Attention, Consciousness, and the Semantics of Questions

Abstract *Attention influences the character of conscious perceptual experience in intricate and surprising ways, including our experience of contrast, space, and time. These patterns of influence have been argued to cause trouble for the attractive thesis that differences in the character of conscious experience flow from differences in what we represent (Block 2010). I present a novel theory of the functional role of attention that has the resources for a systematic representationalist account of these phenomena. On the erotetic theory of attention, we bring an interest to the task of perception, captured as a question we seek to answer. Questions, as understood here, are contents that cognitive systems can represent rather than sentences. We process perceptual input as a putative answer to our question in a way that is modulated by attentional focus; attentional focus aims to pick out something that matches what our question is "about." In certain cases, this yields a form of predictive coding: if the contribution of focus matches what our question is about, we take it to select one of the possible answers we are entertaining, even though our perceptual input by itself does not supply a full answer. The proposed account also provides a new account of the phenomenology of salience.*

(Abstract: 206 words, Body: 10,486 words)

It is fairly uncontroversial that changes in attentional focus can yield changes in phenomenal character – what our conscious experience is like – even when stimuli and ocular fixation are held constant (Block 2010; Chalmers 2004; Peacocke 1998). For example, attentional focus can influence our perception of space, time, brightness, and contrast in a variety of striking ways (Stelmach & Herdman 1991; Peterson and Gibson

1991; Tse 2005; Carrasco et al. 2004). Our theories of phenomenal character need to be able to account for the possibility of this influence. I present aspects of a theory of attentional focus that provides avenues for a systematic account of the varied influences of attention on phenomenal character in terms of what is being represented in conscious experience.

The key idea will be that attentional focus involves a richer representational structure than would be involved in just selecting things from the visual data stream or in representing things as more or less salient, or more or less determinate. Attentional focus is driven by a form of *interest* that structures our experience, in a way that can be made formally precise. I propose that attentional focus has to do with processing our perceptual input for a purpose: we deploy attentional focus to try to use our perceptual input to answer a *question* determined by our goals. Questions, as they are understood here, are in the domain of things to be represented, just like propositions or probability distributions, so there is no special role for language in attention on the proposed account and attention does not depend on linguistic capacities. Fundamentally, questions are sets of alternative possible answers (Hamblin 1958), which could be thought of as sets of alternative propositions or states.

The status of the claim that aspects of attention can be understood as manipulating questions and answers is similar to the status of the more familiar claim in vision science that aspects of attention can be understood as updating probability distributions (Itti & Baldi 2005; 2009); these are claims about the functional role of attention in terms of epistemic aims. Both questions and probability distributions fundamentally consist in structured sets of alternatives, though probability distributions involve far more complicated structure than questions. There is no reason to take it to be less plausible that

various areas of the brain could represent and manipulate questions in the sense at issue than that various areas of the brain could represent and manipulate probability distributions.

Attentional focus can influence our perception of space, time, and brightness, among other things, in rather surprising ways. It has been argued that this poses serious problems for representationalism (Block 2010). Certainly, any systematic account of this influence in terms of the representational contribution of attention must have considerable flexibility. Now, if attentional focus is involved in evaluating perceptual input as a potential answer to a question that encodes our interest, we immediately gain some flexibility in how attention could influence what we represent. An example from language provides a ready illustration. If we evaluate an utterance as an answer, rather than just as a bare statement, there is a considerably greater range of propositions we might take the utterance to assert, varying with the question we take to be answered. In fact, we can take utterances to assert propositions, if we evaluate them as answers, in cases in which those utterances would not assert anything on their own.

(1)   John

An utterance of (1) in isolation does not assert anything, even if we stipulate that there is only one John in the world.

(2a)   Who around here smokes?
(2b)   John.

In contrast to (1), (2b) is taken to assert that John smokes, in response to the question in (2a). Questions clearly influence what putative facts we come to represent as a result of processing linguistic stimuli. I will argue that once we think of attentional focus as fundamentally concerned with structuring our perceptual input to match our *interest*, and once we model that interest in the form of questions, we can begin to use this type of pattern to make sense of some of the surprising ways in which attention influences phenomenal character.

The foregoing could be seen as a particularly simple version of the idea that aspects of perception involve *predictive coding*. Predictive coding is the idea that the brain constructs a set of hypotheses about possible causes of perceptual input and then adopts the hypothesis that best fits this input, instead of directly constructing a model of the world from perceptual input. In a Bayesian implementation, predictive coding has already been put to productive use to make sense of various perceptual phenomena (Hohwy et al. 2008). I suggest that attentive perception involves a question-based form of predictive coding: it involves instances of taking perceptual input and using it as the strongest possible answer to the question that encodes the interest that drives one's attention. The core proposal that attention is about the relationship between questions and answers suggests the name "erotetic theory of attention." With the erotetic theory, I will argue that certain recent challenges to representationalism (Block 2010) can be answered.

1   Attention and Phenomenal Character

I have conscious perceptual experience. Right now, as I look around, I see a monitor, a tree behind a window, books, and a collection of office furniture. What is the nature of conscious experience? It seems plausible from where I am sitting that every aspect of the

character of my conscious experience is telling me something about how things are, and perhaps about how they relate to me spatially (Harman 1990). It seems plausible that if something appears horizontally oriented in my conscious experience, then this is tightly related to *representing* something as being horizontally oriented. This representation could then turn out to be correct or incorrect.

How we represent the world and how we make decisions using those representations arguably has a crucial role in determining our success in surviving and reproducing. Now, the fact that representing the world at least approximately well is necessary for human survival, together with the fact of success, does not yet show anything about the nature of conscious experience. Cognitive science suggests that not all representations are conscious. However, it is worth investigating if all differences in phenomenal character – all differences in appearance – correspond to differences in what we represent. If we were to learn that they do, then we would learn something significant about the nature of conscious experience. It is an interesting and substantive hypothesis that all differences in the character of our conscious experience correspond to differences in what we represent. Call that hypothesis "representationalism."[1]

To develop a representationalist account of phenomenal character, we need to say what the representational differences are that correspond to experienced differences in phenomenal character. Differences in attentional focus can yield differences in phenomenal character. An unspectacular example can be obtained from Fig. 1. Maintain ocular fixation on the cross in Fig. 1 and shift your attentional focus between one or more of the black Ts and one or more of the grey Ts. As you shift attentional focus,

---

[1] What I will call "representationalism" in this paper does not include *a priori* restrictions on what we could potentially represent in experience, e.g. objective properties in the world, properties we merely "spread onto the world," fictional objects, sets, relations, etc.

phenomenal character changes. To capture this change in representational terms, we could take a view discussed by Chalmers (2004):
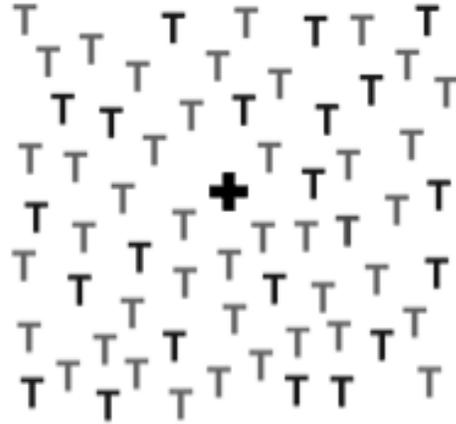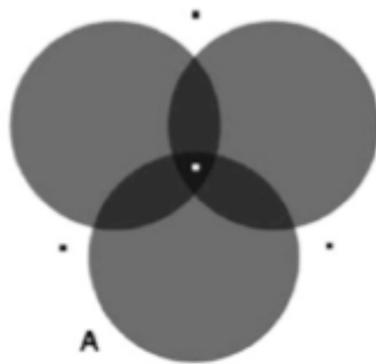

Figure 1

(S)   The contribution of attention to phenomenal character is a representation of salience.

