Sparse, Narcissistic, and Under Attentional Control
Comments on Carolyn Dicey Jennings’ The Standard Theory of Conscious Perception
Pacific Division Meeting of the American Philosophical Association, Seattle, WA, April 2012
Bill Seeley, Bates College

1. Introduction. I think I can say (genuinely) that I agree with (nearly everything) that Carolyn has to say about the relationship between conscious perception and attention. In particular, I agree that subject level interest determines the conscious content of perception, that attention does not merely direct, but constitutively transforms the processing of sensations into the content of perception, and that task salience is the critical variable that drives these processes. Furthermore, I agree that all of this is a matter of neural tuning. What I want to do is to present a model for how this might work derived from a biased competition model for selective attention (Desimone and Duncan, 1995; Pessoa, Kastner, and Ungerleider, 2002). Theorizing and metaphysics of mind can be fun. It's the stuff from which sound hypotheses are born. But my interest, and I get the impression that Carolyn and I are of like minds here, is in whether our best theories describe the kinds of minds that could be run on the actual bodies we have. Against this backdrop, my plan is to raise some questions about what it might mean to say that attentional processes are within a subject’s control, about whether or not sensory processes have the structure to account for the unified experience of conscious perception, and about the potential structure and role of a global subjective standard in attention. My suspicion is that there is substantial agreement between biased competition models and The Standard Theory of Conscious Perception. However, the view of conscious perception that emerges from the former is spatiotemporally sparse and scaffolded by stereotyped sensiromotor contingencies that obviate the need for a common spatio-temporal framework. So, although I think Carolyn is right about the relationship between attention and conscious perception, it may not be for all the right reasons.

2. A short story about the constitutive role of attention in perception. The nature and structure of perception is, has, and perhaps will always remain a slippery topic. The trouble is that our introspective awareness of this crucial aspect of conscious experience doesn't seem to match up to the information processing story we get from either cognitive science or neuroscience. What is a philosopher of mind with naturalistic leanings supposed to do with this kind of a conflict? At the extremes we could either reify the content of conscious experience, treat it as an independent variable in behavior, or dismiss it. However, a more reasonable strategy, the one I take it Carolyn has in mind, would be to a) figure out how the unified experience of consciousness emerges from the best information processing story we have and b) modify our views perception and conscious experience to
accommodate this information. With this goal in mind I'd like to briefly sketch a story about the nature and structure of perception. This story emerges from Dana Ballard, Michael Land, and Mary Hayhoe's work on *visual routines* for everyday activities, Phillipe Schyns' *diagnostic recognition framework* for object recognition, and a biased competition model for selective attention.

2a. Visual Routines. Perceptual and cognitive systems evolved in lock step with, and in the service of the needs of, our bodies. In this context the primary purpose of perception is to provide information to support everyday physical and cognitive behavior in an environment that has remained, by and large, stable over the duration of the history of the species. The key here is that the stability of our physical environment supports stereotyped behaviors, and more importantly, stereotyped perceptual support strategies – we don't need a global spatial model of the environment to accomplish our everyday goals, we only need to know where to direct our attention to recover the information we need when we need it. For instance, in studies of everyday activities like tea- and sandwich-making, participants make very few task irrelevant fixations (Land and Hayhoe, 2001). A short initial scan of the room is made at the start of the task in order to locate salient objects. During this phase participants fixate on task relevant objects about half of the time. After this initial period, however, participants rarely engage in orienting, or looking around, behavior. In fact, fewer than 5% of observed fixations were focused on task irrelevant objects during the performance of the prescribed everyday activities in these experiments. Further, eye movements preceded the motor acts that they were associated with by up to half a second, and the focus of attention tended to move on to the subsequent location prior to the completion of a motor act. Similarly, expert cricket batsmen do not continuously track a pitch across its trajectory. Rather they fixate on the bowler's release point and then shift attention directly to a point just before where the ball will bounce (Land and McLeod, 2001). These studies indicate that perception in ordinary contexts is guided by shared internal scripts that produce stereotyped patterns of attention targeted to the sensiromotor requirements of everyday activities. The purpose of these scripts is to provide just the perceptual information needed at just the right time to support current goal directed behaviors. Dana Ballard has thereby referred to them as *just-in-time-strategies* (Ballard, Hayhoe, and Pelz, 1995). But, more importantly, everyday activities do not appear to depend upon a global world map to unify these diagnostic features into an integrated searchable common spatial framework. Qualitative representations of the relative orientation of salient features within the environment will do (see Mataric, 1991).
Sparse, Narcissistic, & Under Control

