An efficient coding approach to the debate on Grounded Cognition*

Abel Wajnerman Paz

Facultad de Filosofía y Letras

Universidad de Buenos Aires

awajnerman@filo.uba.ar

Abstract: The debate between the amodal and the grounded views of cognition seems to be stuck. Their only substantial disagreement is about the vehicle or format of concepts. Amodal theorists reject the grounded claim that concepts are couched in the same modality-specific format as representations in sensory systems. The problem is that there is no clear characterization of (modal or amodal) format or its neural correlate. In order to make the disagreement empirically meaningful and move forward in the discussion we need a neurocognitive criterion for representational format. I argue that efficient coding models in computational neuroscience can be used to characterize modal codes: These are codes which satisfy special informational demands imposed by sensory tasks. Additionally, I examine recent studies on neural coding and argue that although they do not provide conclusive evidence for either the grounded or the amodal views, they can be used to determine what predictions these approaches can make and what experimental and theoretical developments would be required to settle the debate.

1. Introduction

It seems that the ongoing debate between amodal and grounded views of cognition has reached a deadlock. The available data is insufficient to favor one of the views and there is no agreement regarding what kind of additional evidence is required to move forward in the discussion. Grounded cognition affirms that the main structures underlying our cognitive abilities (crucially, mental concepts) are not amodal but rather components of modality-specific systems for perception, action and emotion. In turn, the classic theories of cognition assume that representations are amodal data structures. The grounded view has been

* Forthcoming in Synthese

supported by cognitive, behavioral, linguistic and neurocognitive evidence. However, it has been argued that none of these sources of evidence is sufficient to rule out relevant versions of the amodal approach (e.g., Machery 2007, Mahon 2015). Both amodal and grounded theorists agree now that this evidence shows, at best, that conceptual processing involves some degree of interaction between modality-independent and modality-specific systems, which is compatible with alternative (either amodal or pluralists) proposals (e.g., Dove 2009, Leshinskaya and Caramazza 2016).

The only claim which amodal theorists would reject is about the vehicle or format of concepts. They deny that concepts are couched in the same modality-specific code as representations in sensory systems. The problem is that both grounded and amodal theorists agree that there is no clear characterization of modal or amodal format or the kind of neurocognitive evidence required to determine format. Therefore, in order to make their disagreement empirically meaningful we need a neurocognitive criterion for representational format.

In this paper, I will suggest that the theoretical framework of efficient coding explanations can be employed both to draw the distinction between modal and amodal format and to determine what kind of evidence could favor the amodal or the grounded view. In section 2, I develop a dilemma for grounded theorists presented by Mahon (2015), which implies that they endorse a view that is either compatible with the amodal approach or incompatible with neurocognitive evidence. I argue that a grounded view that emphasizes format can provide a way out of the dilemma because it constitutes a middle ground between weak and strong versions of the proposal. As a view about format, the grounded approach can be both substantial and empirically plausible.

In section 3, I address the question of what neurocognitive evidence can be used to determine format. Theorists from both sides of the debate agree that location (i.e., being inside a perceptual system), interaction (with a perceptual system) and (modality-specific) content are independent (and therefore not reliable indicators) of modal format. This suggests that we need a criterion directly based on the intrinsic structure of representations. I propose that modal format can be identified by determining the relation between intrinsic structure and perceptual processing and claim that efficient coding models can be useful precisely to understand this relation. The main idea is that a code is modal if and only if it satisfies some special informational and/or computational demand imposed by a perceptual task. In turn, a code is amodal if and only if it does not satisfy any demand of this kind but rather *only* general demands imposed by any information processing task or a specific demand imposed by a non-sensory task.

In section 4, I present the notion of neural coding and characterize different kinds of neural codes. I will also explain how these codes are related to the structures that are relevant for the debate on grounded cognition. In section 5, I examine recent studies on efficient neural coding and argue that although they do not provide conclusive evidence for either the grounded or the amodal views, they can be used to determine which predictions they can make and which experimental and theoretical developments would be required to move forward in the debate. That is, the efficient coding framework can provide a neurocognitive characterization of the disagreement between these approaches.

I suggest that grounded theorists can make use of the hypothesis that sparse codes in sensory systems evolved or were developed to represent the statistical structure of their input signals. Based on this idea, they can predict that codes in downstream areas dedicated to higher cognition preserve sparseness. On the other side, amodal theorists can seek support in the proposal that neural structures dedicated to higher cognition developed localist or selective codes in order to satisfy demands imposed by specific learning tasks. In this case, the predictions that must be tested are related to the implementation of specific computational operations and task parameters presupposed by the relevant efficient coding model.

2. The grounded view as a proposal about format

As mentioned above, the grounded view has been supported by different sources of evidence. Originally, diverse behavioral and linguistic experiments have been proposed to show that cognition requires perceptual representations (e.g., Wu 1995, Barsalou et al. 1999, Solomon and Barsalou 2001, Neininger and Pulvermüller 2003, Hauk et al. 2004; Pulvermüller 2005). Here I will focus on findings at the neural level(s). This is because although (as we will see briefly) the grounded view is underdetermined by current evidence from neuroscience, it is in this domain where we can find additional data to move forward in the discussion.

The neurocognitive experiments offered to motivate the grounded view show that modality-specific regions dedicated to vision and action are activated during concept related tasks that do not overtly require sensorimotor processing. For instance, early brain-imaging studies showed that retrieving the name of the typical color of an object elicited activity near a region in occipital cortex which is activated during color perception (Martin et al., 1995). Similarly, the word ‘kick’ causes the activation of the motor representation of the leg (Hauk et al., 2004) and saying ‘hammer’ to a picture of a hammer activates information about how to manipulate the object (Chao and Martin, 2000). Some of the examples more recently proposed include activation of motor-processing regions by reading about motion (Deen & McCarthy, 2010, Saygin, McCullough, Alac, & Emmorey, 2010) and activating somatosensory cortex by viewing pictures of graspable objects (Smith & Goodale, 2015). These findings have an obvious implication: Sensory-motor and conceptual processing are closely related. However, this is not denied at all by amodal theorists. If there is a disagreement between grounded and disembodied views, it is not about whether sensory-motor systems play a role in conceptual processing but rather about which role they play.