On this view, if we attend to a particular grey T, we represent that T as salient. If we attend to a black T, we represent that T as salient, and so on. A worry one might have about (S) is that salience is not an objective feature of the world, but at best something we project onto the world. However, it might turn out to be adaptive to systematically misrepresent the world as having certain properties, so this should not worry an open-minded representationalist.

Changes in attentional focus can change phenomenal character in ways that go beyond changes in salience. As you shift attention between discs in Fig. 2, while maintaining fixation on one of the dots, the apparent brightness of the discs changes and some discs become apparently transparent (Tse 2005 cited in Block 2010). This means that (S) is not enough to account for differences in phenomenal character engendered by

differences in attention, since we get a change in brightness and transparency and not *just* a change in salience. If we claim that our notion of "salience" entails relative brightness and transparency for the purposes of (S), we have to falsely claim that shifting attentional focus between grey and black Ts in Fig. 1 leads to a corresponding change in apparent brightness and transparency.
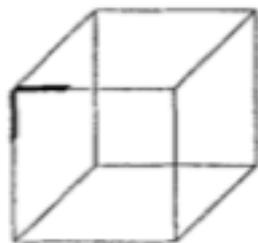

Figure 2

The effect attention can have on phenomenal character extends to the temporal domain. Attentional focus can make an attended stimulus appear as simultaneous with an unattended stimulus that appeared 40ms before the attended stimulus (Stelmach & Herdman 1991; Spence & Parise 2009). The experimental setup is illustrated in Fig. 3. A dot is flashed on the right of the display, while subjects are instructed to maintain attentional focus (schematically illustrated in the diagram as a box labeled "Foc") to the left of the display. The dot disappears again and 40ms later, another dot is flashed on the left of the display where subjects were instructed to maintain their attentional focus.
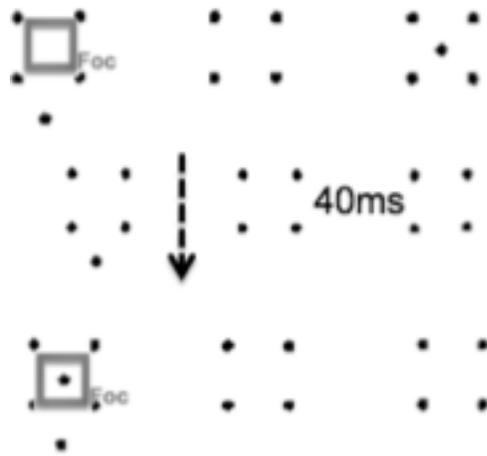
Figure 3

The two dots are subjectively perceived by participants as appearing simultaneously if attentional focus is deployed in the way described, but not otherwise. This is another case in which the influence of attentional focus on phenomenal character outstrips differences in what is represented as salient.

Attention can also influence apparent spatial orientation. There is considerable evidence to suggest that shifts in attention, voluntary and involuntary, can drive changes in perceived depth orientation in figures like Fig. 4. The attended corner tends to be perceived as convex and in the foreground (Peterson & Gibson 1991; Kawabata 1987; Kawabata & Yamagami 1978; Meng & Tong 2004).

Figure 4

As Block points out, perceptual changes like the ones discussed, and many others engendered by shifts in attention, like changes in apparent contrast, speed, size, color saturation are "not just a matter of looking more salient" (Block 2010, p.44). In fact, it looks like no *single* putatively represented property other than salience could account for the variety of the influence of attention on phenomenal character.

So far, we do not have a problem for the thesis that there is some change in representational content corresponding to every change in phenomenal character. We can just say that when appearances seem to change as a result of changes in attentional focus, we represent things as having different properties, where we tailor to each instance what those properties are. However, an unsystematic stipulative account of this sort is unsatisfying and it is worth trying to do better.

If we had an account of the functional role of attention in terms of representational aims that yields enough resources to capture the varied differences in phenomenal character just discussed, we would have an account of changes in phenomenal character in terms of changes in representational content that is explanatorily interesting. I will shortly suggest a new avenue for providing such an account.

Before moving on to my account of attention, we will have to consider a few more cases that this account will need to handle. Block (2010) discusses examples of the influence of attention on phenomenal character that he takes to pose a more general challenge to representationalism. Both voluntary and involuntary attention to the left patch in Fig. 5, while fixating on the black dot, give the two patches the same apparent contrast, though the two patches differ in contrast (Carrasco 2009; Carrasco et al. 2004; Liu et al. 2009).
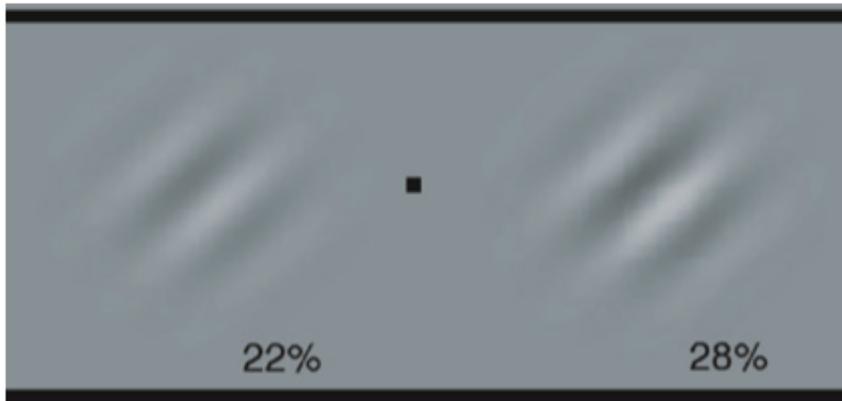
Figure 5

Setting aside the aim of a systematic account for the moment, we might just say that we represent the left patch as having whatever contrast it seems to have after we deploy our attention. If we deploy our attention one way, it will seem more contrasty, and we correspondingly represent it as more contrasty. If we deploy our attention another way, it will seem less contrasty and we will correspondingly represent it as less contrasty. Block argues that a representationalist would now have to say that at least one of those percepts is illusory. After all, the patch cannot be both more and less contrasty, so at least one of the representations fails to correspond to the facts. He also argues that neither of those percepts is in fact illusory.

Block presents two arguments against the idea that one of the percepts of the left patch in Fig. 5. is illusory. The first argument is that there is "no way to pick which distribution of attentional resources engenders veridical perception and which engenders illusion" (Block 2010, p.45). He argues for this claim by elimination of alternatives, though the considerations brought forward, as I understand them, do not seem decisive.[2]

---

[2] Block gives only brief consideration to a number of alternatives. If it is a high absolute degree of attention that guarantees veridicality, then, Block concludes, "if the subject were speaking on a cell phone, *nothing* would be veridically perceived," since talking on a cell phone takes up some attentional resources. That "*nothing*" would be veridically perceived in such a case only plausibly

A representationalist who is happy to postulate a new representational content for every change in phenomenal character engendered by attention will presumably also be happy to stipulate what distribution of attentional resources corresponds to veridical perception for each particular case of interest. Block concludes his argument on a more restrained note, claiming that proposals about what distribution of attention engenders illusion would be "unmotivated." This is not enough to defeat representationalism as a