2b. Diagnosticity & Biased Competition. Diagnostic features can be defined as sets of sensory features sufficient to enable an organism to perceptually recognize the identity, shape, location, and affordances of objects and events in the environment (Schyns, 1998; see also Palmer, 1999). Some views of objects are more informative than others. For instance, in our ordinary interaction with clocks we want to know what time it is. A view orthogonal to the plane of the clock face is the most informative for this task. A frontal view of a horse, on the other hand, provides very little information about its size and posture. A side view rectifies some of this, but it obscures information about the breadth of the animal across the shoulders. A three quarter view thereby provides us with the most information about the size, strength, and demeanor of the horse. Of course if you want to wind the clock, set the alarm, or check the horse's teeth, a different view would be better. There are two relevant points here. First, the kinds of sparse, directed perceptual strategies adopted in the context of stereotyped everyday activities generalize to a broader range of contexts associated with object recognition and directed search. This makes sense. After all, the structure of object types has also remained generally stable over time, and object recognition is itself a stereotyped everyday activity. Second, diagnosticity is a task specific notion. Altering our behavior alters the way we categorize objects and events. Changing the category under which we identify a perceptual stimulus can alter the way we assign salience to its parts and features. This, in turn, entails that, as discussed above, perceptual systems do not need to generate and search a global model of the detail of an organism's environment in ordinary contexts. A representation of task salient aspects of objects and events that encodes their relative orientation to the organism and each other will do.

2c. Biased competition: unimodal, crossmodal, and affective attention.

The natural environment is replete with sensory information. However, cognitive systems are limited capacity processing systems. In this regard selectivity is a critical feature of perception. Smooth, efficient responses in a replete dynamic environment require a capacity to ignore distracters and focus attention on task-salient behaviorally-significant environmental features, aspects of the local environment salient to both the apical and instrumental goals of an organism. This raises a question. How do perceptual systems assign salience to environmental features. One means is perceptual salience. Some features stand out simply because they contrast starkly with their surround, e.g. sharp contrasts in brightness or color, sudden changes in brightness, or abrupt movements against a static backdrop. However, behaviorally significant features in the environment are not always the most perceptually salient. Therefore cognitive systems must employ an independent means to flexibly bias perception in real time.
Sparse, Narcissistic, & Under Control

Biased competition models for selective attention suggest that in goal directed contexts fronto-parietal attentional networks bias perception by priming sensory systems to the expectation of features diagnostic for task salient objects or events. In other words we attribute salience to sensory features on the fly relative to an interpretation of current behavioral contexts. This entails that perception is naturally tuned to diagnosticity. We perceive just what we need for the current behavioral context just as we need it.

How does this model work? Feedback projections from prefrontal areas associated with spatial working memory, e.g., dorsolateral prefrontal cortex (dIPFC), and premotor areas associated with motor planning and preparation, e.g. supplementary motor areas and premotor cortex (SMA, PMC), to topologically organized areas throughout the visual system enhance the firing rates of populations of neurons that would encode behaviorally salient features, objects, and object parts at expected locations in the local environment and suppresses, or inhibit the firing rates of populations of neurons that would encode task irrelevant features, objects, and etc. (Kastner, 2004; Schubotz and von Cramon, 2003). These cognitive processes influence the sensory encoding of task salient information as early as the lateral geniculate nucleus (LGN) in the thalamus. LGN is the primary relay station between the retina and the visual system. This entails that cortical feedback from prefrontal areas influences the very earliest stages of visual processing. Analogous stories can be told about auditory and somatosensory processing. Further, visual, auditory, and somatosensory attentional networks converge on the superior colliculus (SC). SC is a structure critical to directing eye movements and the crossmodal perceptual integration of visual, auditory, and somatosensory information, bringing discrete unimodal spatial maps into spatial register with one another in order to coordinate attention, synchronizing unimodal processing across these distinct sensory systems via reciprocal priming, and thereby facilitating the production of a coherent multisensory perceptual representation of the environment (see Beck and Kastner, 2009; Grossberg, 1999, 1980; Stein, Sanford, Wallace, Vaughan, and Jiang, 2006; Spence and Driver, 2006).

