

# Perception, Attention and Demonstrative Thought: In Defense of a Hybrid Metasemantic Mechanism<sup>1</sup>

Felipe Nogueira de Carvalho

Universidade Federal de Minas Gerais

Forthcoming in *Manuscrito* 43(2), 2020

**Abstract:** Demonstrative thoughts are distinguished by the fact that their contents are determined relationally, via perception, rather than descriptively. Therefore, a fundamental task of a theory of demonstrative thought is to elucidate how facts about visual perception can explain how these thoughts come to have the contents that they do. The purpose of this paper is to investigate how cognitive psychology may help us solve this metasemantic question, through empirical models of visual processing. Although there is a dispute between attentional and non-attentional models concerning the best metasemantic mechanism for demonstrative thoughts, in this paper I will argue in favor of a hybrid model, which combines both types of processes. In this picture, attentional and non-attentional mechanisms are not mutually exclusive, and each plays a specific role in determining the singular content of demonstrative thoughts.

**Keywords:** demonstrative thought; object perception; object representation; attention

## I – Introduction

A visual perception of a particular object in our external environment puts us in a position to engage in a series of cognitive activities in relation to that object. We can identify the object to a hearer with an ostensive act or a demonstrative expression, we can plan a course of action in relation to it, image what it would look like from a different spatial perspective, speculate about its hidden properties and dispositional behaviors, estimate whether it would fit in the space between two other objects, wonder whether it is the same object we have previously encountered on other occasions, and so on.

Thoughts and other cognitive activities directed at particular objects in the world are called “demonstrative thoughts”. The most obvious reason for this terminology is that such thoughts can be linguistically articulated with a demonstrative expression such

---

<sup>1</sup> This research is funded in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (Capes) - Finance Code 001. I would like to thank Ernesto Perini, Eduarda Calado, Carlos Barth, Samuel Maia and Francisco Lages for comments on an earlier draft that led to this article.

as ‘this’ or ‘that’, as a way of identifying the object to a hearer, or to internally articulate an inferential reasoning involving the object (“if this is 30cm in length, and that is 45cm in length, then this will fit inside of that”). But, more importantly, this terminology highlights an important metasemantic question: the singular content of these thoughts is determined “demonstratively”, i.e., through a perceptual relation that is unmediated by concepts and does not depend on the attribution of descriptive material to the referent. It is because demonstrative thoughts reveal this direct connection between subject and object that they have been deemed philosophically interesting.<sup>2</sup>

That is to say, although I can refer to a perceived object with a conceptually complex demonstrative such as “that chair” or “that fig tree on top of the tallest mountain seen in the northern direction”, philosophers generally agree that there is a form of reference that is more simple and direct, something that visual perception makes possible, even in situations where I am not in a position to attribute conceptual material to the object my thought concerns.<sup>3</sup> If I visually perceive a flying object in the sky, I can think, through  $t1$  to  $t3$ , “that’s a bird...that’s a plane...that’s superman”,<sup>4</sup> and still manage to single out a particular object in thought from  $t1$  to  $t3$ , even if I am wrong in my conceptual attributions. This shows that the reference of demonstrative thoughts is not determined in a descriptive manner through conceptual material associated with the object, but by the very fact of my being perceptually related to it, a relation which allows me to visually select the object in my perceptual experience.

On the basis of these observations, philosophers have sought to elucidate the nature of the perceptual relation that puts in a direct (i.e., conceptually unmediated) relation with objects in the world, and which determines the singular content of demonstrative thoughts. In this picture, the “metasemantic problem” of demonstrative thought is to elucidate how certain facts about visual perception can explain how these thoughts come to have the singular contents that they do.

According to Campbell (1997, pp. 56-58), the fundamental problem to be solved in this respect is to explain how the propositional content of a demonstrative thought

---

<sup>2</sup> Philosophical investigations about demonstrative thoughts have their origins in Strawson’s work on demonstrative identification (1959) and Burge’s notion of *de re* belief (1977). But in its current form, the terminology dates back to Peacocke (1981) and Evans (1982). More recent notions of demonstrative thoughts, closer to cognitive psychology, can be found in Campbell (2002), Levine (2010), Wu (2011), and Stazicker (2011). For a critical discussion of these latter views see De Carvalho 2016.

<sup>3</sup> Strawson (1959), Burge (1977), Bach (1987), Smith (2002).

<sup>4</sup> The example comes from Kahneman et al. (1992).

can select an object in an iconic perceptual representation, when both have very different structural properties. Campbell's solution consists in positing conscious attention as the mechanism responsible for selecting objects in an iconic representation of the visual scene, so that this object may be further processed by the agent's cognitive system.

However, the metasemantic problem of demonstrative thoughts isn't fully solved by elucidating how propositional mental contents combine with iconic perceptual contents. After all, even if we manage to show how both kinds of content can interact, all we've done was connect one kind of mental content with another; but we still leave open how, in turn, the iconic content of perception connects to particular objects in the world, which are the referents of our demonstrative thoughts. If we don't want the same problem to arise at every level of analysis by positing further and further levels of content, at some point the world must impose itself onto our perceptual systems in a purely bottom-up manner. In this respect, solving the metasemantic problem of demonstrative thoughts is connected to the task of explaining the intentionality of thought via visual perception.

On the basis of these considerations, it has become commonplace to borrow from cognitive psychology empirical models of object perception, which are supposed to bear the theoretical burden of explaining how objects can be visually selected in the world in a non-conceptual and bottom-up manner. These mechanisms would be responsible for establishing the fundamental perceptual relation that puts us in contact with external objects, explaining how demonstrative thoughts based on this perceptual relation come to have the singular contents that they do.

The purpose of this paper is to investigate how cognitive psychology may help us solve the metasemantic problem, through empirical models of visual processing. With the advance of our scientific knowledge about the visual system, this approach has become increasingly popular in the philosophy of language and mind, so that an explanation of how the mind, through visual perception, connects to the world, acquires scientific status by being grounded on perceptual mechanisms of object representation. In this picture, we resort to the empirical sciences in order to complement philosophical explanations of the intentionality of thought, and, simultaneously, to help us solve the metasemantic problem of demonstrative thoughts.