According to grounded theorists, this evidence implies that the resources required to perform the relevant tasks are (in whole or in part) modality-specific and not amodal. The expression ‘in whole or in part’ suggests two different readings of the grounded approach. Mahon (2015) distinguishes between strong and weak grounding (or embodiment). Strong grounding is the claim that modality-specific representations are necessary and sufficient to support conceptual processing. In turn, weak grounding is the thesis that concept-related tasks require both modality-specific and modality-independent structures (Binder and Desai, 2011; Hauk and Tschentscher, 2013; Kiefer and Pulvermüller, 2011; Lambon-Ralph, 2014; Meteyard et al. 2012; Zwaan, 2014).

Given that, according to the strong view, the only representations underlying conceptual processing are modality-specific, this approach implies that concepts *are* modality-specific representations. Mahon (2015) points out that it is not clear whether this approach has ever been adopted. Leshinskaya and Caramazza (2016) argue that it was defended by Barsalou et al. (2003). Also, Meteyard et al. (2012) claim that Pulvermüller (2001) and Barsalou (1999) often suggest a strong reading of their views. Barsalou (2016) rejects this interpretation and presents strong grounding as one of the usual caricatures of the grounded approach. Although perhaps Barsalou’s proposal can be sometimes read as consistent with amodal symbols, there are other grounded views which are more openly strong, such as the ones presented by Gallese & Lakoff (2005), Allport (1985) and Glenberg & Gallese (2012). Also, Prinz (2002) is very explicit in his endorsement of a strong view (at least under one of the possible readings of his proposal[1]). He is very exhaustive in showing how all conceptual tasks can be performed by using only perceptual representations. Anyway, it is not important whether strong grounding was actually held. Even if it was a mere distortion of actual proposals, it implies a dilemma for grounded theorists.

Mahon (2015) argues that on the one side, strong embodiment is a thesis that is clearly different from any disembodied proposal. However, there is strong neuro-imaging evidence against it. On the other side, current evidence is compatible with weak embodiment. The problem is that this thesis is not different from disembodied cognition. This means that, if the weak and strong views are the only available versions of the grounded approach then there is no version which is both different from an amodal approach and consistent with neurocognitive evidence. In a similar vein, Leshinskaya and Caramazza (2016) argue that the idea that conceptual representations can be reduced to sensory-motor structures is untenable on the face of evidence of neuropsychological damage that give rise to distinct disorders, some modality-specific and some modality-general. They also claim that the weaker versions of the grounded approach, which claim that conceptual processing requires either the interaction between amodal and modality-specific representations or the activation of structures that overlap with sensory-motor regions, are compatible with a disembodied approach.

Evidence for some degree of independence between conceptual and perceptual processing was originally proposed by Warrington (1975). She described subjects that were able to match pictures of different viewpoints of an object but could not name nor describe those objects, nor match names to descriptions. These findings led Warrington to conclude that perceptual classification and semantic classification depend on systems which can be differentially impaired. This conclusion was later supported by, for instance, Hodges (1992)

[1] Prinz (2002) claims that concepts are copies of perceptual representations. This idea can have two different interpretations that depend on two alternative characterizations of the notion of 'copy' (Prinz 2002 pp. 108-109). According to one reading, copies involve instructions to reactivate representations stored in perceptual systems in the absence of external stimuli. Under this view, concepts are not strictly speaking copies because they are identified with the reactivated perceptual representations (i.e. they are not the instructions). This proposal is a version of strong grounding. Under a second reading, copies are duplicates of perceptual representations stored outside perceptual systems. This is a version of the form of weak grounding that, I will argue, is more plausible.

and Hillis & Caramazza (1995). Also, Caramazza & Mahon (2006) show that there is consensus regarding the fact that category-specific deficits (involving categories such as ‘animals’, ‘fruit/vegetables’, and ‘artifacts’) are not associated with modality-specific deficits. Likewise, McCaffrey & Machery (2012) indicate that patients with semantic dementia, a form of frontotemporal dementia, lose all feature knowledge (visual, auditory, tactile, etc.) for specific concepts, while knowledge about those same kinds of features for related concepts remains intact. This kind of evidence suggests that, against strong grounding, we cannot identify conceptual representations with modality-specific representations. Although perceptual representations may play a relevant role in conceptual processing, conceptual representations constitute a different set of structures.

However, we saw that many grounded theorists accept that perceptual and conceptual representations are different. Weak grounding is now the dominant view. The problem is that if grounded cognition is merely the thesis that conceptual processing involves some kind of interaction between conceptual and perceptual representations, it would not be inconsistent with an amodal approach. This is because disembodied cognition is essentially a thesis about format. Machery is quite explicit on this point:

> “In the present context, neo-empiricist and amodal theories of concepts disagree, not about the origins of concepts (whether concepts are innate or acquired) or about the content of concepts (what concepts represent), but rather about the representational code, or format, of concepts […] Neo-empiricists hold that concepts and perceptual or motor representations have the same code; amodal theories deny it.” (Machery 2016, p. 1091)

In Mahon’s terms: “if the issue of whether or not concepts are represented in a modality-specific format has been resolved, then there is no longer any debate about embodiment” (Mahon 2015, p. 423). Weak grounding is consistent with this view because interaction between conceptual and perceptual representations has no implication regarding conceptual format. As Leshinskaya and Caramazza (2016) point out, interaction is not a threat to the amodal view unless one accepts “the false assumption that by interacting with sensory-motor representations, concepts have a different nature than if they did not do so” (p. 993). As Mahon (2015) puts it, the format of a concept and the format of the representations with which it is connected in the input and output systems are independent empirical questions. Weak grounding is a substantial claim about how conceptual processing is performed and which neural systems it requires. However, it is not a thesis about the nature of concepts, which seems to be the core of the amodal view.