An analogous story can be told about emotional attention. Reciprocal connectivity linking the orbitfrontal cortex (OFC), amygdala (AMG), and pulvinar (PUL) mediates the influence of affective appraisals, categorical assessments of the predictive value of a stimulus, on bottom-up, low road emotional processing (Pessoa and Adolphs, 2010; Pessoa et al, 2002). Further, reciprocal connectivity between ventromedial prefrontal cortex (vmPFC), anterior cingulate cortex (ACC), and AMG mediates the influence of categorical assessments of affective salience on visceromotor responses constitutive of our gut reactions to emotionally charged stimuli (Duncan and Barrett, 2007). Feedback projections from
AMG influence all levels of visual processing from PUL in the thalamus through area TE in the inferior temporal cortex (associated with complex pattern and object recognition). This entails that affective appraisals influence what we perceive and how we perceive it. Further, reciprocal connectivity between AMG and PUL integrates cortical influences in affective and visual processing at the very earliest stages. The result of all of this is an integrated, crossmodal, biased competition model for selective attention that is a means to direct attention, encode the semantic, task, and biological significance of objects and events in the distal environment, implement working memory, and maintain stable, enduring perceptual representations over iterations of perceptual processes.

2d. Some preliminary conclusions. The suggestion that emerges from this discussion is that the perceptual content of conscious experience is sparse and spatiotemporally disjoint. How can we reconcile this with introspective phenomenological intuitions about the unity and repletefulness of conscious experience. First, notice that perception is closely coupled with behavior and fine tuned to the environmental features diagnostic for the successful performance of those behaviors. In other words, in ordinary contexts the cadence of attention and the content of perception are determined by the goals and task demands of everyday behavior. These goals and task demands themselves structure the conscious expectations that frame our awareness of the dynamics of our environment – we see just what is needed just as we need it. Second, visual acuity falls off significantly as we move away from the central, or foveal region of the visual field – at an eccentricity of 10 degrees from fixation there is approximately an 80% loss in visual acuity. This entails that access to the kind of fine detailed information necessary to object recognition and the performance of fine motor behavior is limited to an area in the center of the visual field approximately the size of one's thumbnail held at arm's length. Taken together these two facts entail that the explicit conscious content of perception is smoothly coupled to our cognitive expectations about the dynamics of our ordinary environment. We don't notice the spatiotemporally disjoint nature of perception because we neither experience nor perceive the jump cuts. Analogously, it feels like the visual field is richly detailed all the way out to its periphery because whenever we look to confirm this intuition we move our fovea and what we find is fine detailed visual information.

3. Three sets of comments:

3a. The structure of sensory systems. One argument that Carolyn makes for the need for a global subjective standard in perception comes from the structure of sensory processing systems. The claim is
that the content of perceptual experience is unified in contrast to the structure of sensory systems, which are both intra- and cross-modally modular and discrete. For instance, the visual system is divided up into discrete parallel subsystems that process color, form and motion information separately and asynchronously, and auditory and visual information are processed by discrete, unimodal processing systems that are not directly interconnected with one another. Therefore, Carolyn argues, the kinds of crossmodal sensory interactions present in perception cannot be accounted for at the level of sensory systems. Rather, a) attention provides the glue that hangs all of this together by b) implementing a for-me-ness a common pointing back towards the subject, or a global subjective standard necessary to integrate the discrete elements of perceptual experience into a unified, interactive whole.

The model of perceptual processing that I have sketched out agrees that what binds perceptual experience is something like a for-me-ness. Further the mechanism for implementing this egocentric, narcissistic perceptual filter is selective attention. However, on this account sensory and attentional systems are not discrete processing systems, but rather form two aspects of an integrated cortico-thalamic circuit. Sensory processing is crossmodally tuned and synchronized directly via attentional feedback, rather than integrated by some aftermarket, post-production processing. This attentional feedback critically involves subcortical areas responsible for sensiromotor integration. Further, these subcortical areas are reciprocally connected with an analogous cortico-amygdala affective attentional circuit so that the emotional salience of environmental features is analogously integrated into sensory processing. Therefore, when conceptualized as a process that unfolds over time, sensory processing has the structure adequate to account for interactions among the elements of perceptual experience precisely because of its internal sensitivity to the for-me-ness of perception, the task salience of diagnostic features in the environment.