The structure of the paper is the following: in the next section I will introduce two theoretical constraints that a perceptual mechanism must meet, in order to be considered a direct and non-conceptual metasemantic mechanism for demonstrative

thoughts. Section III will examine a first candidate, based on Pylyshyn's FINST hypothesis (2007), incorporated into a philosophical theory of demonstrative thoughts by Joseph Levine (2010). Once this mechanism is discarded due to lack of scientific evidence, section IV will examine another candidate, namely, object segmentation processes (Rensink 2000, Lamme 2003), incorporated into a philosophical theory of demonstrative thoughts by Athanassios Raftopoulos (2009a,b). The output representations of this mechanism, however, will be too unstable and short-lived, requiring attention in order to be able to refer successfully to objects in the world. But if that is true, it seems that the resulting mechanism fails to meet the theoretical constraints of section II.

Section V will propose a solution to this problem, by reformulating both theoretical constraints in a way that gives us more space of maneuver without losing sight of their main motivation. On the basis of this new formulation, section VI will present a hybrid mechanism composed of both attentional and non-attentional elements, and make precise the role of each in determining the singular content of demonstrative thoughts, as well as sketch some final considerations.

## **II – Two theoretical constraints**

If we will borrow from cognitive psychology perceptual mechanisms of object representation to help us solve the metasemantic problem, there are some conditions such mechanisms must conform to. In order to clarify this point, we can borrow Levine's distinction between direct metasemantic mechanisms, or DMM's, and intentionally mediated mechanisms, or IMM's (2010, pp. 173-75). IMM's are mechanisms that select their referents through the semantic content of other representations. A paradigmatic example would be a descriptive name like Evans' 'Julius', stipulated to refer to "the inventor of the zipper, whoever he is (Evans, 1982, p. 31). DMM's, on the contrary, select their referents directly, by which Levine means with no representational intermediaries (2010, p. 174). The first condition, therefore, concerns the absence of representational intermediaries in the way these mechanisms select their objects. Applied to object representation systems, the first constraint can be formulated in the following manner:

- **DIRECT**: any putative perceptual mechanism must yield as output the lowest representational level where objects are represented in the visual system

In addition, we've seen that these mechanisms must select their objects in a purely bottom-up manner, independent of the application of concepts. On the basis of these considerations, Raftopoulos argues that a second constraint can be formulated along the following lines (2009a, p. 340):

- **NON-CONCEPTUAL:** any putative perceptual mechanism must be cognitively impenetrable, i.e., instantiated by a modular system encapsulated from higher cognition.<sup>5</sup>

On the basis of these two conditions, some mechanisms that have been proposed in the literature may be immediately discarded. According to a popular theory developed by Joseph Campbell (1997/2002), the fundamental perceptual relation that puts us in a direct contact with external objects is an attentional relation. Campbell finds empirical support for this view in Treisman and Gelade's Feature Integration Theory of attention (1980), according to which attention serves as the "glue" that binds various sensory features (such as color or orientation) as features of one and the same object, when attention is consciously allocated to the location occupied by the object. This attentional relation supposedly yields as output the lowest representational level where objects are represented in the visual system, since attention is what makes object representation possible in the first place.

However, it seems that this attentional model does not meet these theoretical constraints. First of all, there is evidence that attention is directed primarily to objects, not locations. These objects are supposed to be pre-attentively represented, and attention is directed to these pre-attentive representations. If this is true, attentional processes cannot yield as output the lowest representational level where objects are represented in the visual system, violating DIRECT above.

Important evidence in this respect comes from the work of Steven Yantis and collaborators, which seeks to explain the automatic capture of attention by sudden object onsets. Yantis considers two hypothesis as to why this happens (1998): perhaps low-level visual processes detect changes in sensory features like luminance, brightness, color or movement in certain locations of the visual field where an object suddenly appears, which causes attention to be automatically drawn to that location. Or, alternatively, as soon as a new object appears in the scene, a pre-attentive representation may

---

<sup>5</sup> The term "cognitive impenetrability" comes from Pylyshyn (1999).

be automatically created for that object, which would prompt the visual system to automatically direct attention to this object in order to extract more information from it.

What would make us decide one way or another? If the sudden appearance of an object is not accompanied by any changes in luminance, brightness, color or movement, but still causes an automatic attentional capture, it would be a good indication that attention is primarily directed at objects, and not locations where certain changes in sensory features are detected. Yantis & Jonides (1984), Yantis & Hillstrom (1994) and Yantis (1998) tested this hypothesis controlling and keeping constant various features such as luminance, brightness, color and movement, whenever a new object appeared in the scene. Even under these conditions, the sudden onset of a new object always captured attention in an automatic manner. Yantis' final conclusion is that attention must be directed to pre-attentive object representations, which would eliminate attention as the metasemantic mechanism we are looking for, since it violates DIRECT above (Yantis, 1998, p. 251).

In addition, there is evidence that attention is not a cognitively impenetrable process. Based on electrophysiological recordings and fMRI studies conducted by Victor Lamme (2003), Raftopoulos argues that the effects of attention are first registered at 200ms after stimulus onset, at a temporal scale where there is already significant interactions between the visual system and higher cognitive centers in the brain (2009b). Attention, in this picture, serves to integrate pre-attentive representations into the whole cognitive context of the agent, which violates NON-CONCEPTUAL above.

Both DIRECT and NON-CONCEPTUAL are reasonable constraints, as they help restrict putative perceptual mechanisms of object representation to direct and non-conceptual metasemantic mechanisms. Although these constraints will be further clarified in section V, they will be provisionally accepted as formulated in this section, and will be used to evaluate putative models of object perception throughout this paper. As an alternative to attentional models, in the next two sections I will present two non-attentional models that have been proposed by philosophers as possible metasemantic mechanisms for demonstrative thoughts, and critically examine them in relation to the theoretical constraints established in this section.

### **III – The FINST model**

The first model to be examined will be Pylyshyn's visual index system, or FINST's<sup>6</sup>, posited as a mechanism of object selection in the cognitively encapsulated early vision system<sup>7</sup>, which automatically "captures" objects in the world through a brute causal relation with no representational intermediaries. This definition makes it an excellent candidate for a direct, non-conceptual metasemantic mechanism, according to the theoretical constraints of section II.

According to Pylyshyn's hypothesis, the FINST system was shaped by evolutionary pressures to be causally sensitive to certain clusters of properties in the world, for these clusters tend to correspond, in the kind of world where our visual system has evolved, to ordinary material objects. As a result, whenever we are confronted with a visual scene, particular objects in the world will "grab" up to four visual indices (which is the maximum number of indices available) automatically and simultaneously, enabling the visual system to individuate and keep track of these objects independently of attention (Pylyshyn, 2001, 2007). The most important evidence in favor of FINST's comes from the Multiple Object Tracking (MOT) experimental paradigm. For if Pylyshyn's hypothesis is correct and the visual system has its own means of individuating and tracking up to four objects independently of attention, it predicts that something like multiple object tracking should be possible, even in conditions where attention cannot be directed to each item to be tracked.