---

follows if we assume that the same absolute degree of attention is required for the veridical perception of *all* properties. However, it could be that different properties require different degrees of attention for veridical perception. Whether the attention-drain of talking on a cell phone could impact the veridicality of our contrast representations is an empirical question; it certainly impacts the veridicality of some representations (Scholl et al. 2003). There seems to be no easy *reductio ad absurdum* argument against the "degrees of attention" option here. Moving on, if the distribution of attentional resources that engenders veridicality has to do with the *relative* deployment of attentional resources, Block holds that we run into the problem that differences between points in the visual field in terms of relative attentional resources will often be very small. Assigning a cutoff for veridicality to such small differences would be "arbitrary." (Block 2010, p. 46). However, it is not obvious why a representationalist would have to stipulate the necessary magnitude of the difference in advance. It is not obvious why this cannot be non-arbitrarily anchored in some way to what makes a difference to the percept. Another possibility Block considers is to take equal distribution of attentional resources between the two patches to yield veridical perception. He argues that we will find that there are numerous ways of assigning our attentional resources that give equal consideration to both patches, but that will still have different phenomenal character associated with them. For illustration, he suggests the example of the two patches in Fig. 5 with focus at varying points along a vertical line through the fixation point. The phenomenal character of focusing on different points on the line will be different, but the veridicality of the comparison of contrast between the patches will be the same. The idea is that this will regenerate the problem of having multiple distinct phenomenal characters that should not all be equally veridical, if they are to be cashed out in terms of different representational contents. However, Block does not provide an argument to rule out the possibility of accounting for additional differences in phenomenal character in this case in terms of differences in what is represented as more salient in the overall visual scene, while the contrast representations stay the same.

metaphysical thesis, but it does seem that a representationalist account of the impact of attention on phenomenal character might only be *explanatorily interesting* if it had independent motivation – more motivation than mere stipulation for individual cases.[3] In this regard, it seems to me that Block succeeds in presenting a challenge for representationalism that is not easily met.

However, if we had a systematic account of the representational function of attention that has the resources to generate, as a function of attentional focus, differences in representational content corresponding to experienced differences in phenomenal character, then there would be no *further* problem about saying what distribution of attention provides veridicality. In this case, there would be nothing unmotivated in saying that whatever distribution of attention yields a veridical representation in a particular case is the distribution of attention that yields veridical representation in that case. I will suggest below how an appropriate representationalist account of the representational role of attention might be constructed.

There is no reason to think that a representationalist with a systematic account of the influence of attention on representational content would have to produce veridicality conditions for attention stated independently of the veridicality conditions of the representations we obtain as the result of the influence of attention. Moreoever, I will argue in a later section that it is unclear that any *non*-representationalist account of the

---

[3] One might observe that if our project is to find out something interesting and substantive to say about the nature of conscious experience, finding that extant representationalist approaches to attention are "unmotivated" is only helpful for Block's non-representational mental paint proposal if that proposal turns out to be better motivated. As far as I can see, he provides no clear reason to think that it is. Things may be different if we regard mental paint as our null-hypothesis.

influence of attention on appearance could provide a motivated account of some of the patterns Block discusses.

Toward the end of his paper, Block provides additional phenomenological considerations against illusion that could be seen as posing a further challenge.

"Take a look at [Fig. 5], fixating at the fixation point and moving your attention around. It does not look as if anything is *really* changing in contrast. That is, these changes don't have what Burge (2009) is calls the phenomenology of objectivity. The change looks unreal (to me). … The upshot is that it is a mistake to treat the change in phenomenology wrought by the change in attention as equivalent in its effect on phenomenology of a change in contrast in the world" (Block 2010, p. 54-5).

The apparent change in contrast is not illusory, on this view, because we do not take it to even putatively apply to the world. There is no misrepresentation of the world because the relevant aspects of our perceptual experience are not aiming to represent the real world.

I believe that there is something importantly correct about the contrast change in Fig. 5 seeming "unreal." Yet, it seems to me that the feeling of unreality depends on context. I suggest that if one were to look at a display like Fig. 5 for the first time, presented using sufficiently complicated experimental apparatus to leave it unclear whether the display is physically changing, one would not perceive the change as unreal on a first look.

I propose that whether we experience perceptual changes influenced by attention as real or unreal in part depends on factors beyond the proximal stimulus (i.e. the image on the retina). Consider another example. The two depth percepts we can obtain from

Fig. 4 seem about as unreal to me as the contrast changes Block describes. Yet, if we monocularly view a hand-held cube made of wires in appropriate lighting conditions, we obtain a situation in which our depth perception is strongly driven by the same depth cues that are triggered by the flat Fig. 4. The proximal stimulus we process is not so different from what we get from Fig. 4. In such a setup, which parts of the cube we focus our attention on still influences the perceived depth orientation of the cube, one of which will now be veridical and one illusory. It seems wrong to say that phenomenal character associated with the correct orientation seems unreal in this case. Similarly, the perceptual possibility with the "wrong" apparent orientation does not seem unreal either. The feeling of illusion, *of perceptually misrepresenting the world*, seems quite strong in the latter case, at least initially. The point here is that it is possible to experience the perceptual influence of attention as real or unreal depending on the circumstances, in a way that is not fully determined by our momentary perceptual input.

If we experience changes produced by shifts in attention as unreal, then we are not experiencing them as illusions and we cannot capture them in terms of misrepresentations of the real world. However, we do not have to take it that cashing out differences in phenomenal character in terms of representational differences limits us to (mis-)representations of the real world; we can represent something unreal, as we might in fiction. The view of the functional role of attention I will propose provides resources to accommodate this.

2   The erotetic theory of attention

To find an account of the functional role of attention, begin by asking about the *objects* of attention. Consider a classic visual search paradigm as used in the study of attention

(Treisman and Gelade 1980). The task is to report presence or absence of a target in a display of distractors. The question is how attention relates to this task. It is useful to consider what stays the same and what changes as the task is performed. These considerations lead to a prima facie distinction between attention as a form of cognitive control or monitoring that is applied to tasks, and attention as a form of orientation or focus, or "perception scientist's attention."

In a cognitive-control sense of attention, the object of attention stays the same throughout normal performance in the search task. If the participant is paying proper attention to the task, she will continue to remain sensitive to whether she can identify the target in the display and realize that the task is completed if she has identified the target and reported it. If the participant ends up identifying the target but then engages in mere contemplation of the target, we would take it that she has decided to abandon her task or that she is not following the task attentively.

In another sense of "attention," the proper deployment of attention in the search task is to move one's attentional *focus* between different elements in the display until the target is identified. It is an interesting result in perceptual psychology that it is not always possible to immediately identify a target without first using attentional focus to scan the display, even if the projection of the display fits on our retinas. Thus, in our example, keeping the object of attention constant in one sense (attention to the task) involves regularly changing the object of attention in another sense (attentional *focus* to objects, features, and positions).

How can we give an account of attention that brings together both the "control" and "focus" aspects of attention? I propose that in the case of perception, attention is concerned with evaluating perceptual input with respect to an *interest* we bring to it. I

propose that this interest can be understood in representational terms as a question. In the visual search example, our question, setting aside the bit about reporting detection, is something like "where is the target?" The object of our attention from a control perspective stays the same in the sense that our interest/question stays the same. Attentional *focus* (which can pick out objects, features, and locations) is then understood as part of what it takes to evaluate part of our perceptual input with respect to our interest. More specifically, I propose that attentional focus is to be understood as part of what it takes to evaluate something as an answer to our question that encodes our interest.

I take attentional focus to be an aspect of cognition the mechanisms of which are studied by perceptual psychologists under the rubrics of "attention," "focal attention," or "attentional focus" (for a recent ambitious review of the empirical literature, see Wright and Ward 2008). What I am proposing is a new account of what those mechanisms are aiming at, or of their *functional role*: attentional focus serves the evaluation of parts of the perceptual input as potential answers to the questions encoding our interest. This account will be fleshed out in the sections to follow.

The hypothesis that attention is about the relationship between questions and answers suggests the name "erotetic theory of attention" (cf. erotetic *logic* as the logic of questions). Questions here are understood as representational contents rather than linguistic expressions, so non-linguistic creatures are not barred from having attention. To flesh out the erotetic theory we will first need to develop a suitably precise notion of questions and answers.

2.1   The structure of questions

"Question," as I use the term, denotes something in the domain of representational contents. One way to represent questions is through language. The semantic contribution

of an interrogative utterance is a question, just as the semantic contribution of an assertive utterance is a proposition. Thus, "Wer raucht?" "¿quien fuma?" and "who smokes" semantically contribute the same question, though they are different interrogative sentences.