The obvious way out of this dilemma is to find a version of the grounded approach that is strong enough to be different from the amodal approach and permissive enough to be compatible with the mentioned evidence. Barsalou (2016) describes a proposal which fits this description. Unsurprisingly, his characterization depends on a distinction between modal and amodal format. He claims that modal representations have an analogical format. A representation is analogical if and only if it is isomorphic with the represented object, property or event. For instance, a mercury thermometer is an analogue representation of temperature because the column of mercury expands and contracts as a function of the temperature. Likewise, a map is an analogue representation of an area, because greater distances on the map represent greater distances in the represented area (Machery 2016). On the contrary, an amodal representation does not exhibit this kind of isomorphism but rather has a structure that is arbitrarily related to what it represents.

Barsalou identifies the amodal approach with the claim that conceptual representations (A1) are arbitrarily related to their corresponding categories (that is, they are not analogical) and (A2) are sufficient to perform the computations underlying conceptual processing. In contrast, his grounded view is that conceptual representations (G1) are analogical and (G2) are not sufficient for conceptual processing, i.e., the activation of modality-specific systems is also required. This is a version of weak grounding because (G2) implies that concepts are not identical with the representations stored in perceptual systems. As we saw, this is consistent with the evidence against strong grounding but it is also insufficient to distinguish between the amodal and grounded views. Condition (G1), which specifies the format of conceptual representations, can be used to draw this distinction. This condition implies that although conceptual and perceptual representations are numerically different (that is, they are different sets of representations located in different systems) they are representations of the same kind: They share a common format. Concepts are not representations stored in perceptual systems, but they have a perceptual format. Some version of G1 is one of the main tenets of paradigmatic versions of the grounded view (see Machery

2007). Also, given that amodal format is a constitutive aspect of amodal approaches, (G1) is incompatible with them.

It is confusing that Barsalou (2016) dismisses format as an optional or lateral aspect of his view. He affirms that “the focus of grounded researchers has typically been on the neural reuse of these regions for conceptual processing, *not* on their representational format.” (p. 1130, his emphasis). The problem with denying that a view on conceptual format is a constitutive aspect of the grounded approach is not only that grounded theorists would be trapped in Mahon’s dilemma, but also that they would be shying away from a discussion they were not losing. Barsalou (2016) is aware that the idea of an amodal format is a “black hole in conceptual space.” He affirms that amodal theorists never provide concrete descriptions of what amodal concepts are or how they are supposed to work. In contrast, we saw that he does provide a characterization of the distinction between modal and amodal format. This is why endorsing (some version of) (G1) seems the best option for a grounded theorist.

Once we managed to formulate a plausible and substantial version of the grounded approach, the next step is to figure out what kind of evidence we need in order to test it. It is clear that the neurocognitive evidence mentioned in this section will not do the trick. Activation of structures inside perceptual systems is irrelevant to the question of whether representations *outside* perceptual systems have a modal format. In the next section, I will evaluate what kind of neurocognitive evidence is relevant to determine representational format and outline a proposal, which will be further developed in sections 4 and 5.

3. What counts as evidence for modal or amodal format?

It is acknowledged by theorists from both sides of the debate that finding evidence for representational format at the neural level is a very challenging task. The main problem is that many accessible variables of neural processing are independent of format. In the first place, we saw that Leshinskaya and Caramazza (2016) affirmed that evidence about the interaction between sensory and non-sensory systems has no implications regarding the format of structures in non-perceptual systems.

In the second place, neural location is not indicative of format either. Martin (2016) affirms that the regions where we store information about specific object associated properties are located within (or overlap with) perceptual and action systems. Lenshiskaya and Caramazza (2016) claim that this is not a problem because neural location does not determine the format of a representation. Furthermore, Martin (2016) himself admits that location and format are unrelated: “Even in the earliest lowest-level regions of the visual-processing stream, the format could be depictive on the way up, and propositional on the way back down.” (Martin 2016, p. 984). A more substantial reason for rejecting this locationist criterion is that it would make grounded cognition conceptually impossible. We saw that grounded theorists have to claim that concepts are modal structures located *outside* perceptual systems. It is interesting to notice that this idea backfires on amodal theorists. The mere fact that concepts are structures located outside perceptual systems (something that the grounded theorist has to concede) does not imply that these have an amodal format.

In the third place, grounded theorists propose that conceptual representations retain lower-level information (e.g., information coding a particular shape, color, or body action) represented in modality-specific systems (e.g., Martin 2007, 2009, 2016, Binder 2016). However, even grounded theorists admit that format and content are independent (e.g., Binder 2016). That is, one can represent the same information in a perceptual or amodal format. In this vein, Mahon affirms that “if information is about a given modality (e.g., information about the visual properties of objects, or about object manipulation), then it may or may not also be assumed to be represented in a modality-specific format.” (Mahon 2015, p. 924).

Given that we cannot determine format indirectly by using either location, interaction or content, it seems that we need a criterion directly based on the intrinsic structure of a modal representation. This would require to identify aspects of neural structure that are not only accessible but that can also be plausibly considered *modal*. A proposal of this kind is Barsalou’s idea that perceptual representations are analogical. This seems a promising criterion because isomorphism is a relatively accessible property of some sensory neural structures. Specifically, topographic maps are neural structures implemented by different sensory modalities which exhibit some degree of isomorphism with what they represent (Swindale 2008). However, this idea has faced different objections related to the conceptual problem of what modal format is. Prinz (2002) mentions Wittgenstein’s (1919) idea that there is an abstract isomorphism between true sentences and the world. If there was a language of thought, utterly removed from perception, it could have this property (Prinz 2002, p. 112). In a similar vein, Machery (2016) points out that analogical representations manipulated by analogue computers are not perceptual in any clear sense. They may not be employed in any perceptual task.

Instead of trying to identify a different *specific* defining property of modal intrinsic structure I propose that modal format depends on the relation between structure and perceptual information processing. Specifically, we can say that a representation is modal if and only if its intrinsic structure is determined in some way by special demands imposed by a perceptual processing task. This idea was suggested by grounded theorists. For instance, Barsalou (2016) claims that different perceptual systems are likely to employ different operations and representations which are best suited to perform their different information processing tasks. He affirms that:

> "Regions of a modality-specific pathway, such as the ventral stream, have probably evolved to perform specific types of computations on relevant information. As a consequence, the representations and processes in these regions are likely to reflect such constraints. As a further consequence, representations and processes are likely to differ across modalities (e.g., vision vs. audition vs. action vs. affect vs. taste)." (Barsalou 2016, p. 1131).