In a typical MOT experiment, the goal is to track four targets as they move randomly among qualitatively identical distractors. The experiment begins as the four targets are identified by a cue (such as blinking on and off), and then move across the screen amidst a number of distractors. At the end of experiment all objects come to a stop and one of them is randomly identified, and the subject is supposed to say if this object is a target or a distractor.<sup>8</sup> This experiment has been widely replicated in many laboratories, and results indicate a high success rate of 85% on average, which invalidates an explanation in terms of random selection of targets at the end of the experiment

---

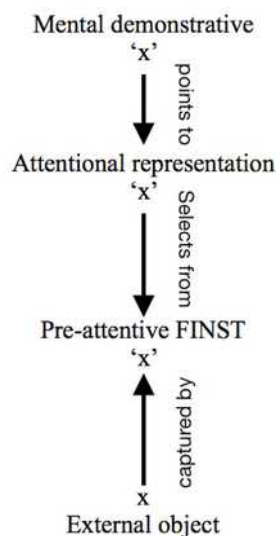
6 FINST's stand for "FingersFINgers ontiation", in order to capture Pylyshyn's imaginative analogy with the superhero "Plastic Man", who can stick his fingers on particular objects as they move around him, affording him means to refer to the objects on the tips of his fingers without having to attend to each of these objects (Pylyshyn, 2007, pp. 13-14).

7 The early vision system is functionally defined by Pylyshyn as the part of the visual system that is encapsulated from the remainder of cognition (1999). Raftopoulos proposes a definition of this system in terms of its temporal properties (2009b), as the processing that occurs up until 150ms after stimulus onset, when processing is still restricted to visual areas (see section IV).

8 The reader is invited to try an online version of the experiment at: <http://perception.yale.edu/Brian/demos/MOT-Basics.html>

(Pylyshyn, 2007, p. 36). With five targets, however, performance drops drastically, which corroborates Pylyshyn’s hypothesis about the set-size limitations of this mechanism.

On the basis of this model, Joseph Levine develops a mental semantic for demonstrative thoughts with a representational hierarchy structured into three levels (2010). On the top level we find mental demonstratives such as ‘this’, whose content is a “mental pointer” that points to an underlying perceptual representation. But rather than pointing directly to visual indices, it points to an attentional representation – the intermediary level – where only one object is visually selected in experience. Attentional processes, in turn, select one of the four available visual indices in the lower level pre-attentive representation, which are captured in an automatic, direct and non-conceptual manner by objects in the world. The reference of mental demonstratives, in this model, is determined by the visual index captured by an external object, but in order for one to be able to think of this object, attention must be drawn to it. Levine’s theoretical model can be captured in the following figure:



(Figure 1)

Although at first sight the FINST model seems like an excellent candidate for a direct and cognitively impenetrable metasemantic mechanism, when we look at other evidence and other explanations for these experimental results, the appeal of the model weakens significantly. As we shall see, the same results may be equally explained by more parsimonious attentional models, based on well-established scientific facts about the benefits of attention and the limits of working memory. This evidence raises serious



problems not only for the pre-attentive status of the FINST mechanism, but for its very relevance to a philosophical theory of demonstrative thoughts.

Pylyshyn's main reason for characterizing FINST's as a pre-attentive mechanism is that an attentional mechanism could not possibly explain the high success rate of 85% observed in MOT experiments. For suppose a subject must direct her attention to each target to be tracked in a serial manner, so as to encode its location; then, as targets move among distractors, the subject must quickly revisit each encoded location, shift attention to the object immediately adjacent to it, update the encoded location, and so on successively for each target to be tracked. Computer simulations have showed that even with very conservative estimates on the timescales of these attentional shifts, the success rate of this strategy would not surpass 30% (Pylyshyn, 2007, pp. 36-37).

This argument, however, presupposes a spotlight model of attention (Posner et al. 1980), where attention moves like a spotlight that scans the visual scene in a serial manner. But there are other models where attention does not work like a single spotlight but can be divided among multiple foci. In an adaptation of Posner's classical spatial cueing paradigm, Awh and Pashler have shown that cues simultaneously presented in multiple regions of the visual field yielded benefits for all these regions, but not for intermediary regions (2000). These results cannot be explained in a spotlight model, which would predict attentional benefits in intermediary regions as attention moved from one cued location to another.

On the basis of these observations, we can propose an alternative explanation for MOT based on multifocal attention. In Cavanagh and Alvarez's model (2005), for example, targets are simultaneously tracked by independent foci of attention, guided by a control process that keeps selection centered over the targets as they move across the screen. This process is supplemented by an encoding stream transmitting target information to higher cognitive processes, which control verbal reports at the end of the task. In this model, the set-size limitation of four items observed in MOT tasks is not explained by the number of available visual indices, but by working memory limitations, which can only deal efficiently with an average of four items at a time.<sup>9</sup>

Finally, there is a curious fact about MOT that seems to be a problem for the FINST model. As we have seen, at the end of a MOT task it is possible to distinguish a target from a distractor in a very efficient manner, with a success rate of 85% on aver-

---

<sup>9</sup> Kahneman et al. (1992).

age. However, it is extremely difficult to indicate which particular target that is, among the four indicated. That is to say, if we mentally label each target to be tracked with the letters A, B, C and D, at the end of the task we would know if a given object is a target or a distractor, but we would be unable to indicate whether it is target A, B, C, or D, or whether “this target” (identified in the beginning of the task) is identical to “this target” (identified at the end of the task).<sup>10</sup>

But if the high success rate of MOT tasks is explained by the automatic capture of visual indices by each object to be tracked, this shouldn't happen. After all, one of the main motivations for positing visual indices is to give the visual system the means to individuate and track objects in an automatic manner, where each object is individuated by a numerically distinct visual index. It is precisely for this reason that Pylyshyn compares his visual indices to “fingers” that point to particular objects, as in the analogy with “Plastic Man”:

It seemed to me that the superhero (...) had what we needed to solve the identity-tracking or reidentification problem. Plastic Man would have been able to place a finger on each of the salient objects (...). Then no matter where he focused his attention he would have a way to refer to the individual parts (...) so long as he kept one of his fingers on it. Even if we assume that he could not detect any information with his finger tips, Plastic Man would still be able to think “this finger” and “that finger” and thus be able to refer to individual things that his fingers were touching. (Pylyshyn, 2007, p. 13)

But if Plastic Man is simultaneously tracking an object with his index finger and another with his ring finger, he should have no problem distinguishing, at the end of the tracking period, one object from another; each finger, in Pylyshyn's metaphor, provides a unique address for each target to be tracked, which should provide means for the superhero to distinguish “this object” (on the tip of his index finger) as distinct from “that object” (on the tip of his ring finger). But, on the contrary, it seems that this mechanism is systematically confusing targets for one another. It is still possible to maintain the identity of the targets as a whole, but not the identity of individual targets.