An insight due to Hamblin is that while propositions determine truth conditions, questions determine what counts as an answer. So a question determines a set of alternative answers. Letting '$[\![X]\!]$' stand for the semantic contribution of X and '$\rightarrow$' stand for 'determines':

(3) $[\![$Who won?$]\!]$ → {Jack won, Pete won, Sarah won, …, Jack and Pete won, …}

Since questions correspond to (on some views are *identical* to) sets of alternative answers, answering a question can be seen as *eliminating* the question, in the sense of eliminating all alternatives save one, leaving us with a proposition.

In order to fully count as an answer to a question, it is not enough to supply material to pick out one of the alternative answers. In order to be what has been called a *congruent* answer, *focus marking* has to correspond to the question in the right way. In English, focus marking is normally indicated with contrastive stress; other languages have special particles or other devices (Beaver and Clark 2008).

(4)   Q: Who smokes?

(5)   A: [$_{Foc}$*John*] smokes.

(6)   A: *John [$_{Foc}$*smokes*].

Answer (5) with focus marking on 'John' is a congruent answer to (4), while (6) is not.

To capture this, Krifka has proposed to analyze a question as a pair of a function and a set

of elements (the "domain" of the question) to which the function can be applied (Krifka

2001). Applying the function of the question to an element of its domain yields one of the

alternative answers of the question, on this analysis. The general pattern (8) is illustrated

in example (7) below.

(7)   ⟦Who smokes?⟧ = <λ*x*.Smoke(*x*), {*y*: Person(*y*)}>

(8)   ⟦S_Q⟧ = <*Function$_Q$, Domain$_Q$*>

Krifka proposed that we can capture the contrast between (5) and (6) by the following

requirement: *it is a necessary condition for a congruent answer to have focus marking*

that corresponds to the domain of the question. In (5), focus marking is on John, which

corresponds to something in the domain of the question (viz. the set of persons, including

groups), while there is no such correspondence in (6).

For a statement to be a congruent answer, focus needs to pick out something in

the domain of the question, and the statement actually needs to supply one of the

alternatives of the question. This in effect means that the statement has to assert what we

would get by taking the focused element and applying the function of the question to it.

We can put this more formally as follows.

(*Congruent Answerhood*) Relative to a question <*F, D*> with a function *F* and a domain

D, a statement *S* is a congruent answer if and only if what is focused in the statement

picks out an element of *D,* and *S* encodes the proposition one obtains by applying *F* to what is focused in *S*.

For example, '[$_{Foc}$John] smokes' is a congruent answer to the question $<\lambda x.\text{Smoke}(x)$, $\{y:$ Person$(y)\}>$, since John is an element of $\{y: \text{Person}(y)\}$ and '[$_{Foc}$John] smokes' encodes the proposition one obtains by applying $\lambda x.\text{Smoke}(x)$ to John.

In sum, questions, for my purposes, are representational contents that consist in a pair of a function and a set. In the case of language, evaluating something as an answer requires a notion of focus marking, where focus marking needs to pick out something that corresponds to the domain to the question. I will suggest that the functional role of attentional focus mirrors that of focus marking in language.

To develop the erotetic theory of attention, I will rely on the analysis of questions just described. As was made clear, questions are not linguistic items, so there is no sense in which language will be required as a precursor to attention. However, I will freely assume that all sorts of cognitive states involved in attention can bear pseudo-propositional content (Burge 2009). As noted in the introduction, there is no clear reason take it to be less plausible that various areas of the brain could represent questions than that various areas of the brain could represent probability distributions, which have found various uses to model aspects of perception and attention (Itti and Baldi 2005; 2009; Hohwy et al. 2008).

Moving on, I will use the foregoing analysis of questions to make sense of "interests" we bring to attentive perception. I will then say more about the nature of focus and how it is assigned.

2.2 Interests and structured questions

On the erotetic theory, attentive perception involves an interest we bring to perception and this interest is encoded in the form of a question. I propose that the questions that encode our interests are structured questions in the sense just developed, with a function and a domain. The domain encodes, in a sense, what our interest is "about."

Return to the example of searching for a target in the display of distractors. In this sort of case, the interest we bring to perception can be captured in terms of the question, "where in the display is the target?" Whenever attentional focus is assigned to a particular position and we find that the target is not at that position we have thereby ruled out one of the alternative answers to the question and we have made partial progress towards fulfilling our task. If we encounter the target, we have answered the question and we have completed the task (as before, we are setting aside reporting requirements). We can represent the relevant question using the formalism developed above.

(9) $<\lambda x.\text{Target\_is\_at\_}x, \{y: y\_\text{is\_a\_position in the display}\}>$

Ignore the issue of focus in determining answerhood conditions for the moment. Then, for any given position in the display p, (9) admits the proposition that the target is at p as an answer. For the purposes of this paper, I will leave open what sorts of questions could correspond to interests we bring to attentive perception, since I will be primarily interested in exploring what the model can do on the assumption that we bring certain questions to the table.[4]

---

[4] Elsewhere, I argue that we can make sense of all tasks we may attentively pursue as determining questions in terms of which we can monitor performance on those tasks (REFERENCE

2.3 Focus and answers

I proposed that we can model the interest we bring to perception when we attentively search for a target as a question, roughly, "where in the display is the target?" I noted that while the task we attend to stays constant (we keep looking for the same target), our attentional focus shifts. With a structured analysis of questions as an account of the interest we bring to perception, we can now give an account of the functional role of attentional focus. In the search task, we might have everything available in the visible display that would allow us to answer the question "where in the display is the target?" but we cannot process this question as answered until our focus has picked out something in the domain of the question, a position on the display (see (9)), and it turns out that the target is at that position.

On the erotetic theory, attentive perception involves bringing an interest to perception, encoded in the form of a question. In order to satisfy our interest, we have to answer the corresponding question. Since our questions are structured questions, in order for us to be able to cognize something as an answer to our question, focus – *attentional* focus – has to pick out something that corresponds to the domain of the question. The functional role of attentional focus in representational terms mirrors that of linguistic focus in language. This proposal is supposed to be an account of the functional role of what is called "attentional focus" or "selective attention" in vision science. Focus "selects" certain positions, objects, or features in the cognitive economy. For simplicity, I will refer to all possible objects of focus as "features". On the erotetic theory, focus marks features as potentially corresponding to the domain of a question.

SUPPRESSED *forthcoming*). However, the main points I try to make in the present paper do not depend on this broader picture.

How focus is assigned is a question for empirical perceptual psychology. Though we can consciously influence the trajectory of attentional focus, search for a target is in fact more efficient if we do not try to control where our focus goes (Wolfe 2000). To a large extent, how focus is allocated is driven by automatic processing. Central aspects of this processing seem to be driven by fairly low-level processing of stimuli and the history of recent perceptual input, since computer models of the trajectory of attentional focus that only rely on such factors yield surprisingly good predictions (Itti and Baldi 2005; 2009). This is not to deny that particular tasks and expectations can clearly bias attentional focus. What aspects of the visual scene are processed in a way that influences attentional focus continues to be under investigation (Jiang et al. 2006).

What I hope to have accomplished in section 2 is to have made it reasonably clear how we might understand attention in perception as an interplay between questions and answers, where we bring questions to perception as a form of interest and where our ability to answer those questions depends on the assignment of attentional focus. What I now want to consider how we can use this model to make sense of some of the surprising ways in which attention impacts phenomenal character. Before we can return to perception, we first have to consider additional aspects of how questions relate to answers.

3   Question answering as predictive coding

Predictive coding is the idea that rather than entirely relying on getting a model of the world through inferences from perceptual input, the brain constructs a set of hypotheses about the causes of perceptual input and then adopts the hypothesis that best fits the perceptual input (Hohwy et al. 2008). The basic pattern required for predictive coding,

the selection of an element from a set of alternatives, does not require something as highly structured as a probability distribution. The relationship between questions and answers itself involves patterns of inference that can be seen as a form of predictive coding.

