.174-203.) New York: Oxford University Press).

Prinz, Jesse. (2012). *The Conscious Brain: How Attention Engenders Experience*. (New York: Oxford University Press).

---

Sergent, C., Wyart, V., Babo-Rebelo, M., Cohen, L., Naccache, L., and Tallon-Baudry, C. (2012) Cueing attention after the stimulus is gone can retrospectively trigger conscious perception. *Current Biology*. 23, 150-155.

Smithies, D. (2011). Attention is rational access-consciousness. (In C. Mole., W. Wu and D. Smithies (Eds.), *Attention: Philosophical and Psychological Essays* (pp.247-273). New York: Oxford University Press).

Stazicker, J. (2011). Attention, visual consciousness and indeterminacy. *Mind and Language*. 26, 156-184.

Taylor, J. H. (2013). Is attention necessary and sufficient for phenomenal consciousness? *Journal of Consciousness Studies*. 20 (11-12), 173-194.

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

Watzl, S. (2010). *The Significance of Attention*. PhD Dissertation, Columbia University.

Watzl, S. (2011a). Attention as structuring the stream of consciousness. (In C. Mole., W. Wu and D. Smithies (Eds.), *Attention: Philosophical and Psychological Essays* (pp.145-173). New York: Oxford University Press).

Watzl, S. (2011b). The philosophical significance of attention. *Philosophy Compass*. 6/10, 722-733.

Watzl, S. (2011c). The nature of attention. *Philosophy Compass*. 6/11, 842-853.

Wu, W. (2008). Visual attention, conceptual content, and doing it right." *Mind*. 117, 1003-1032

---

Wu, W. (2011a). "Attention as selection for action." (In C. Mole., W. Wu and D. Smithies (Eds.), *Attention: Philosophical and Psychological Essays* (pp.97-116). New York: Oxford University Press).

Wu, W. (2011b). "Confronting many-many problems: attention and agentic control." *Noûs*. 45, pp.50-76.

Wu, W. (2011c). What is conscious attention? *Philosophy and Phenomenological Research*. 83, 93-120