These observations weaken considerably the motivation for positing visual indices in the first place. A more apt analogy would be a “closed hand”, which “holds” the targets to be tracked, distinguishing them from other objects outside the hand, but concealing individuating information about targets inside the closed hand. This is exactly

---

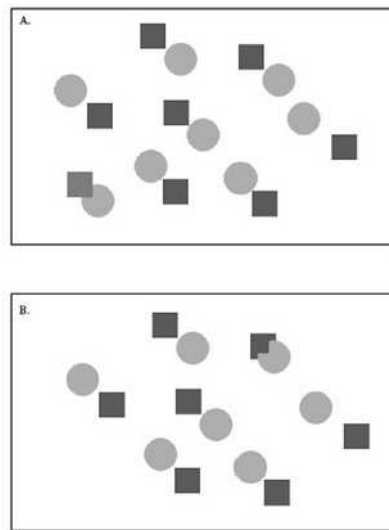
<sup>10</sup> This curious fact was first noticed by Scholl (2009).

what Rensink proposes with his coherence theory of attention (2000), where attention works like a hand that holds up to four visual units, allowing a subject to track them as they move across the visual scene. Rensink even suggests that the term FINST (fingers of instantiation) should be replaced by HANST (hand of instantiation), which describes in a more appropriate manner how attention is focused on the targets as a set (Rensink, 2000, p. 27).

On the basis of these observations, it is reasonable to suppose that a multi-focal attentional model, or a coherence theory of attention, explain the same data from MOT as the FINST model, while explaining further facts that the latter has trouble accommodating. In addition, these attentional models are more parsimonious, as they are based on well-established scientific facts about the benefits of attention and the limitations of working memory, rather than positing a pre-attentive mechanism for which we have no other independent evidence. This leads us to conclude that the main evidence in favor of the FINST model, obtained through MOT tasks, does not favor the existence of a pre-attentive metasemantic mechanism for demonstrative thoughts.

Of course, this does not mean that such a mechanism does not exist. After all, even if these attentional models are correct, we still need to explain how attention is simultaneously directed to objects, and not regions of the visual field (as suggested by Yantis and collaborators). Some pre-attentive mechanism must be responsible for parsing the visual scene into discrete units, to which attention may be allocated. There is empirical evidence, for example, that the visual system amodally completes partially occluded objects during the very first stages of perceptual processing, before the allocation of attention.

Take, for example, the two images represented in figure 2 below. If the goal is to find the notched “pac man” shape among the other shapes, this can be done effortlessly and easily in image B, no matter how many additional shapes are added to the image (a feature mark of automatic and parallel processing). The visual search in figure A, however, is slower, requiring one to serially attend to each item until the notched figure is found. Search time also increases progressively with the amount of shapes added to the image, which is a feature mark of a serial attentional process (Driver et al., 2001).



(Figure 2)

This leads us to conclude that the visual field over which attention roams already contains amodally completed objects. This explains the difficulty in finding the notched shape in image A, since the shape is already represented pre-attentively as a full circle. What this evidence reveals, however, is not a pre-attentive FINST mechanism, but low-level processes of object segmentation, responsible for organizing the initial visual input into discrete units before the allocation of attention. Even Pylyshyn is ready to admit that the assignment of visual indices would presuppose object segmentation processes, as can be seen in the following passage:

In assigning indexes, some cluster of visual features must first be segregated from the background or picked out as a unit (...). Until some part of the visual field is segregated in this way, no visual operation can be applied to it since it does not exist as something distinct from the entire field. (Pylyshyn, 2001, p. 145)

To conclude this section, visual indices cannot be the perceptual metasemantic mechanism we are looking for in a theory of demonstrative thoughts. If we want to find support in cognitive psychology for a direct and non-conceptual metasemantic mechanism, we must look to an even earlier level of perceptual processing, where segmentation processes parse the visual scene into discrete units in a purely bottom-up manner. This is precisely Raftopoulos' proposal, which will be examined in the next section.

#### **IV – Segmentation process and proto-objects**

We've seen in section II that according to the NON-CONCEPTUAL constraint, any putative mechanism must select objects in the world in a purely bottom-up manner. According to Raftopoulos (2009a,b), such a mechanism can be found in object segmentation processes. In order to show that this mechanism satisfies the NON-CONCEPTUAL constraint, Raftopoulos presents evidence of a level of visual processing that is unaffected by top-down signals from higher cognitive centers in the brain. This evidence comes from the work of Victor Lamme (2003), obtained through electrophysiological recordings and fMRI studies, which show that up until 150ms after stimulus onset, information processing is restricted to visual areas.

On the basis of this evidence, Raftopoulos defines 'perception' properly speaking as the kind of processing that occurs at this timescale, and identifies the representational content of perception with neural states in the early vision system during this interval (Raftopoulos 2009a, p. 341). In this picture, questions about the content and structure of perception become purely empirical questions, to be resolved by cognitive science. Only scientific investigation will tell us what these neural states are sensitive to and what they encode, before the modulatory effects of higher cognition reach perceptual processing.

Evidence from Lamme (2003) and Rensink (2000) shows that neural populations in the early vision system, at temporal scales up until 150ms after stimulus onset, encode a structural representation of the scene where particular objects – or proto-objects<sup>11</sup> – are segregated from the background and represented as discrete visual units. This evidence allows Raftopoulos to include objects in the content of perception, and to put forward the processes responsible for representing objects in this manner – object segmentation processes – as a direct and non-conceptual metasemantic mechanism for demonstrative thoughts.

In Lamme's model of visual processing, which Raftopoulos presupposes in his theory, there are three processing stages, distinguished by temporal properties: the feed-forward sweep (FFS), local recurrent processing (LRP) and global recurrent processing (GRP). The FFS begins at 40ms after stimulus onset, when the first patterns of activation are registered in V1, and lasts until 100-120ms with the activation of most visual areas in the dorsal and ventral streams. As the name indicates, neural activity at this level moves only forward, never laterally or backwards. There is very little perceptual organization at this point, and no segregation between figure and background. Some

---

<sup>11</sup> The nature of these proto-objects will be discussed shortly.

sensory properties are detected, but not attributed to particular visual elements. Stimuli at this temporal scale are not consciously perceived (Lamme, 2003, pp. 14-15).