Questions can contribute part of the intuitive content of answers that would otherwise be too incomplete to count as answers on their own. Consider so-called term answers as in (11):

(10)   Q: Who smokes?

(11)   A: [*John*]_{Foc}

~> John smokes.

The intuitively conveyed content of the answer in (11) as a response to (10) is that John smokes. We can think of this pattern as involving a form of predictive coding in the following sense: In raising the question of who smokes, we are representing a set of alternative situations, including the alternative situation that John smokes. When we hear "John" and interpret what we hear as an answer, we select the alternative situation that best matches what we hear, viz. the alternative situation that John smokes.

The pattern just described only holds in language if the term answer could be taken as elliptical for a full answer. In languages like German with overt case marking, it can easily be illustrated that the pattern does not hold when grammatical features of the would-be answer prevent ellipsis, as in (14).

(12)   Q: Wen hat Johannes gesehen?   [*Who did Johannes see?*]

(13)   A: [Foc Den Karl]   [*Karl*$_{Accusative}$]

~> Johannes hat den Karl gesehen.   [*Johannes saw Karl*]

(14)   A: *[Foc Der Karl]   [*Karl*$_{Nominative}$]

This constraint also rules out taking (16) as a congruent answer to (15), since the presence of a verb phrase irrelevant to the question prevents us from taking the answer as elliptical in a relevant way. However, the problem is not that there is irrelevant information provided by (16). As the acceptability of (17) shows, it is fine to include extra information that is strictly speaking irrelevant to the question, provided it is included in a way that does not prevent adding to the answer trough ellipsis.

(15)   Q: What did you buy?

(16)   A: *[Foc *This*] is cheap and sounds great.

(17)   A: [Foc *This*], which is cheap and sounds great.

The above observations suggest a general pattern of ampliative interpretation of perceptual input relative to a question, subject to certain constraints. This pattern can be stated in a way that does not make essential reference to language and thus becomes potentially applicable for question answering in other cognitive domains. The idea is that we can get a form of predictive coding that amounts to treating perceptual input as the strongest possible answer that does not violate any constraints.

(18) Relative to a question $<F, D>$, if an answer $F(X)$ could be obtained by making additions to $[_{Foc} X]$ as part of what is evaluated as a potential answer, without violating constraints at the level of representation supplying X, then the following pattern holds:

Q: $<F, D>$

A: …, $[_{Foc} X] \in D$, …

$\sim> $ A': …, $F(X)$, …

In the case of language, the substance of (18) is that we can treat a focused term as a full answer, provided the full answer could be arrived at by adding more syntactic constituents to the term without violating syntactic constraints.

In the case of vision, the substance of (18) would be that we can treat something that is picked out by focus as corresponding to the domain of our question as a full answer, if the full answer could be arrived at by activating additional feature representations without violating constraints imposed by other feature representations.

4   Phenomenal character, questions, and predictive coding

I will now explore if we can give a systematic account of the effects attention can have on phenomenal character with the erotetic theory of attention. I will begin with the intuition that shifts in attention can engender shifts in apparent salience.

4.1   Salience and interest

Recall the simple way in which attention can influence phenomenal character discussed in section 1 with the example in Fig. 1. As we shift our attentional focus from the gray Ts to the black Ts, our phenomenal character changes. There is a sense in which what seems most salient to us changes in this case. As noted, changes in salience do not exhaust the

variety of ways in which attention impacts phenomenal character, but it does seem plausible that changes in what appears salient are *among* the ways in which attention impacts phenomenal character. The question then is what salience comes to.

One option would be to take salience to be a property that we can represent things as having or lacking. It would be attractive to have an account of salience that fulfills the following desiderata:

1. We are not representing properties of external objects that do not objectively exist.

2. We are not representing our own mental states in representing something as salient. Experience remains 'transparent' (Harman 1990).

3. Since salience is a representational content, it has non-trivial accuracy conditions (see Siegel 2010).

Now, if we take salience to be a property of external objects, like shape, we will have to conclude that we always misrepresent when we take it that some things have this property while others do not, since there is no corresponding asymmetry in the distal layout. If we take it that salience is a property of our own mental representations reflected in phenomenal character, we can no longer claim that our perceptual phenomenology is in all its aspects transparent onto the non-mental world.

The erotetic theory of attention allows us to identify a representational content as corresponding to salience that satisfies our three desiderata. We can say that the phenomenology of salience supervenes on an independently motivated representational contribution of attentional focus. Attentional focus represents things as corresponding to the domains of our questions. Those questions themselves fundamentally correspond to

sets of alternative situations; questions are contents not mental states. If what is picked out by focus does not match any of our questions, focus momentarily misrepresents. Thus, the erotetic theory of attention can give us accuracy conditions for salience without having to insist that we represent perceived distal properties as changing with changes in what we perceive as salient or having to insist that saliency represents mental states. Moreover, the relevant representational content in fact plays a crucial role in the 'gear box' of attention and is not just stipulated to defend representationalism.

I will now move on to aspects of the influence of attention to phenomenal character that may involve the kind of predictive coding effects described in (18). First, we need concrete questions. I believe we have to distinguish cases in which our phenomenal character is relevantly influenced by unusual or special *adopted interests* that often are consciously accessible, and cases in which our phenomenal character is relevantly influenced by what we may call *default interests* that may escape conscious access either in principle or in practice. Ultimately, there may be multiple interests at play simultaneously.

4.2   Adopted interests

In a chapter on attention that remains insightful to this day, James notes that "every stir in the wood is for the hunter his game; for the fugitive his pursuers," and "every bonnet in the street is momentarily taken by the lover to enshroud the head of his idol" (James 1890, Ch. XI). James seems to be describing the phenomenology of momentarily taking something that captures our attentional focus to be whatever it is that we are intently searching. On the proposed view of attention, together with the notion of question-based predictive coding, we can make sense of this influence of attention and interest on phenomenal character.

Suppose we are frantically trying to spot a tiger in the wood. When our attentional focus is captured by surprising rustling, it marks a certain location, something in the domain of the question we use to monitor our search task ("At *what location* is there a tiger?"). Attentional focus can be taken to suggest, "location such-and-such!" in this sort of case. We then treat this suggestion as the strongest possible answer to our question, as long as the sensory input already processed does not rule out that what is attentionally focused corresponds to a tiger. On a rapid first glance, visual processing may not yield enough detail to rule out as a tiger whatever is at the attended location, especially if the texture at the location is complex enough to allow camouflage (recall the constraint in (18)). As we process our sensory information more, it is likely ruled out that there is a tiger at the relevant location, which blocks the pattern in (18). As a result, the phenomenology of taking a tiger to be at the location of the rustling fades.

The foregoing captures a pattern that seems to me to be common enough so that most people will have experienced some version of it; if not with a tiger in the wood, then perhaps while looking for a friend at a busy station in the knowledge that the train is about to depart. Anecdotal evidence about interest impacting phenomenal character of the kind just presented does not show that there is a narrowly *perceptual* effect. However, we do not need those cases to involve perceptual effects in order for them to count as instances in which attention influences *phenomenal character* in a way that needs to be explained.

I suggest that the reality/unreality cases discussed at the end of section 2 fall into broadly the same category as the "tiger in the wood" cases, involving adopted interests. Recall from section 2 that, at least in the right context, the apparent change in contrast in Fig. 5 as the result of a shift in attentional focus does not have the "phenomenology of

objectivity." I argued that whether we experience apparent changes due to shifts in attention as apparent changes in the world depends on factors beyond the stimulus. I suggest that what explains these differences is that we can interpret the deliverances of our perceptual systems as answers to questions about the actual world, or as answers to questions about non-actual worlds or fictions. I suggest that when we experience the apparent contrast change in Fig. 5 as unreal, our interest has shifted from the ordinary interest in what the space in front of us is like to something akin to an interest in a fiction.