3b. Within Subject Control: presubjective/volitional/voluntary. Of course, Carolyn need not disagree with any of this, I don't think. Rather, she may argue that the processes that implement these influences in perception are pre-subjective, sub-personal, or automatic. As a result they do not suffice to accommodate the voluntary, or volitional quality that defines the global, subjective standard as within subject control. The question here is perhaps a question of clarification. The implementation of visual routines in the context of everyday activities is certainly not voluntary or volitional in the sense I take it that Carolyn means. It is not felt as an increase in the phenomenal priority of some set of perceptual features, nor is it associated with the introspectively volitional quality of our cognitive lives. Nonetheless, the targets of visual routines are not determined by either the perceptual salience of
environmental features or the evolved computational structure of sensory systems (see Marr and Nishihara, 1978). They are within subject control in the sense that they are cognitively mediated via top-town, endogenously cued attentional processes. If what Carolyn means by within subject control is simply under the top-down control of neurophysiological systems ordinarily associated with the executive control of behavior, then I think her model is consistent with what I have suggested. But I don't see this as synonymous with her concept of phenomenal priority. In any event, on the account I have given it looks as if the content of conscious experience is under attentional control, and within the subject's control, independent of the kind of subjective phenomenal priority we often associate with explicit conscious experience.

So, what is meant by within a subject's control. Consider DF, Milner and Goodale's patient who suffers from a form of apperceptive agnosia called visual form agnosia. Apperceptive agnosia is, loosely, caused by a breakdown of the connection between sensation and perception. DF doesn't see the world as composed of shapes and forms. As a result she does particularly poorly on visual tasks that require her to report what she sees or match objects and patterns by their visual appearance. However, she seems to be able to use sensiromotor function to overcome this perceptual deficit in some contexts. For instance, Milner and Goodale report that she often used motor rehearsal to improve her performance in manual orientation tasks like posting a card into a slot or tracing a line (Dijkerman and Milner 1997; Milner and Goodale, 2005; Milner and Goodale 2006;). Interestingly, this manual rehearsal does not need to be in the same orientation as the target, e.g. she might trace in the air by her side before copying the orientation lines onto the paper. When asked to stop this behavior, she would report naturally imagining this kind of rehearsal, a process that likewise increased her performance. But when asked to perform a high attentional load distractor task designed to interfere with the capacity for motor simulation her performance would fall off to chance. It would seem, therefore, that DF has a sensiromotor awareness of the environment that is voluntary, volitional, within subject control, exhibits a for-me-ness, and is underwritten by sensiromotor attentional feedback loops (perhaps associated with intact dorsal stream processing). Dennis Proffitt reports a similar result for slope perception in normal perceivers. We ordinarily exaggerate the slopes of hills (for a review see Proffitt, 2006). The degree of exaggeration can be mapped as a power function scaled to our behavioral capacities, e.g. a 5 degree slope is perceived as be 10 degrees, a 10 degree slope is perceived as 25-30 digress, and above 30 degrees – after which we can no longer walk up the hill and need our hands to scramble, crawl, or climb – exaggeration falls off significantly) However, if participants in his studies are asked to place their palms flat on a hinged board (like the plate of a tripod) and angle it to match the slope of a hill perceived face
Sparse, Narcissistic, & Under Control

on, they are remarkably accurate, even though they do not look at the plate to check their performance. These seem to be cases of conscious perceptual awareness of the environment that are within subject control in the right way, and under attentional control, but don’t match the intuitions of felt phenomenal priority or a common spatio-temporal framework that drive Carolyn’s paper.