The first signs of recurrent processing (LRP) are registered only at 100-150ms after stimulus onset, when lateral and feedback connections are established in the same visual areas activated during the FFS, strengthening the connections between different neural populations that represent various sensory properties. According to Lamme, a perceptual representation during the LRP consists in “tentatively bound features and surfaces” (2003, p. 17), which may be overridden or strengthened by subsequent attentional processes. When visual information reaches areas of executive and mnemonic control (i.e., frontal, prefrontal and temporal cortices), at about 200ms after stimulus onset, this information is inserted into the overall cognitive context of the agent, becoming integrated with plans, beliefs, intentions, background knowledge, etc. This is the level of global recurrent processing (GRP), where the effects of attention are first registered.

More importantly for Raftopoulos’ proposal, information processing during the LRP is still restricted to the visual system, and therefore cognitively impenetrable. But as long as discrete visual units, which correspond to particular objects in the world, are represented by populations of neurons during the LRP, as the outputs of object segmentation processes, this process qualifies as a direct and non-conceptual metasemantic mechanism for demonstrative thoughts. As recurrent processing for Lamme is the neural correlate of consciousness, at this level of processing the perceptual representation is already conscious, although in a format that is iconic, short-lived, and not easily reportable (Lamme, 2003, p. 16). To borrow a distinction from Ned Block (1995), we would have phenomenal consciousness of this representation, but not access consciousness, which requires attention and global recurrent processing. As Raftopoulos and Müller put it:

We argue that causal chains relating the world with mental acts of perceptual demonstration single out the demonstrata and attach mental particulars to things. In a linguistic context our claim is that these causal chains fix the reference of the perceptual demonstratives in a nonconceptual and nondescriptive way. The causal relation is provided by the nonconceptual contents of perceptual states that are retrieved in bottom-up ways from a visual scene by means of preattentive object-centered segmentation processes (Raftopoulos & Müller, 2006, p. 253).

Although at first sight Raftopoulos’ model seems to satisfy both DIRECT and NON-CONCEPTUAL constraints, a more careful examination will reveal some prob-

lems regarding the first. The main problem, as we shall see, is that although the first condition states that any putative mechanism must yield as output the lowest representational level where objects are represented in the visual system, in Raftopoulos' model the outputs of object segmentation processes are only proto-objects, and it is not clear they can bear this theoretical burden.

Raftopoulos' notion of proto-object comes from Rensink<sup>12</sup>, where they are defined in the following terms:

1. Proto-objects are the highest-level outputs of low-level vision;
2. Proto-objects are the lowest level operands upon which attentional processes act (Rensink, 2000, p. 22).

In Rensink's model, the function of low-level vision is to provide a "quick and dirty" interpretation of the visual scene, a rough sketch that provides the basic "gist" of the structure of the scene. In this rough structural sketch, visual units – or proto-objects – are simultaneously represented, although at this point these representations are unstable and short-lived. The function of attention in Rensink's model is to endow these unstable representations with greater spatiotemporal coherence. Attention, as we've briefly seen in section III, works like a "hand" that "holds" a small number of proto-objects – around four – in order to form a "coherence field" around them, a more stable representational structure that persists as long as attention is sustained over these items, allowing them to enter visual short-term memory. Once attention is disengaged, the coherence field dissolves into its unstable constituents (the proto-objects).

So far this model is compatible with Lamme's, where pre-attentive processing during the FFS and the LRP provides a rough structural sketch of the visual scene constituted by discrete visual units. Moreover, Rensink also agrees that we have only phenomenal consciousness of this representation, which is constantly regenerated as our eyes move across the scene. As attention for Rensink is necessary in order to see change<sup>13</sup>, we are not aware of the way this representation is in constant flux; we are only phenomenally aware of the basic structural aspects of the scene, a virtual representation that seems stable and constant to us but that is constantly dissolving and regenerating.

---

<sup>12</sup> As Raftopoulos himself is ready to admit (2009b, p. 21).

<sup>13</sup> One of the goals of Rensink's coherence theory is to provide an explanation of inattentional blindness, or the incapacity to perceive change when they occur outside the focus of attention. (Rensink, 2000, p. 19).

However – and here is where Raftopoulos’ model runs into trouble – in Rensink’s theory proto-objects have an extremely limited spatiotemporal coherence, decaying after a few hundred milliseconds or being immediately replaced whenever a new stimulus appears in the same retinal location where a proto-object was previously detected (Rensink, 2000, p. 20). Rensink’s main conclusion is that attention is required for this representation to persist for more than a few hundred milliseconds (Rensink, 2000, p. 23).

These observations strongly suggest that proto-objects cannot meet the DIRECT constraint from section II. After all, if proto-object representations last no longer than a single eye saccade of a few hundred milliseconds, and are immediately replaced by the representation of another proto-object that appears in the same retinal location, this mechanism cannot, on its own, pick out particular objects; it would constantly equivocate between two distinct objects that appear in the same retinal location, and it wouldn’t be able to track a single object that moves from one adjacent location to another. A perceptual representation of an object, at the very least, is something that persists in time, allowing us to track the object in space during a period of observation, and grounds our capacity to affirm that “this object” at position  $p1$  and time  $t1$  is the same as “this object” at position  $p2$  and time  $t2$ . Proto-objects do not meet this requirement, and therefore these representations do not constitute the lowest representational level where objects are represented in the visual system. We are thus led to conclude that object segmentation processes cannot, on their own, solve the metasemantic problem of demonstrative thoughts.

But if Rensink is right and attention is required to maintain the numerical identity of an object in time, then perhaps we should reconsider the outputs of attentional processes as the lowest representational level where objects are first represented in the visual system. But if this is the case, then we seem to have reached an impasse: on the one hand, genuine object representations are only possible with attention. On the other hand, attentional processes are not cognitively impenetrable according to evidence from Victor Lamme (2003). How do we resolve this impasse?

A possible conclusion would be that none of the mechanisms examined so far are capable of meeting both theoretical constraints at the same time, and therefore we should seek further alternatives from cognitive psychology. This conclusion, however, would be too hasty. In the next section I will argue that the observations put forward in



this section point to a reformulation of both theoretical constraints from section II. Although these are reasonable constraints that should not be abandoned, some distinctions and clarifications are in order for the conflict to dissipate. This will be the main goal of section V.