The fact that a perceptual system employs a given set of representations is determined in some way by special constraints imposed by a sensory task. Likewise, Prinz endorsed Kosslyn's idea that different kinds of representations may be better suited for different sensory tasks and therefore different systems may employ different representations (Kossylyn 1980). He affirms that "the representations ideally suited for deriving information from light are not ideal for deriving information from sound." (Prinz 2002, p. 117). In this paper, I will endorse a weaker version of this idea. What makes a format perceptual is being determined by an informational demand that is unique to perceptual processing (i.e., that is not imposed by any other cognitive task). This is consistent with the possibility that different sensory systems share a common perceptual demand and, as a result, they share a common perceptual code. As we will see in section 5, this possibility is supported by current evidence.

An apparent problem with this proposal is that, as we saw in the previous section, the only viable version of the grounded view claims that systems which are not dedicated to perception but rather to higher cognitive functions employ representations that have a perceptual format. If a representation is employed by a non-perceptual system, then it seems unlikely that its format will be determined by the demands of a perceptual task (that is, a task for which it may not be recruited). However, this is not problematic if we notice that many grounded proposals claim that non-perceptual systems *preserve* modal format. Concepts are modal because transmission of information from the modalities to non-sensory systems is not an arbitrary transduction into a different code but rather a process that retains (or reproduces) sensory coding. If this is the case, non-sensory systems would employ a format initially or previously developed to perform perceptual tasks.

Barsalou (2016) reviews different ways in which concepts could represent abstract properties without departing from a modal coding strategy. For instance, conjunctive representations proposed by Binder (2016) perform multi-modal data-compression rather than arbitrary transduction. Also, distilled abstraction is a process described by Jamrozik et al. (2016) through which conceptual representations can acquire abstract content without losing modal format. Likewise, Prinz claims that concepts are copies of perceptual representations. As I mentioned above (see footnote 1), one of the readings of this idea is that representations produced in perceptual systems are duplicated in other systems. Prinz also points out that this sort of causal/temporal priority of perceptual representations is a hallmark of classic British empiricism (Prinz 2002, p. 108). This means that non-perceptual systems could replicate a format which was initially developed to satisfy modality-specific demands. Representations in a non-perceptual system could be tokens of a format type that was previously (in evolutionary or developmental time) instantiated in a perceptual system to satisfy modality-specific demands.

Besides the fact that this idea has been proposed by grounded theorists, there are reasons to affirm that it is also biologically plausible. It has been pointed out that a feature common to engineering and biological systems is the reuse of available circuit elements. In the design of an electronic device the same circuit elements, such as transistors and logic gates, are reused many times. Biological systems display the same principle. For instance, Alon and colleagues have modeled key wiring patterns, called ‘network motifs’ that are repeatedly implemented in a biological network (Alon, 2007a, 2007b; Milo et al., 2002; Shen-Orr et al. 2002). Within the domain of cognitive neuroscience, it has been recently noticed that different neural systems reuse standard computational modules, called 'Canonical Neural Computations' (CNCs). These modules apply the same operations for different information processing tasks in different brain areas (Reynolds 2009). For instance, divisive normalization (one of the most well-studied CNCs) is reused in different systems to maximize sensitivity, achieve invariance with respect to some stimulus dimensions or optimize stimuli discrimination (Carandini & Heeger 2012). Different neural systems often reuse components to perform different information processing tasks.

The question now is: How can we determine that a given code is the result of modality-specific demands? Is this an accessible aspect of neural processing? I will argue that his idea can be computationally and neurally tested by using the efficient coding approach in cognitive neuroscience. This constitutes a line of investigation which aims to explain precisely the presence of a given coding regime in a neural system by its ability to optimize specific parameters of information processing. I will say that the code implemented by a system is modal if and only if it satisfies special demands imposed by a perceptual task. On the contrary, a code is amodal if and only if it does not satisfy any perceptual demand but rather *only* general informational demands (that is, demands imposed by any information processing task) or special informational demands imposed by a non-sensory task. In the next two sections, I will clarify this criterion by employing the efficient coding approach and consider what implications it has regarding the present debate.

4. An efficient coding approach to modal and amodal format

In the previous section, I proposed to identify modal and amodal representations by assessing which demands determine the implementation of a given format. However, it is not obvious what format should look like at the neural level.

Martin (2016) offers a clear description of this problem:

> "The problem, however, is that we do not know how to determine the format of a representation. What we do know is that at the biological level of description, mental representations are in the format of the neural code. No one knows what that is, and no one knows how it maps onto the cognitive descriptions of representational formats (i.e., amodal, propositional, depictive, iconic, and the like), nor even if those descriptions are appropriate for such mapping. What is missing from this debate is agreed-upon procedures for

determining the format of a representation. Until then, the format question will remain moot. It has no practical significance." (Martin 2016, p. 984).

Following Martin, we can break down the format problem into three sub-problems. In the first place, we must determine what neural code is. Then, we must figure out how to map neural code onto the kinds of representations that are relevant for the debate. Lastly, we must determine what kind of evidence can be used to determine that a given modal or amodal code is implemented. In what follows, I will suggest how these three questions can be addressed.

Regarding the first question, although neural coding is a developing field of research with many unanswered questions, we have at least some ideas about what neural codes are or can be and there are more or less standard procedures to characterize coding in a given neural population. A common approach is population analysis, which aims to determine coding regime by discovering patterns in the combined activity of different neurons. A pattern of spike trains produced by a neural population is often recorded by using electrode arrays or (more recently) fluorescence microscopy and interpreted with decoding algorithms or information theory. Once its informational properties are understood, the pattern is systematically explored to determine which features of the spike trains carry the relevant information, i.e., which code the population implements (Quian Quiroga & Panzeri, 2009, 2013).