3c. A global subjective standard. Carolyn defines conscious perception as the experience of integrated information arranged within a common spatio-temporal framework provided by a global subjective standard. There is a sense in which the model I have sketched is a straightforward mechanism for implementing a global subjective standard that is a constitutive constraint on the structure and content of perception. Relative to the goals of our current behavior, in our current context, top-down feedback prioritizes unimodal sensory processing, synchronizes sensory processing across modalities, and thereby integrates crossmodal sensiromotor processing. This model therefore accounts for the experience of integrated information – sensory information is integrated relative to the goals of the agent. However, Carolyn also argues that a common spatio-temporal framework is necessary to locate the disparate elements of conscious experience relative to the egocentric, narcissistic perspective of the agent. The claim is that this is needed to unify the disparate elements of perception into a common spatio-temporal frame for action on the one hand and to discriminate distinct particular instances of common types on the other, to solve the twin problems of unity and diversity. My interest is in the former claim – or the idea that the first operation of attention is to create for itself a field that can be surveyed. On the account I have suggested this phenomenal intuition would seem to be an illusion. We certainly employ some sort of frame, but it need not have the rich global spatial structure suggested by Carolyn. A sparse set of deictic markers pointing towards the relative positions of spatio-temporally discontinuous locations expected to contain task salient features will do (Ballard, Hayhoe, Pook, and Rao, 1997). The illusion of the unity of consciousness, on this account, emerges from the fact that we never look at, are never aware of, the in-betweens, we ignore the jump cuts just as we do the edited structure of a contemporary Hollywood movies in the theater. Likewise the diversity of conscious experience emerges from the fact that these top-down processes tag each of the particular perceptual features that we become aware of with differential salience, salience for a task in a context. In fact, it is only because of this salience that we become aware of them at all. This salience is the critical constitutive feature of conscious perceptual experience and our smooth, flexible, cognitively mediated perceptual sensitivity to it accounts for both its unity and diversity. Perhaps this is just what Carolyn means by a global subjective standard. But note, it is not a common spatio-temporal frame, it is a qualitatively unified narcissistic
frame, bound by the goals of the agent as opposed to an egocentric spatial perspective.

It's fair to say that I agree with at least half of what Carolyn has to say about the relationship between conscious perception and attention. The prioritization work of attention is constitutive of the structure of conscious perception. In this case conscious perception can be defined as the experience of task salient information narcissistically integrated into the goal structure of our current behavior. However, the application of this global subjective standard does not rest on the prior assignment of a common spatio-temporal framework derived from the inheritance of some presubjective neural tuning. Rather, all aspects of perceptual processing, from the thalamus (and SC) on, are under attentional control, directed by learned sensiromotor schema that obviate the need for a prior searchable field. Further, we develop, implement, and update these visual routines on the fly, flexibly adapting to variance in stereotyped sensiromotor contingencies in the local environment in a way that supports the illusion of a replete, universal, objective, spatiotemporal framework.

Appendix - Three quick demonstrations (I’ll use truncated versions of these in the talk).
1. Transformational apparent motion differs from its more familiar cousin, translational apparent motion. In the latter case, figures presented alternately at one location and then another are seen to move back and forth between them. In the former case, a second of two spatially overlapping figures appears to grow out of the first as if the display were animated using a sequence of shapes. Peter Tse & Patrick Cavanaugh have demonstrated that visuomotor interactions influence the perception of transformational apparent motion (Tse and Cavanaugh, 2000). Participants in Tse and Cavanaugh's study sat in front of a computer screen that displayed the Chinese character below. They were instructed that the kanji would disappear and then reappear stroke by stroke. The order in which the strokes reappeared was the order they would have naturally been drawn if one were writing in Chinese. The experimental task was to maintain fixation on the dot in the middle of the display and report on the appearance of the last stroke. Critically, and unbeknownst to the participants, the strokes were not drawn in any direction, but rather appeared on the screen all at once.
Ten American and ten Chinese students studying in the United States participated in the experiment. The American students were all native English speakers and knew no Chinese. The Chinese students were all native Chinese speakers, had been in the United States fewer than ten years, had been at least 18 years old when they came to America, and knew how to write the character. All participants were naive to the purpose of the study. Tse and Cavanaugh hypothesized that native Chinese speakers would see the final stroke appear from left to right, as if it had been drawn in, and that American participants would see it grow out of the long dog-leg stroke from right to left, consistent with basic perceptual grouping processes (Tse et al, 1998). The results of the study confirmed this prediction. Interestingly, Chinese participants with psychophysics backgrounds that were familiar with the relevant perceptual grouping processes all saw the final stroke drawn in from left to right as it should be drawn. Likewise, American students continued to perceive the direction of motion of the appearance of the last stroke from right to left after they were told how to draw the Kanji in debriefing. The conjunction of these latter results demonstrates that motor skill and not declarative knowledge is the productive variable in this perceptual effect.