## **V – Two theoretical constraints, clarified and reformulated**

An important clarification concerning DIRECT was already introduced in section IV. As we've seen, it is not enough for a structural representation of a visual scene to contain discrete perceptual items; these representations also need to persist in time as the agent and object move in space, under the risk of continuous referential equivocation. Therefore, when we ask cognitive psychology how objects are represented in the visual system, there are two different things we want to know:

1. **Individuation**: how are visual units segregated from the background and from one another in a visual array?
2. **Maintenance of numerical identity**: how can representations of these visual units persist in time, through successive movements of the object and the sensory organ during a period of observation, so that the object's numerical identity is maintained?

The second question naturally presupposes the first, since an object needs to be segregated and discriminated from the background before the representation can persist in time. Therefore, when we say that a mechanism of object representation should not be representationally mediated, we are talking about the individuation question. The moment when external objects first impose themselves onto the visual system is when the visual system is able to spatially differentiate them from one another in a structural representation of the visual scene. This mechanism must in fact be unmediated by other representations, if we want to connect mind and world through visual perception.

However, this is not yet the lowest representational level where we find object representations in the visual system, since these representations still lack a minimal spatiotemporal coherence to be able to refer to objects properly speaking. The DIRECT theoretical constraint can therefore be distinguished into two sub-conditions, each concerning one aspect of object representation:

- **DIRECT**<sup>i</sup>: Mechanisms of **individuation** must be direct, i.e., with no representational intermediaries;
- **DIRECT**<sup>m</sup>: Mechanisms responsible for the **maintenance of numerical identity** must yield as output the lowest representational level where objects are represented in the visual system.

These observations point to a hybrid metasemantic mechanism for demonstrative thoughts, combining both attentional and pre-attentive elements in each sub-condition specified above. It is important to notice, however, that not any attentional or pre-attentive model can be used as part of this hybrid mechanism. We could not find convincing evidence for Pylyshyn's FINST model, for example, since the main evidence in its favor could be explained by more parsimonious attentional models, that are also able to explain other phenomena that the FINST model has trouble accommodating. We were, however, able to find good evidence for pre-attentive processes of object segmentation, responsible for individuating perceptual units (proto-objects) in a visual array in a purely bottom-up manner. These processes will be presupposed as mechanisms of individuation.

Similarly, Campbell's attentional model, briefly discussed in section II, must also be discarded, since in this model attention is directed to locations, so that the various sensory features detected at that location can be bound together as properties of a single object. This model, and the empirical theory it presupposes, does not conform to the evidence produced by Yantis and collaborators (section II), according to which attention is directed to pre-attentive (proto)object representations. In Rensink's theory, on the other hand, the function of attention is to endow unstable pre-attentive proto-object representations with greater spatiotemporal coherence. This theory will therefore be presupposed as an attentional mechanism of maintenance of numerical identity.

But before this hybrid mechanism can finally be explained in more detail in section VI, an important question remains open. According to the NON-CONCEPTUAL constraint from section II, a mechanism of object representation must be cognitively impenetrable, independent of the application of concepts. But attention, as Lamme has shown, does not meet this constraint. How, then, can the output of an attentional process be the lowest representational level where objects first appear in the visual system? If this is the case, then this mechanism does not meet NON-CONCEPTUAL, and the

whole model is compromised.

But here we should make a distinction between a mechanism mentioning the application of concepts in the explanation of its basic operation, and a mechanism operating simultaneously to an application of concepts that is external to it. To go back to Levine's example, the intentionally mediated metasemantic mechanism behind the name 'Julius' mentions the application of concepts in the explanation of its basic operation, since the name refers in virtue of the conceptual content of the representation "the inventor of the zipper." But in Rensink's coherence theory, the function of attention is just to endow unstable proto-object representations with greater spatiotemporal coherence, and nothing in the explanation of the basic operation of this mechanism mentions the application of concepts. Even if at the temporal scale this mechanism operates there are already recurrent connections with higher cognitive centers in the brain, this at most shows that concepts may be applied to perception at the same temporal scale, but it does not show that this application takes place through the mechanism in question. Indeed, in Rensink's theory attentional representations acquire greater spatiotemporal coherence merely in virtue of entering visual short-term memory, and they can be iconic and non-conceptual (Rensink 2000: 26). On the basis of these observations, we can reformulate the NON-CONCEPTUAL constraint in the following terms:

- **NON-CONCEPTUAL'**: A perceptual metasemantic mechanism for demonstrative thoughts must not mention the application of concepts in the explanation of its basic operation.

Thus reformulated, Rensink's theory can now satisfy this theoretical constraint, insofar as the function of attention is just to endow iconic proto-object representations with greater spatiotemporal coherence, by allowing them to enter visual short-term memory. This move allows attentional processes to be incorporated into the hybrid mechanism that will be presented in the next section. It is important to notice that even after both theoretical constraints were reformulated, the main motivation behind them was nonetheless preserved, which is to restrict putative perceptual mechanisms to direct and non-conceptual metasemantic mechanisms. Reformulating the two constraints in this manner has therefore been proven advantageous, affording more space of maneuver without losing sight of the main motivation behind them.

## **VI – Conclusion: in defense of a hybrid metasemantic mechanism for demonstrative thoughts**

In this paper I introduced the philosophical notion of “demonstrative thoughts”, as cognitive activities directed at particular objects in the world, based on the visual perception of these objects. One of the main functions of this terminology is to indicate that the singular content of these thoughts is not determined satisfactorily, through the attribution of descriptive material to the object, but “demonstratively”, through a perceptual relation between subject and object established at the time of the perception. It is precisely because they reveal this “direct” (i.e., conceptually unmediated) relation between subject and object that demonstrative thoughts are philosophically interesting (section I).

A fundamental task of a theory of demonstrative thoughts is to elucidate this fundamental perceptual relation that puts us in a direct contact with objects in the world, which explains how demonstrative thoughts come to have the contents they do. I’ve called this the metasemantic problem of demonstrative thoughts. An approach that has become increasingly popular in the last two decades is to borrow empirical models of visual processing from cognitive science. The basic presupposition behind this approach is that perceptual mechanisms of object representation may help us solve the metasemantic problem, according to some pre-established theoretical constraints (section II).