There are two aspects of population activity that are used to characterize neural codes and which will be relevant for the present discussion. The first one is density. The density of a code is determined by the average fraction of neurons of a given population that are (more or less) simultaneously triggered to represent a given condition. Density can vary from close to 0 to about 1/2. When density is higher than 1/2 it can be decreased without loss of information by replacing each active neuron with an inactive one, and vice versa. Codes with lowest density are local codes, in which each condition is represented by only one active neuron. The highest density is given by holographic coding. These codes represent each condition by the combination of activities of 1/2 neurons of a population. Sparse codes are a compromise between dense and local codes. Under this regime, multiple (but few) neurons participate in the representation of each condition (Földiák 2002, Földiák & Endres 2008, Földiák 2013).

A second aspect is the distinction between selective and distributed codes. In distributed codes different subsets of cells in a population encode different conditions but, crucially, these subsets share cells. In contrast, in selective[2] coding a population represents different conditions by using non-overlapping subsets. This implies that single cells are interpretable only if the code is selective. Under this coding regime each unit or neuron responds highly selectively to a given condition (e.g., specific faces, objects, or words) such that the output of that unit can be interpreted unambiguously. Units of this sort are sometimes called localist units or ‘‘grandmother cells”. On the other hand, in distributed representations, given that each unit responds to a wide range of different conditions, individual units cannot be interpreted without knowing the state of other units in the network (Thorpe 1995, p. 550, Page 2000, Bowers 2016). Figure 1 presents an illustrative comparison between a selective and a distributed code.

John	1	1	0	0
Paul	0	1	1	0
George	0	0	1	1
Ringo	1	0	0	1

(a)

John	1	0	0	0
Paul	0	1	0	0
George	0	0	1	0
Ringo	0	0	0	1

(b)

Figure 1. Adapted from Page (2000). Four names represented by (a) distributed coding and (b) selective coding. Each 1 and 0 represents an active or inactive cell, respectively. Rows represent the population response to each stimulus.

The dense/sparse and the distributed/selective distinctions are related. A distributed code requires some degree of density (i.e., more than one cell has to respond to each stimuli)

[2] Selective coding is often also called ‘localist’. Given that low density codes have a similar name, I will avoid using this expression.

in order to form overlapping sets of neurons. Also, populations in visual areas are known to be both sparse and selective (Lehky et al 2005 and see section 5). However, these notions are relatively independent. A given population can implement a sparse and distributed code if overlapping subsets of cells fire in response to different stimuli and each of those subsets are constituted by very few cells. In turn, a selective and dense representation could be implemented by a population in which many redundant neurons respond selectively to the same input (Bowers 2016)[3].

The second problem mentioned by Martin is the mapping problem. The described codes appear at the level of neural populations. How do we know that structures at this level can be identified with the representations that are relevant for the present debate? There are two aspects to this question which can be addressed separately. First, we must determine whether structures at the cellular or population level can be identified with the kinds of representations posited by grounded and amodal theorists.

As it happens, this can be done for many relevant representations. For instance, some grounded theorists have proposed that conceptual processing employs distributed representations. These structures are not distributed in the sense specified above but rather in the sense that they have no central point of convergence and involve dispersion of perceptual and motor features across modal association cortices (e.g., Allport 1985; Gage & Hickok 2005). One way to understand the implementation of distributed representations is by using the Hebbian notion of a cell assembly. This is a set of strongly connected cells formed when neurons in different cortical areas are frequently active at the same time (Palm et al. 2014, Holtmaat & Caroni 2016). This characterization of distributed representations, which has been proposed by Pulvermüller (e.g., Pulvermüller 1999, 2013) situates them at the same cellular level as neural codes. Another kind of grounded structure is cross-modal conjunctive representation (CCR) (e.g., Binder 2016). These are structures that respond preferentially to a particular combination of inputs. It has been pointed out that CCRs can be instantiated in single cells having a local coding regime (Quiroga, Reddy, Kreiman, Koch, & Fried, 2005) and also in distributed neural ensembles or networks (O'Reilly & Busby, 2001). A third example are amodal structures that exhibit cross-modal responses to a given class of stimuli.

---

[3] I thank an anonymous reviewer for stressing the relevance of distinguishing between these two parameters of neural coding.

In contrast with distributed representations (and consistent with the so called amodal hubs), these structures constitute central points of convergence for different modal inputs. The most striking findings supporting the existence of these amodal concepts come from the research on number representation (e.g., Dehaene et al. 1998, Piazza et al. 2006). These representations can respond to a given cardinality independently of stimulus modality. Interestingly, cross-modal number representations have been found at the cellular level. Andreas Nieder has studied ‘number neurons’, which have a maximum discharge rate to a preferred numerosity across different modalities (Nieder 2016). In a similar study, Vergara et al. (2016) showed that individual neurons in the pre-supplementary motor area (pre-SMA) exhibit cross-modal responses to specific frequencies. These examples imply that at least some of the relevant structures proposed both by modal and amodal theorists can be found at the same level as neural codes.

After determining that a given neural code constitutes the format of, for instance, a conjunctive representation, we still have to assess whether this code is modal or amodal. The second aspect of the second question is to determine whether neural codes can be mapped onto known modal or amodal formats discussed in the literature. We do not know which of the mentioned neural codes could implement, for example, propositional (e.g., Pylyshyn, 2003) or depictive (e.g., Kosslyn, Thompson, & Ganis, 2006) format. However, this is not necessary. Once we endorse the criterion proposed in the previous section, we only need to figure out whether a given code satisfies special demands required by a perceptual task in order to determine if it is modal or not. This leads us to the third problem: How can we determine that a code is modal in this sense? There is a tradition of investigation in cognitive neuroscience which studies how neural coding contributes to optimize the transmission of information. Efficient coding models have been recently introduced in the philosophical literature by Mazviita Chirimuuta in order to discuss the nature of computational explanation in cognitive neuroscience (Chirimuuta 2014, 2017). My aim is to apply this view to the present discussion on representational format.