How is experiencing perceptual input as unreal fiction different from experiencing it as an illusion? Consider the experience of watching a Western on TV. We could look at the TV as presenting us with a rather serious illusion. Try to imagine what it would be like to experience an illusion of a shrunken prairie complete with miniature cows and horses stuffed into a bottomless box with a window to the living room. It is not so easy to imagine what this experience would be like, since we do not normally experience our TVs as illusory in anything like this way. I suggest that the reason we do not experience this sort of illusion is that we directly interpret the relevant bits of visual input as telling us about a fiction.[5] We skip any intermediate step of experiencing an illusion, just as we skip interpreting a novel as a pack of lies before interpreting it as telling us about a fiction. It seems natural to capture the difference between interpreting something as fiction or as a statement about the actual world as a difference between questions we attempt to answer. We may try to use what we hear or see as answering a question about the actual world or as answering a question about a fiction (or one fiction rather than another).

---

[5] If we watch live programming, we might be interpreting the input as telling us about the actual world from a different center or perspective. If you think that the obvious flatness of non-3D TV images is what blocks the illusion, imagine I am addressing myself to a one-eyed audience.

Why might it not be intuitively obvious that our interest applied to perception can shift? Some experiences of uncertainty between reality and unreality may be necessary to become more reflectively aware of the possibility of a shift. Yet, even in our display-rich modern environments, there only rarely arises uncertainty about whether we should interpret a given bit of perceptual input as answering questions about the actual space in front of us or as answering questions about some fiction. With the exception of still rare cases of augmented reality applications, we ordinarily know that we are not to interpret what happens on a screen as telling us about what is happening in front of us *right now from our own perspective*.

I suggest that contemporary art may help with increasing reflexive awareness of shifts in the type of interest we apply to our perceptual input. A striking aspect of recent works by Trisha Baga, notably *Flatlands 3D* (2010) and *Madonna y El Niño* (2009) is that her coordination of live performance and 3D video can in fact produce experiences of uncertainty between reality and unreality, beyond the issue of illusion. These works produce in the viewer successive moments of uncertainty with respect to how to interpret different parts of visual and auditory input - whether to interpret them as answering questions about what is going on in physical gallery space or about what is going on in different layers of depicted fiction.[6] For example, there are moments when it is unclear whether a moving speck of light should be interpreted as telling us something about an actual disco-ball on the gallery floor or about a fictional reflection of light depicted in a video. Similarly, there are moments when it is unclear whether the appearance of a water droplet corresponds to something depicted in a video, or to a real water droplet on the

---

[6] Live performance seen at *Greene Naftali Gallery*, New York, 2010. The video component can be seen at http://www.bagalab.biz/work/VIdeos.html, accessed on 10/20/12.

filter of the projector or on the physical screen (the screenshot in Fig. 6, if viewed with red-green 3D glasses cannot entirely reproduce this effect, but provides a rough idea of the setup). As the performance progresses, which interpretation is dominant shifts, making a striking difference to the character of the experience.

Figure 6: Still from Baga (2010). *Flatlands 3D*.

On the erotetic theory, we can model the foregoing as follows, with a slight modification in how we regiment questions. We say that the questions that correspond to the interest we bring to perception encode not only a function and a domain, but the intended target, I, of the effort at representation.

(19) $<F, D, I>$

On this view, the question encodes whether it is a question about the actual world in front of us, or about some fiction. The pseudo-propositions contributed by perceptual processing that we evaluate as potential answers may simply not mark such distinctions. Whether we take our experience to be "unreal" depends on whether the question we try to answer with our perceptual input aims at something unreal or something real. An advantage of this way of capturing the difference between "unreal" and "reality directed" phenomenology is that we are not forced to make the narrow contribution of visual processing itself ambiguous or indexical with respect to different representational aims. Since the phenomena discussed in the sections to follow do not hinge on the real/unreal distinction, I will suppress the target of the effort at representation in my regimentation of questions for expository simplicity.

4.3   Default interests and their effects on perception

If the picture developed so far is to account for the variety of perceptual effects discussed in section 2, we cannot only rely on questions generated by consciously adopted interests, or on questions that depend on some special context. The perceptual effects surveyed in section 2 are independent of particular task instructions. To what questions could we appeal in these cases? I suggest that visual perception involves a default interest that may not be fully conscious, and that involves a question roughly paraphrased as "what is most novel in the foreground?"

4.3.1   An argument for default questions

My first task will be to argue that something in the area of "what is most novel in the foreground?" is a plausible candidate for a default question for visual perception. My argument will hinge on the following claims:

(20) By default, the assignment of attentional focus is anarchic (Wolfe 2000), not directed by conscious control

(21) By default, the assignment of attentional focus is largely based on stimulus-dependent aspects of the visual input (Itti and Baldi 2005; 2009).

(*Default congruency hypothesis*) The attention system is optimized to yield congruent answers in the default case.

Claims (20) and (21) find support in empirical literature already discussed. The crucial claim for my purposes is the default congruency hypothesis. If (20) and (21) hold fixed what sorts of things attentional focus is drawn to in default cases, the default congruency hypothesis basically requires that the default case includes a default question that matches what focus is drawn to by default.

If we take the foregoing seriously, we can develop a decent hypothesis about what might be a default question by figuring out what aspects of a natural scene autonomous focus is drawn to by default, and then exploring questions that would make it likely for that kind of assignment to yield a congruent answer.

Bottom-up models of attentional focus are getting increasingly good at predicting the movement of our focus as evidenced by eye-tracking data for people viewing natural scene videos (Itti and Baldi 2005; 2009). Models suggest that attentional focus is, by

default, assigned to those parts of the scene that are the most novel or surprising in a sense that can be made precise. Other work suggests that attention also appears to be attracted to those parts of a scene that are likely to correspond to the spatial foreground (Mazza et al. 2005). If the default congruency hypothesis is correct, there should be a default question for vision that corresponds to that default focus assignment, viz. "What is most novel/surprising and in the foreground?":

(22) $\langle \lambda x.(\text{Foreground}(x) \ \& \ \text{MostSurprising}(x)), D \rangle$

This will likely have to be revised. However, it serves as a useful working hypothesis to explore how we might make sense of the perceptual effects of attention from section 2.

4.3.2   Foreground and correlated properties

Reconsider figures like the Necker cube in Fig. 3. The explanandum was the default tendency to perceive the focused corner as in the foreground. Suppose we have attentional focus on one of the corners, as indicated in Fig. 7, with corner A focus marked. We could get attentional focus assigned this way by conscious effort or by drawing the corner in such a way that anarchic attentional focus assignment will be more likely to be drawn to it.
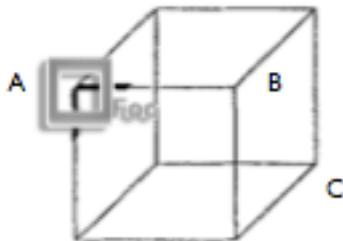

Figure 7

On the theory developed so far, we have:

Default question   Q: $<\lambda x.(\text{Foreground}(x)$ & $\text{MostSurprising}(x)), D>$

Attentional focus assignment   A: …,[$_{\text{Foc}}$ Corner_A], …

Pattern in (18)   A': …, Foreground(Corner_A) & MostSurprising(Corner_A), …

Thus, the prediction is that by default, the focused corner, in this case corner A, will be perceived as in the foreground. The other features involved in this display seem to impose further constraints. If corner A is represented as convex, and we can identify a cubic object, it appears that corner B must be represented as convex as well, perhaps because the activation of a cube representation requires co-activation of the right set of features diagnostic of cubes, like co-oriented corners. On the present proposal, this constraint is not engendered by attention itself but by the way the relevant visual feature representations relate to each other.