I then examined two putative mechanisms in light of these theoretical constraints, starting with Pylyshyn’s FINST model (2001/2007), as incorporated into a philosophical theory of demonstrative thoughts by Joseph Levine (2010). After arguing that available evidence does not support the existence of this mechanism, and that the same experimental results may be explained by more parsimonious attentional models (section III), I looked to an earlier level of perceptual processing, involving object segmentation processes (section IV). This was Raftopoulos’ proposal to solve the metasemantic problem of demonstrative thoughts (2009a,b). The proto-object representations at this level of processing, however, were too unstable and short-lived, being incapable of determining the singular content of demonstrative thoughts. One possible solution, based on Rensink’s coherence theory of attention (2000), is to posit attention as the process responsible for endowing these unstable representations with greater spatiotemporal coherence. Attentional mechanisms, however, do not seem to meet the NON-CONCEPTUAL theoretical constraint, which leads us to an impasse: either an atten-

tional mechanism meets the first but not the second theoretical constraint, or a pre-attentive mechanism meets the second but not the first.

A solution to this impasse was found by reformulating both theoretical constraints, so as to allow a more flexible space of maneuver but without losing sight of the main motivation behind these constraints (section V). Finally, on the basis of this reformulation, and on the empirical evidence presented throughout this paper, we can propose a hybrid metasemantic mechanism that perceptually determines the singular content of demonstrative thoughts:

First of all, pre-attentive processes of object segmentation discriminate visual units in a visual array in a purely bottom-up manner with no representational intermediaries, connecting mind and world in a direct and conceptually unmediated manner. These units, however, are not yet object representations, but proto-objects with very limited spatiotemporal coherence. With the allocation of attention these representations are endowed with greater spatiotemporal coherence by entering visual short-term memory, allowing the visual system to represent a particular object that retains its numerical identity through time and movement during a period of observation. The result is a spatiotemporally coherent perceptual representation that represents particular objects in the world with an iconic structure in visual short-term memory.

On the basis of these perceptions, an agent can engage in a series of cognitive activities in relation to the particular object perceived (demonstrative thoughts). In this case, the singular content of these thoughts is determined by the perceptual relation between subject and object established when the object was first segregated from the background by object segmentation processes, and the resulting representation endowed with greater spatiotemporal coherence through attention, allowing the agent to select just that object in experience.

These observations lead us to conclude that Joseph Levine is basically correct in postulating a hierarchy of three representational levels, although he is mistaken as to the pre-attentive mechanism specified at the first level, is vague as to the attentional mechanism presupposed in the intermediary level, and construes conceptual content as abstract symbols in a language of thought, a view we need not endorse.<sup>14</sup> We can, however, stick

---

<sup>14</sup> The Language of Thought, or LOT, is a representational theory of mind developed by Jerry Fodor (1975), where thinking consists in the manipulation of abstract symbols in a mental language. This theory, which is assumed by Levine, remains controversial, although for reasons of space I will not engage with it here.

to the basic idea of a three level hierarchy as a useful schema to capture the structure and function of each level, as well as the interactions between them. Adapted to the present discussion, this model can be reconstructed and reinterpreted in the following terms:

LEVEL	CONTENT	STRUCTURE	FUNCTION
Thought/Language	“This(x) is F”	Conceptual	Communication/inferential reasoning/conscious deliberation/etc.
Attentional representation	(x)F	Non-conceptual	To endow pre-attentive representations with greater spatiotemporal coherence
Pre-attentive representation	(x)F	Non-conceptual	To segregate and discriminate visual units from the background

Property ‘F’ in the table above should be understood as a basic sensory feature, such as ‘rectangular’ or ‘red’, that can figure in the content of perceptual representations already at the lowest pre-attentive level. The attentional level immediately above it refers to attended object representations that enter visual short-term memory, which retain the iconic structure from the pre-attentive level but gains greater spatiotemporal coherence. The choice of representing the external object as x(F) is to mark a structural isomorphism to the pre-attentive and attentional iconic representations, while simultaneously marking a structural difference from the conceptual representation “this is F”.

According to Burge (2010), only conceptual contents exhibit a genuine predicative structure, where the application of the predicate ‘...is F’ can be separated from the subject ‘this’ in a way that both can be individually combined with the content of other conceptual representations: the property ‘F’ can be applied to other objects, at the same time that other properties may be applied to the object that the demonstrative ‘this’ refers to.<sup>15</sup> In perception, however, general elements (sensory features) and singular elements (object representations) are always applied together. What we perceive, in other words, are objects bearing properties, and properties *as in* particular objects. These two elements cannot be “peeled off” from one another so as to individually combine with other representations. This non-conceptual structure, according to Burge, can be captured with a noun phrase such as ‘this x F’ (i.e., ‘this red object’), in contrast with a gen-

<sup>15</sup> This is similar to Evans’ “generality constraint”, posited as a characteristic feature of conceptual thought (1982).

uine predicative structure like ‘this x is F’ (2010, pp. 541-4).

Burge’s proposal to structurally demarcate conceptual and non-conceptual contents is compatible with the table above, where the perceptual representation  $x(F)$  marks the inseparability of the singular element ‘x’ and the general element ‘F’. When we engage in cognitive activities directed at particular objects in the world, however, the object attentively selected in experience can be referred to with a demonstrative such as ‘this’, and one of its sensory features with the concept ‘F’. We need not, however, take the elements ‘this’, ‘is’ and ‘F’ in the conceptual representation to be abstract symbols in a language of thought, as Levine proposes. Rather, this predicative structure, following Burge, serves only to capture certain cognitive abilities on the part of the subject, where these elements can be separately combined with other conceptual representations in the form of deliberations, suppositions, inferential reasonings, etc., as a characteristic feature of demonstrative thoughts. The object these thoughts concern is none other than the object represented in an iconic and non-conceptual manner by the hybrid mechanism described above, which anchors these cognitive activities to the world.

In this manner, I hope to have showed how empirical models from cognitive psychology may complement philosophical questions concerning the intentionality of thought and the determination of singular mental contents. Before concluding, however, it must be admitted that I have treated the maintenance of numerical identity question in a simplified manner. In this paper I focused on perceptual abilities to track the spatiotemporal trajectory of an object during a period of observation, but it is clear that this question may acquire increasingly higher levels of conceptual complexity, as more sophisticated cognitive strategies are required to identify and reidentify an object through space and time. This is particularly clear during longer periods of non-observation or through substantial qualitative changes, where the capacity to maintain the numerical identity of an object will mobilize cognitive resources that are more complex than mere attentional abilities.