5. Implications from the efficient coding approach

The efficient coding approach is a tradition that can be traced back to classical theoretical work by Horace Barlow. He hypothesized that optimizing information transmission is a driving force in the evolution and/or development of neural codes (Barlow 1959). Informational and computational demands can be used to explain why the brain implements a given coding strategy. Briefly, an efficient coding explanation requires first specifying which demands must be satisfied in order to perform efficiently a given information processing task. Then, we build models of different hypothetical coding strategies which the brain could plausibly implement and which satisfy to some degree the specified demands. If the most efficient strategy lines up with the actual one, we have an explanation of why it was implemented (see Chirimuuta 2014). In this section I will suggests how grounded and amodal theorists can use efficient coding models in order to make predictions about neural coding.

5.1. Amodal codes

Some theoretical results and experimental data provided within this framework seem to favor the amodal view. At the end of section 3 I mentioned that a code is amodal if and only if it is not the result of demands which are unique to sensory tasks. This is compatible with two possibilities. An amodal code could be determined either by specific demands of a non-sensory task or by general demands shared by all neural information processing tasks[4]. Some early studies suggest that the development of neural codes is driven by general requirements, such as maximizing representational capacity (i.e., the number of conditions or the amount of information that a system can represent) and minimizing metabolic cost. Considering that the brain is one of the metabolically most active organs of the body (Sokoloff 1989), Levy and Baxter (1996) proposed that neural coding must result from an optimal compromise between energy and informational efficiency. In this vein, Attwell and Laughlin (2001) offer a characterization of the optimality of distributed neural coding on the basis of a detailed energy budget for signaling in the grey matter[5].

In order to determine the impact of different coding strategies on neural energy consumption, the authors consider a system that must represent 100 different sensory or motor conditions. One possibility is using a selective code in which each of the 100 conditions is represented by a single active cell. This would require to recruit 100 cells. As

[4] I thank an anonymous reviewer for suggesting the possibility of dedicated but non-sensory codes

[5] Attwell & Laughlin (2001) call this code 'sparse'. As we will see below, this expression is often used to refer to very different coding regimes. However, as I mentioned, it is relevant to distinguish between sparse and distributed coding.

soon as we begin to depart from this selective coding regime towards a distributed one, an increase in energy efficiency is patent. If each condition is represented by the simultaneous firing of 2 cells (with the others not firing) and different representations can share cells, only 15 neurons are needed to represent 100 conditions. This is given by the equation (which I will call the "capacity/code/components equation" or "3C equation") that relates representational capacity or number of conditions represented (*R*) with the number of cells or components of the system or population (*n*) and number of active cells that represent each condition (*np*):

$$R: n!/[(n-np)!\,(np)!]$$

In our case, 3C implies that 15!/(13! 2!) = 105. Similarly, if each condition is represented by 3 firing cells and different representations can overlap, 3C implies that only 10 cells are needed to represent 100 conditions (given that 10!/(7! 3!) = 120), which is a further improvement of energy efficiency[6]. These results imply that the implementation of distributed coding is not motivated by special perceptual demands (i.e., unique to sensory tasks) but rather only by general demands. Any information processing task can be improved by minimizing metabolic cost without reducing representational capacity. This is why distributed codes can be considered amodal.

The main problem with this approach is that current evidence suggests that codes in neural structures underlying higher cognition are selective and not distributed. We saw that in distributed codes each cell is used to represent different conditions and therefore these conditions cannot be decoded by recording a single cell but rather all (or part of) the population has to be taken into account (Thorpe 1995, p. 550). Several studies imply that relevant information can be extracted from single units in key areas associated with higher cognition.

Although the role of the prefrontal cortex is difficult to characterize, it has been often associated with functions that play a crucial role in higher cognition and intelligent behavior.

---

[6] This is an abstract and schematic characterization of energy optimization. See Attwell and Laughlin (2001) to understand how the biochemical variables that constitute the metabolic cost of neural signaling are involved.

It is considered that its basic function is the representation and execution of new forms of organized goal-directed action. This main function depends on the so-called executive functions of the prefrontal cortex: planning, decision-making, working memory (the memory required for the performance of acts in the short term), preparatory set (the priming of sensory and motor neural structures for the performance of an act contingent on a prior event), and inhibitory control (Fuster 2015, ch. 1).

What do we know about the prefrontal representations that subserve these functions? A relevant model affirms that prefrontal populations implement what has been called 'adaptive coding'. The central idea is that throughout much of prefrontal cortex the response properties of single neurons are highly adaptable. Any given cell has the potential to change its tuning properties. In a particular task context, many cells become tuned to code information that is specifically relevant to this task (e.g., Duncan 2001, Woolgar et al. 2011, Stokes et al. 2013). A second and related aspect of prefrontal cells is mixed selectivity (MS). The firing patterns of individual MS neurons are modulated by each of multiple task dimensions. Prefrontal cells often code some combination of sensory stimuli, task rule and motor response which are relevant for a task being performed (see Fusi et al. 2016).

These aspects of prefrontal representations (adaptive coding and mixed selectivity) are not related to neural coding but rather only to neural information. The facts that prefrontal cells can change the information that they carry and that they can represent many task-relevant features at the same time do not inform us *per se* about which response properties carry this information, i.e., which code they implement. However, physiological studies aimed at characterizing these properties of prefrontal information processing do reveal something about prefrontal codes. In these studies, the relevant information can be very often decoded from single cells (e.g., Freedman et al. 2001, Rigotti et al. 2013). As we saw, this is possible only if a selective code is employed. These findings suggest that prefrontal codes are selective and therefore they are not determined solely by the general demands of minimizing metabolic cost and maximizing informational capacity. An additional demand may preclude the implementation of a distributed code.