2. Imagine that this is lecture on Abstract Expressionist painting. I might show you the following image to illustrate the way these artists were focused on the flatness of the surface of the canvas, the viscosity of the paint, and the gesture of the brushstroke. If all went well what you would see in the image is what there is, two brushstrokes in an ambiguously defined visual field, or two marks in low relief. Now, let's change the story a little bit and imagine that I had presented this image as an illustration of Sumi-e painting on late modernist abstraction. My understanding is that the goal of Sumi-e painting is to distill the essence of the subject matter in the fewest possible strokes. This painting happens to be a mountain landscape with a small, austere cabin, bathed in sunlight deep in the valley in the foreground. I would
Sparse, Narcissistic, & Under Control

likely point out the larger mountain looming in the shadow in the background and tell a story about the
efficient way the painter had used just two marks to depict the stillness and majesty of the scene. If all
went well you all would transparently perceive these three objects in depth in the landscape. Of course,
this image (actually drawn by John Hyman) doesn't belong in an art history class at all. It is clearly a work
of political satire – a caricature of Adolf Hitler (Hyman, 1992). Now, if the example worked correctly,
changing the perceptual category to which we assigned this sparse set of perceptual cues changed what
was seen from two objects in shallow relief (abstract marks), to three objects in deep perspective (a
 cabin and two mountains framed against a clear sky), to one object in depth defined by three parts (hair,
 face, and moustache). The exercise illustrates that changing how we contextualize a stimulus changes
how we prioritize and perceive its features.

3. In a recent study, Pessoa, Kastner, and Ungerleider (2002) investigated the influence of attention on
affective perception and attention. Participants maintained fixation on the center of a display where a
range of emotional faces were presented. They were asked to either identify the gender of the face
(attended condition) or to report the relative orientation of bars in the upper corners of the display
(unattended condition). Although participants maintained fixation on the center of the face in both
attended and unattended trials, amygdala response was only recorded for attended trials,
Demonstrating that amygdala responses to emotional stimuli are not automatic, but require the
allocation of attentional resources – or that the perception of valenced stimuli requires attention. These
results come with a caveat. The fate of unattended stimuli is tied attentional load. For instance, in low
attentional load tasks like determining whether words are written in upper case or lower case letters,
Pessoa and his colleagues report high MT activation associated with unattended peripheral motion cues
(MT is the area of the visual system associated with the processing of abstract motion cues). However,
in high attentional load tasks like determining whether or not words contain two syllables, this
responsiveness in MT disappears (Pessoa, Kastner, and Ungerleider, 2003).
Sparse, Narcissistic, & Under Control

Figures from Pessoa et al 2002 (a-c), 2003 (d). (A) shows the procedures for experiment discussed in Pessoa et al, 2002. (B,C) show the relative amygdala responsiveness in attended and unattended trials for fearful(FA/U), happy(HA/U), and neutral(NA/U) faces Pessoa et al, 2002. (D) shows the relative MT responsiveness in the high/low attention tasks reported in Pessoa et al 2003.

References:
Sparse, Narcissistic, & Under Control

Edition (Malden, MA: Blackwell), 266.
Cognitive neuroscience of attention (pp. 144-156). New York: Guilford Press.
Neuroscience, 3(12), 1340-1345.
Land, M. F., and Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities?
Vision Research, 41, 3559-3565.
vision (New York: Oxford University Press).
Pessoa, L. and Adolphs, R. (2010). Emotion processing and the amygdala: form a "low road" to "many
of sensory processing to top-down control. The Journal of Neuroscience, (23)10: 3990-3998.
Pessoa, L., Kastner, S., and Ungerleider, L. G. (2002). Attentional control of the processing of neutral and
Pessoa, L., Kastner, S., and Ungerleider, L. G. (2002). Attentional control of the processing of neutral and
Science, 1(2), 110-122.
T. Watanabe (Ed.), High Level Motion Processing (Cambridge, MA: MIT Press).
cortex: evidence from fMRI and PET. Neurimage 20, S120-S131.
Cognition, 67(2), 147-179.