Although some philosophers have said that singular contents are only possible in the presence of this more complex cognitive apparatus<sup>16</sup>, I see no reason to deny that singular contents may already be available at the level of these more primitive perceptual abilities. In this picture, the capacity to maintain the numerical identity of an object through space and time take place in a continuum, and is a matter of degree. It has its

---

<sup>16</sup> Evans (1982), Quine (1995), Hatfield (2009), among others.

origins in more primitive attentional abilities – where singular contents are already available to characterize the mental state of an agent who keeps track of an object of perception – but acquires higher levels of conceptual complexity as the agent’s cognitive system develops along with the kinds of challenges she faces in her external environment. To choose one particular point or another in this continuum, where singular contents suddenly become available, seems like an arbitrary choice to me.<sup>17</sup>

Object segmentation processes and selective attention, which allow us to individuate and track an object during a period of observation, mark the beginnings of our conception of the world as structured into particular objects that persist in time. When we cognitively engage with these objects, we are exercising demonstrative thought characterized by singular contents, which concern objects that have been pre-attentively segregated and attentively selected.

#### References:

- . AWH, E., PASHLER, H. “Evidence for Split Attentional Foci.” *Journal of Experimental Psychology: Human Perception and Performance*, 26 (2), pp. 834–846, 2000.
- . BACH, K. *Thought and Reference*. Oxford: Oxford University Press, 1987.
- . BLOCK, N. “On a Confusion About the Function of Consciousness.” *Behavioral and Brain Sciences*, 18, pp. 227–47, 1995.
- . BURGE, T. “Belief De Re.” *The Journal of Philosophy*, 74 (6), pp. 338-362, 1977.
- . BURGE, T. *The Origins of Objectivity*. Oxford: Oxford University Press, 2010.
- . CAMPBELL, J. “Sense, Reference and Selective Attention.” *Proceedings of the Aristotelian Society*, 71, pp. 55–98, 1997.
- . CAMPBELL, J. *Reference and Consciousness*. Oxford: Oxford University Press, 2002.
- . CAVANAGH, P., ALVAREZ, G.A. "Tracking multiple targets with multifocal attention." *Trends in Cognitive Science*, 9(7), pp. 349-354, 2005.
- . DE CARVALHO, F.N. *Demonstrative Thought: A Pragmatic View*. Berlin, Boston: De Gruyter, 2016.

---

<sup>17</sup> For an elaboration of this point see De Carvalho & Newen, 2019.



- . DE CARVALHO, F.N., NEWEN, A. “A Role for the Prefrontal Cortex in Supporting Singular Demonstrative Reference.” *Journal of Consciousness Studies*, 26 (11-12), pp. 133-156, 2019.
- . DEDRICK, D., TRICK, L. *Computation, Cognition, and Pylyshyn*. Cambridge, MA: MIT Press, 2009.
- . DRIVER, J, et al. “Segmentation, Attention and Phenomenal Visual Objects.” *Cognition*, 80 (1-2), pp. 61–95, 2001.
- . EVANS, G. *The Varieties of Reference*. Oxford: Oxford University Press, 1982.
- . FODOR, J.A. *The Language of Thought*. Cambridge, MA: Harvard University Press, 1975.
- . HATFIELD. G. *Perception and Cognition: Essays in the Philosophy of Psychology*. Oxford: Oxford University Press, 2009.
- . KAHNEMAN, D., et al. “The Reviewing of Object Files: Object-specific Integration of Information.” *Cognitive Psychology*, 24 (2), pp. 175–219, 1992.
- . LAMME, V. “Why Visual Attention and Awareness Are Different.” *Trends in Cognitive Science*, 7(1), pp. 12–18, 2003.
- . LEVINE, J. “Demonstrative Thought.” *Mind and Language*, 25 (2), pp. 169–195, 2010.
- . PEACOCKE, C. “Demonstrative Thought and Psychological Explanation.” *Synthese*, 49(2), pp. 187-217, 1981.
- . POSNER, M., et al. “Attention and the detection of signals.” *Journal of Experimental Psychology: General*, 109, pp. 160-174, 1980.
- . PYLYSHYN, Z. “Is Vision Continuous with Cognition? The Case for Cognitive Impenetrability of Visual Perception.” *Behavioral and Brain Sciences*, 22(3), pp. 341–365, 1999.
- . PYLYSHYN, Z. “Visual Indexes, Preconceptual Objects, and Situated Vision.” *Cognition*, 80, pp. 127–158, 2001.
- . PYLYSHYN, Z. *Things and Places: How the Mind Connects With the World*. Cambridge, MA: MIT Press, 2007.
- . QUINE, W.V.O. *From Stimulus to Science*. Cambridge, MA: Harvard University Press, 1995.
- . RAFTOPOULOS, A. “Reference, Perception, and Attention.” *Philosophical Studies*, 144 (3), pp. 339–360, 2009a.

- . RAFTOPOULOS, A. *Cognition and Perception: How Do Psychology and Neural Science Inform Philosophy?* Cambridge, MA: MIT Press, 2009b.
- . RAFTOPOULOS, A., & Müller, V.C. "Nonconceptual Demonstrative Reference." *Philosophy and Phenomenological Research*, 72(2), pp. 251–285, 2006.
- . RENSINK, R.A. "The Dynamic Representation of Scenes." *Visual Cognition*, 7 (1-3), pp. 17–42, 2000.
- . SCHOLL, B. "What Have We Learned About Attention from Multiple Object Tracking (and Vice Versa)?" In D. Dedrick and L. Trick (eds.) (2009), pp. 49–78.
- . SMITH, A. D. *The Problem of Perception*. Cambridge: Harvard University Press, 2002.
- . STAZICKER, J. "Attention, Visual Consciousness and Indeterminacy." *Mind & Language*, 26(2), pp. 156-184, 2011.
- . STRAWSON, P.F. *Individuals: An Essay in Descriptive Metaphysics*. London: Routledge, 1959.
- . TREISMAN, A., GELADE, G. "A Feature-integration Theory of Attention." *Cognitive Psychology*, 12 (1), pp. 97–136, 1980.
- . YANTIS, S. "Objects, Attention and Perceptual Experience." In: R.D. Wright (ed.) (1998), pp. 187–214.
- . YANTIS, S., JONIDES, J. "Abrupt visual onsets and selective attention: evidence from visual search." *Journal of Experimental Psychology: Human Perception and Performance*, 10(5), pp. 601-21, 1984.
- . YANTIS, S., & HILLSTROM, A.P. "Stimulus-driven Attentional Capture: Evidence from Equiluminant Visual Objects." *Journal of Experimental Psychology. Human Perception and Performance*, 20 (1), pp. 95–107, 1994.
- . WRIGHT, R.D. *Visual Attention*. Oxford: Oxford University Press, 1998.
- . WU, W. "What Is Conscious Attention?" *Philosophy and Phenomenological Research*, 82(1), pp. 93-120, 2011.