The prediction is also that corner A is represented as most surprising or novel. Unlike the case of depth orientation, there is no obvious constraint preventing us from taking the whole apparent object as equally novel, so we do not get an effect of attention that would introduce perceptual differences depending on where we focus.

The account of the effect of attention on phenomenal character in the case of the Tse figure (Fig. 2) follows the same pattern. By default, whatever disk we attend to will be perceived as in the foreground. Transparency and apparent brightness effects are explained as the result of connections between representations of depth and representations of brightness in the visual system. There is independent evidence that representations of depth organization influence representations of more low-level perceptual features like brightness (Adelson 1993). This account is not far from Tse's

own explanation for the brightness and transparency effects. Tse presented a stereoscopic version of the stimulus that partly removes ambiguity in depth organization. For that stimulus, the backmost disk appears as dark and opaque, while the brightness and transparency the other disks depending on whether they appear to be in the foreground (Tse 2005).

There is reason to think that a subtle influence on depth organization is also at the heart of the case central to Block's (2010) discussion, in which attentional focus can make a Gabor patch appear to have increased contrast (Carrasco 2009; Carrasco et al. 2004; Liu et al. 2009). Notice that Gabor patch configurations like the ones being compared in Fig. 5 intrinsically involve depth cues. Within a single Gabor patch, the luminance gradients suggest ridges in depth, and those gradients just are what constitutes the patch. You can obtain something that looks similar to a Gabor patch by placing a drop of water on a thin sheet of paper, causing the paper to warp. Between the two Gabor patches in Fig. 5, we additionally have differences in contrast providing a relative depth cue. More contrast is readily treated by the visual system as a cue for increased spatial proximity, which is what makes it possible to use contrast gradients in painting to provide an impression of depth. It seems hard to avoid the conclusion all percepts we obtain from flat Gabor patch stimuli under normal conditions are fundamentally illusory, where apparent contrast is confounded with subtle apparent depth. Our visual system is not wired to allow us to perceive the physical ink splotches on the page "as they really are" in this case, since this would clash with being optimized for representing objects in depth given flat retinal images as input. However, we still need to account for the possibility of attention influencing apparent contrast and we can ask when attention introduces differences or similarities that are not in the stimuli. Even if there is no absolute

veridicality in Gabor patch percepts, we can ask about the veridicality of same/different judgments.

If attention, by default, can make us represent things as in the foreground, it is not entirely surprising that it can increase perceived contrast. Relative foregroundedness is correlated with increased contrast, in our way of experiencing the world, so relevant detectors in the visual system have a tendency to be activated simultaneously by natural scene stimuli. There is a tendency for neurons that fire together to develop lateral connections (following Hebb's Law). As a result, when attention triggers a foreground representation, then that representation could softly activate representations of features that correlate with foregroundedness, like increased relative contrast, through a standard neural mechanism. This could account for the possibility of seeing an unattended patch as equal to an attended patch that would otherwise be taken to have less contrast. In a case in which we only have a single Gabor patch, there remains room for attention to make the apparent depth of the ridges more pronounced, increasing apparent contrast by separating parts of the Gabor patch more clearly into different apparent depth planes. This could account for the effect of perceiving an attended single Gabor patch as having more contrast than an unattended single Gabor patch (Carrasco et al. 2000). A core illusion involved in percepts obtained from Gabor patch stimuli is independent of attention. However, we can avoid further illusions of spurious similarity or difference in contrast by assigning equal focus to our candidates for comparison.

The idea just proposed is that attention can activate a foreground representation via question-based predictive coding, where this activation can in some cases laterally spread to representations whose normal activation is correlated with foregroundedness. This idea brings some order to what may otherwise seem like a mere natural history of

curious perceptual effects. It may of course turn out some or all parts of the proposed explanations turn out to be mistaken, but at least the erotetic theory provides an integrated account where so far there is no clear systematic alternative. As argued above, representationalism would clearly benefit from a systematic theory.

Block (2010) mentions but fails to explain a contrast between cases that lends further support to the picture provided within the framework of the erotetic theory. Apparent color saturation can be increased by attention, while apparent hue cannot (Fuller & Carrasco 2006). This fits the pattern proposed in this section. The colors of objects close to us tend to be seen as more saturated than those of objects far away, which allows us to use saturation gradients as depth cues in drawings. However, hues are not experienced as changing with distance. Thus, the proposed account of how attention can impact phenomenal character can account for the contrast between the influence of attention on apparent saturation and hue in a principled manner, flowing from an account of the representational aim of attention.

To have provided an account of the asymmetry between apparent hue and apparent saturation with respect to the influence of attention is not trivial. An account of the influence of attention on perception might include a full description of the neural mechanisms underlying the differential effect of attention on hue and saturation, without even beginning to answer the question of why these mechanisms provide for an effect in the case of one feature dimension (saturation) but not the other (hue). It is unclear how any non-representationalist account of the influence of attention on phenomenal character would begin to explain the asymmetry.

In contrast to hue, which does not correlate with object distance and is not affected by attention, other apparent features that do correlate with object distance are

affected by attention: For example, nearby objects subtend a greater degree of visual angle and attended objects can appear larger (Gobell & Carrasco 2005).[7] In addition, nearby moving objects produce more rapid changes on the retina than far-away objects, and attention can increase perceived flicker rate (Montagna & Carrasco 2006). Finally, we perceive nearby object textures in more detail, and attention can increase apparent spatial frequency (Gobell & Carrasco 2005). The present account predicts that for instruction-less experimental setups, apparent features that normally vary with distance are more heavily represented among the cases that involve changes in appearances due to changes in attention than apparent features that do not normally vary with distance, like hue. Of course, the proposed account does not imply that *only* appearances connected with distance can be affected by attention, as evidenced by the cases in the previous and following subsections.

4.3.3   Time and sub-scene representations

I will now turn to the finding that attention can make a flashed dot appear as simultaneous with an unattended dot flashed 40ms earlier (as described in section 2 with the help of Fig. 3). I believe this effect can be made sense of within the present

---

[7] To avoid confusion, note that the account I propose is *not* that we perceive two equally sized projections on the retina and then infer that the object corresponding to one of those projections must have a different size from the other because attention made us represent it as more proximal. If the observed effects were due to an inference of this sort, they should be reversed; a nearby object has to be *smaller* than a more distal object, if it subtends the same degree of visual angle. The mechanism I propose is simpler and more direct and can be implemented by connections between low- or intermediate-level feature representations: A representation of proximity, activated through attention, softly activates representations of features that normally correlate with proximity (e.g., looming larger), through lateral excitatory connections that could be the result of Hebbian learning.

framework. However, this requires a more sophisticated view of how visual scenes are represented, adding a diachronic dimension.

Research in computational models of human attention has suggested that we may spontaneously generate so-called "minimal sub-scene representations" from the contents of a buffer in visual working memory. These representations include a distinguished "anchor" that defines what the sub-scene is about (Itti and Arbib 2006; Peters and Itti 2008). It has been suggested that the putative fact that these representations structure our perception of time explains why we tend to fail to perceive jump-cuts in movies (Carmi and Itti 2008). If our perception of the flow of time is based on the succession of anchored minimal sub-scene representations, our perception of simultaneity is likely mediated by these representations. On Itti and Arbib's view, attentional focus is what supplies the anchor for minimal sub-scene representations, allowing us to connect their proposal with the attention-induced illusion of simultaneity.