It is relevant to mention that the same point applies to another neural structure which is also crucial for higher cognition. The hippocampus is often considered the site of different forms of declarative memory (e.g., semantic and episodic memory) and also contributes to spatial navigation. It is known that information about different domains can be decoded from individual hippocampal cells (Eichenbaum 2016). The classic study by O'Keefe & Dostrovsky (1971) showed that different neurons in the hippocampus of a rat constitute a spatial map in which each neuron responded solely or maximally when it was situated in a particular part of a platform facing in a particular direction. Recent evidence shows that there are also hippocampal 'time cells', each of which fires in sequence when an animal is at a particular moment in a temporally structured experience (Manns et al. 2007, Pastalkova et al 2008, Gill et al. 2011, McDonald et al. 2011. See Eichenbaum 2014 for a recent review). Also, in a task where choice performance is guided by odor cues and not their spatial locations, hippocampal individual cells were selective for specific odors (Muzzio et al., 2009)

Of course, the fact that non-sensory areas implement selective coding does not imply that they use modal codes. Perhaps the additional demand that motivates using a selective (instead of a distributed) code is non-sensory. It has been argued that these codes are determined by the second kind of amodal demand I mentioned (i.e. a demand imposed by a specific but non-sensory task). For instance, Marr (1971) suggested that the non-overlapping representations of selective coding are useful for fast learning tasks performed by episodic memory. The main idea is that when memory representations overlap, new learned memories will often interfere with old memories. More recently, Bowers et al. (2014) argued that learning tasks performed by short-term memory (STM) in the cortex require selective coding for co-activating multiple representations at the same time. The basic claim is that when we represent more than one condition by activating at the same time the activity patterns that represent each condition, distributed codes will often generate ambiguities. We often cannot decode the different conditions when the patterns are blended together (Figure 2). This is the so-called superposition catastrophe (Von der Malsburg, 1986). Bowers et al. (2014) showed that a recurrent Parallel Distributed Processing (PDP) model trained to co-activate and recall multiple words at the same time succeeded by learning highly selective letter and word codes. They claim that this could explain the selective coding of neurons in cortex given that the cortex supports STM in various domains (Cowan, 2001).

Pete	1	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0
Roger	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1
Charlie	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0
Brian	0	1	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0
Keith	0	1	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0
Mick	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0
Ringo	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0
George	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	1	0
Paul	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1
John	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0
Paul + Roger	0	1	1	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	1	1
Paul + Brian	0	1	1	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	1	1

Figure 2. Adapted from Bowers et al. (2014). Each 1 and 0 represents an active or inactive cell, respectively. Rows represent the population response to each stimulus. The population implements a distributed code. The two bottom rows represent conjunctions of conditions by simultaneously activating the components that represent each condition. The resulting patterns are ambiguous.

If this hypothesis was correct, selective codes in non-perceptual areas would be the result of a demand imposed by a non-sensory task and therefore could be considered amodal. However, the mentioned evidence is far from conclusive[7]. A first problem is that PDP models are not intended to be biologically realistic. Plaut and McClelland (2010) have pointed out that, traditionally, the primary goal of the PDP approach is accounting for behavioral data. These models are aimed at reproducing human performance in cognitive tasks as it occurs in real time and how performance changes over the course of normal and abnormal development and in adulthood. Neural verisimilitude is not a priority. Moreover, there are specific aspects of the PDP architecture that seem to undermine its biological plausibility, such as backpropagation. Bowers himself has criticized PDP models on the basis that theorists in the PDP tradition often claim that single hidden units in PDP networks are uninterpretable, which

[7] I am indebted to an anonymous reviewer for pointing out the following shortcomings of Bower's proposal. I will suggest below that these are not irresoluble problems but only difficulties that call for further theoretical and experimental developments.

is inconsistent with neurophysiological data from cortical recordings (Bowers 2009, 2010). Therefore, it seems that more work has to be done in order to present a biologically plausible characterization of the computations that relate learning task demands with selective coding. Lately, biological plausibility has become a more central issue for theorists working in the PDP tradition. It has been pointed out that new methodologies used in cognitive neuroscience (such as multivariate pattern analysis and representational similarity analysis) capture insights derived from PDP models. Also, observations from neuroscience are now being used to constrain the architecture of simulated neural network models (Rogers & McClelland 2014).

A second problem is that, under some conditions, distributed codes can avoid the superposition catastrophe. Critically, in both of the two simulations proposed by Bowers et al. (2014) the models successfully recalled lists of three words relying on few localist codes. His model implies that the number of local representations increases only when the list (i.e., the number of conditions that are simultaneously represented) is longer. This finding suggests that distributed codes have some limited capacity to overcome the superposition catastrophe. In this vein, Botvinick and Plaut (2006) proposed a recurrent PDP model of immediate serial recall that succeeded in avoiding the superposition catastrophe with distributed codes and could support short-term memory. Of course, distributed solutions only work under certain conditions (see also Bowers et al. 2016). However, in order to claim that selective codes in episodic and short-term memory result from this constraint one should offer a precise characterization of these conditions and determine whether these are met by the relevant learning tasks.

To summarize, I described two lines of investigation that can be used to claim that higher cognitive areas use amodal codes. The argument based on Bowers et al. (2014) seems more plausible than the one based on Atwell and Laughlin (2001) because only the former is consistent with neurophysiological data from prefrontal and hippocampal recordings. However, the evidence for amodal selective codes is inconclusive. In order to further explore his proposal there are at least two questions that should be answered. In the first place, it must be determined whether the studied neural structures exhibit relevant similarities with PDP models or, alternatively, whether biologically plausible PDP models also produce selective codes during the relevant learning tasks. In the second place, it should be ascertained whether the specific conditions met by the relevant learning tasks (number of conditions represented, number of components, list length, etc.) are those that make it impossible to provide a distributed solution to the superposition problem. An alternative possibility is that selective coding is not required at all by memory learning tasks but is rather the result of specific demands imposed upstream by sensory processing. In the next section I will explore a line of investigation that seeks to characterize the codes underlying efficient perceptual processing.

5.2. Modal codes

A long-standing hypothesis affirms that the visual system implements sparse coding in order to satisfy special perceptual demands. Classic theoretical work suggests that response properties of neurons in the early visual system are the result of properties of the organism's visual environment. The images received by the retina when viewing the natural world have a relatively regular statistical structure, which arises from the contiguous structure of objects and surfaces in the environment (Olshausen & Field 2004). Attneave (1954) and Barlow (1961) proposed that visual neurons are adapted (through evolutionary and developmental processes) to represent these statistical properties.