By default, surprise or novelty in a visual display, like a suddenly flashing light, attracts anarchic attentional focus and starts a new minimal sub-scene representation. If we have a case in which aspects of successive displays are disposed to be represented in the same minimal sub-scene, we would expect the first thing in temporal order to draw attentional focus and become the anchor of the sub-scene. I suggest that when attentional focus is intentionally kept away from the first flashing dot in the experiment described in Fig. 3 and is forced upon the second flash 40ms later, we get a situation in which two flashes are grouped into the same anchored minimal sub-scene, but where the anchor supplied by attentional focus is the later flash, contrary to the normal course of things. Now, it would make sense to think that the anchor of a minimal sub-scene should be experienced as occurring at the same time as or earlier than the other elements in the sub-

scene, since anarchic attentional focus normally provides anchors for which this is true. This means that if conscious guidance of attention – following the experimenter instruction to focus on a particular place – forces the second flashing dot to be the anchor of a minimal sub-scene that includes the first flashing dot, the two dots would tend to be perceived as at least simultaneous. By contrast, in the ordinary course of perceptual events without consciously forced attention, a flashing dot would immediately attract attentional focus and start a new anchored minimal sub-scene, so there is no reason why a later flashing dot would be experienced as happening at the same time.

To integrate this explanation with the foregoing account of attention, we would change the default background question to something along the lines of "what surprising minimal sub-scene anchor is going on in the foreground?" If we represent a minimal sub-scene as a pair $<A, E>_{\text{Subscene}}$ of an anchor A and a set E of elements in the sub-scene drawn from visual working memory, the default question may be tentatively regimented as $<\lambda x.(\text{Foreground}(x) \& <x,E>_{\text{Subscene}}), D>$. It is then a further question how elements of the minimal sub-scene besides the anchor get filled in.

In sum, the notion that attention in perception involves interpreting perceptual input as an answer to a question, a *default* background question in many cases, can yield an integrated account of a wide variety of effects of attention on phenomenal character in representational terms. In fact, I am not aware of an alternative proposal, representationalist or not, that would account for such a large variety of cases systematically in terms of a compact thesis about the functional role of attention. I suggest that this makes it worth taking the proposal seriously and exploring it further.

5 Conclusion

On the erotetic theory of attention, we bring an interest to perception, regimented as a question we seek to answer. Questions, as understood here, are contents that cognitive systems can represent rather than linguistic items. We process perceptual input as a putative answer to our question in a way that is modulated by attentional focus; attentional focus aims to pick out something that matches what our question is "about." In certain cases, this yields a form of predictive coding: if the contribution of focus matches what our question is about, we can, in certain cases, take it to select one of the possible answers we are entertaining, even though our perceptual input by itself does not supply a full answer. I argued that the erotetic theory of attention provides the resources to develop systematic accounts of a variety of disparate ways in which attentional focus can influence phenomenal character, introducing some order to what may otherwise seem like an arbitrary collection of perceptual curiosities.

References

Adelson, E. (1993). Perceptual organization and the judgment of brightness. *Science*, 262, 2042–2044.

Beaver and Clark (2008). *Sense and sensitivity: how focus determines meaning*. Malden: Wiley-Blackwell.

Block, N (2010). Attention and mental paint. *Philosophical Issues*, 20, 23-63.

Burge, T. (2009) Perceptual objectivity. *Philosophical Review*, 118 (3), 285-324.

Carmi, R., Itti, L. (2005). Why do we fail to perceive jump-cuts in motion pictures?, In: *Proc. Vision Science Society Annual Meeting (VSS05)*.

Carrasco, M. (2009). Attention, cognitive models of. In P. Wilken & T. Bayne & A. Cleeremans (Eds.), *Oxford Companion to Consciousness*. Oxford: Oxford University Press.

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

Carrasco, M., Cigdem, P.-T., & Eckstein, M. (2000). Spatial covert attention increases contrast

sensitivity across the CSF: support for signal enhancement. *Vision Research*, 40, 1203-1215.

Chalmers, D. (2004). The Representational Character of Experience. In B. Leiter (Ed.), *The Future for Philosophy*. Oxford: Oxford University Press.

Fuller, S., & Carrasco, M. (2006). Exogenous attention and color perception: Performance and appearance of saturation and hue. *Vision Research*, 46(23), 4032-47.

Gobell, J., & Carrasco, M. (2005). Attention alters the appearance of spatial frequency and gap effect. *Psychological Science*, 16, 644-651.

Hamblin, C. L. (1958) Questions. *The Australasian Journal of Philosophy, 36*, 159-168.

Harman, G. (1990). The Intrinsic Quality of Experience. *Philosophical Perspectives*, 4, 31-52.

Hohwy, J., Roepstorff, A., Friston, K. (2008). Predictive coding explains binocular rivalry: an epistemological review. *Cognition, 108*, 687-701.

Itti, L., Arbib, M. A. (2006). Attention and the minimal subscene. In: *Action to Language via the Mirror Neuron System*. Arbib, M. A. (Ed.), Cambridge: Cambridge University Press, pp. 289-346;

Itti, L., and Baldi, P. (2005). A principled approach to detecting surprising events in video. *Pro ceedings in IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 631–637.

Itti, L. and Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, 49, 1295-1306.

James, W. (1890/1950). *The Principles of Psychology*. New York: Dover Publications, Ch. XI.

Jiang, Y., Costello, P., Fang, F., Huang, M., and He, S. (2006). A gender- and sexual orientation-dependent spatial attentional effect of invisible images. *PNAS*, 103(45), 17048–52.

Kawabata, N. (1987). Interpretive process of depth in line drawing. *Systems and computers in Japan*, 18(7), 103-9.

Kawabata, N., Yamagami, K. (1978). Visual fixation points and depth perception. *Vision Research*, 18, 853-854.

Krifka, M. (2001). For a structured account of questions and answers. In: C. Féry and W. Sternefeld (Eds.). *Audiatur vox sapientiae. A Festschrift for Achim von Stechow*, Berlin: Akademie-Verlag, 287-319.

Liu, T., Abrams, J., & Carrasco, M. (2009). Voluntary Attention Enhances Contrast Appearance. *Psychological Science*, 20(3), 354-362.

Mazza, V., Turatto, M., and Umilta, C. (2005). Foreground-background segmentation and attention: A change blindness study. *Psychological Research*, *69*, 201-210.

Meng, M., & Tong, F. (2004). Can attention selectively bias bistable perception? Differences

between binocular rivalry and ambiguous figures. *Journal of Vision*, 4(7), 539-551.

Montagna, B., & Carrasco, M. (2006). Transient covert attention and the perceived rate of flicker. *Journal of Vision*, 6(9), 955-65.

Peacocke, C. (1998). Conscious attitudes, attention and self-knowledge. In: C. McDonald, B. Smith, and C. Wright (Eds.). *Knowing our own minds: Essays on self knowledge*. Oxford: Oxford University Press, 63-99.

Peterson, M. A, and Gibson, B. S. (1991). Directing spatial attention within an object: altering the functional equivalence of shape descriptions. *Journal of Experimental Psychology*, *Human Perception and Performance*, 17(1), 170-182.

Peters, R. J., Itti, L. (2008). Congruence between model and human attention reveals unique signatures of critical visual events. In: *Advances in Neural Information Processing Systems, Vol. 20,* Cambridge, MA:MIT Press, 1145-52.

Scholl, B. J., Noles, N. S., Pasheva, V., & Sussman, R. (2003). Talking on a cellular telephone dramatically increases sustained inattentional blindness. *Journal of Vision*, 3:9(156a).

Siegel, S. (2010) The contents of perception. *Stanford Encyclopedia of Philosophy*, Accessed May 2012.

Spence, C., and Parise, C. (2009). Prior-entry: A review. *Consciousness and Cognition*, 19(1), 364-379.

Stelmach, L. B., and Herdman, C. M. (1991). Directed Attention and Perception of Temporal Order. *Journal of Experimental Psychology: Human Perception and Performance*, 17(2), 539–50.

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

Tse, P. U. (2005). Voluntary attention modulates the brightness of overlapping transparent surfaces. *Vision Research*, 45, 1095–98.

Wolfe, J. M. (2000). Attention is fast but volition is slow. *Nature*, 406, 691.

Wright, R. D. and Ward, L. M. (2008). *Orienting of Attention*. Oxford: Oxford University Press.

REFERENCE SUPPRESSED (*forthcoming*). *Mind & Language*.