Barlow (1961) hypothesized that early sensory neurons are able to remove statistical redundancy in the sensory input by representing only the signals that occur most frequently in the natural environment. Many theorists have subsequently claimed that the sparse representations that capture this statistical structure have the same receptive field properties as simple-cells in primary visual cortex (e.g, Field 1987, Olshausen & Field 1996, Bell & Sejnowski 1997, van Hateren & van der Schaaf 1998, van Hateren & Ruderman 1998 and Hyvarinen & Hoyer 2000). Some of the studies advanced by these authors show that after being trained in detecting natural images, sets of model cells only respond to properties that typically occur in those images. Therefore, these sets can represent any given image by using a small number of active units (i.e. by using a sparse code). Crucially, these studies revealed that the receptive fields of trained units resemble those of visual cortical cells. Both are spatially localized, oriented, and bandpass (i.e., selective to spatial structure at a particular scale). This suggests that visual systems actually implement this coding strategy. More recently, it has been suggested that sparse coding may reflect a general principle of sensory processing across sensory modalities. This strategy has been studied not only in vision, but also in audition (DeWeese et al. 2003), and olfaction (Perez-Orive et al. 2002; Stopfer 2007).

Although the similarity between trained artificial units and cortical cells supports the hypothesis that sensory systems implement a sparse code, different attempts have been made to provide more direct evidence. Ideally, the best way to determine the implementation of a sparse code is by recording (often with electrode arrays) simultaneously from the different units in a population and assessing the average number of units that respond to each image in a relevant stimulus set. This measure is what is known as 'population sparseness'. Given that the technical limitations of electrode arrays make it difficult to record simultaneously from all of the units in large populations, population sparseness is often measured by using a different property known as lifetime sparseness. This is a measure of the responses of individual neurons to many stimuli from a given set. A cell with high lifetime sparseness responds to very few stimuli from a given set whereas a population with high population sparseness is one in which very few cells are activated in response to each stimulus (Willmore & Tolhurst 2001).

Measuring lifetime sparseness is supposed to be an adequate way to determine population sparseness because both characterize the 'peakedness' of a neural response. This alleged connection between lifetime and population sparseness is useful for our debate because lifetime sparseness is closely related to selectivity. We saw that a selective representation is only activated by one item from a stimulus set (it is not part of the representation of different stimuli). This means that it has very high lifetime sparseness (see Bowers 2010). We know that responses in sensory systems are selective. For instance, neurophysiological studies showed that as one progresses along the ventral stream of the visual system, individual neurons become selective for increasingly complex stimuli, from spots of light and oriented bars in early visual areas to parts of borders in intermediate areas and complex objects and faces in higher visual areas (Mély & Serre 2017)**.** We also know that higher cognitive structures (such as the PFC or the hippocampus) have selective units. If selectivity is a reliable indicator of high population sparseness then it would seem that sparse codes are implemented in perceptual systems and preserved in higher cognitive areas. This idea could be used to support the grounded view.

Nevertheless, Willmore et al. (2011) showed that lifetime and population sparseness are only related under very special conditions (namely, only if neural responses are independent and identically distributed) and that these conditions are not met by visual populations. Lifetime sparseness does not imply population sparseness. I already mentioned that selectivity and (population) sparseness are independent: a population in which all cells are selective for the same stimulus will have high lifetime sparseness (they only fire in the presence of that stimulus) but low population sparseness (they all fire in the presence of the stimulus). Therefore, lifetime sparseness cannot be used to show either that sensory systems implement sparse codes or that sparse codes are preserved in areas dedicated to higher cognition.

Fortunately, the relatively recent development of optical techniques has provided the necessary tools to overcome the technical limitations of electrode arrays. Transcranial two-photon microscopy (TPM) is a very powerful fluorescence imaging technique that can be applied to living tissue up to about 1 mm in depth and at a spatial resolution of less than 1 μm (e.g., Ustione & Piston 2011). More importantly, two-photon imaging can reach a very high spatial density, recording the activity of hundreds of neurons in small volumes of the visual cortex. This means that TPM can be used to directly measure population sparseness. In a recent study Froudarakis et. al (2014) used TPM to record the activity of nearly all of the neurons in small volumes of the visual cortex in mice and directly detect high population sparseness. This kind of study provides very solid evidence for sparse coding in sensory areas. Additionally, they found that population sparseness was higher for natural images than for control movies in which relevant high-order statistical information was removed. This suggests that sparse codes are not only implemented by visual areas but also correlated with the statistical properties of natural visual signals.

These findings suggest two ways in which the grounded approach can be explored. In the first place, a grounded hypothesis can be formulated in terms of sparseness. Sparse codes are modal because they are the result of constraints imposed by sensory processing. Therefore, grounded theorists can predict that codes employed by neural structures dedicated to higher cognition preserve sparseness. To the best of my knowledge, sparseness has not yet been studied in higher cognitive areas. However, we saw this is now technically possible. Fluorescence techniques could be used in order to measure population sparseness, for instance, in prefrontal and hippocampal structures. A second possibility is to evaluate the idea that selectivity contributes in some way to efficient sensory processing. As I mentioned, we know that neural responses in sensory and non-sensory systems are selective. However, the parameter that is known to optimize sensory processing (population sparseness) is independent of selectivity (or lifetime sparseness) and the contribution of the later to perception is still not well understood (Willmore et al. 2011). What we learnt from Attwell and Laughlin's argument is that selective codes are very expensive. Therefore, it seems plausible that there is some constraint forcing the implementation of a selective strategy. The question is whether this constraint is unique to sensory processing. The grounded view could be supported by showing that selective codes satisfy some sensory demand.

There is a final question that is relevant for understanding how the efficient coding framework shapes the debate on grounded cognition: What is the relation between the amodal and grounded predictions about neural coding? I mentioned that the grounded view would be vindicated if it is found that sparse codes are preserved in areas dedicated to higher cognition and that selectivity is the result of demands imposed by sensory tasks. In turn, the amodal view would be supported by evidence that codes for higher cognition are not sparse and that their selectivity is the result of non-sensory demands. However, we saw that selectivity and sparseness are not incompatible. Therefore, grounded and amodal predictions could be *both* confirmed. It could be the case that central codes preserve sparseness and that their selectivity is the result of non-sensory demands. If this was the case, the evidence would support other positions that already have an important place in the debate.

Hybrid and pluralist views claim that higher cognition requires different kinds of structures (Dove 2009, Lambon Ralph et al. 2010, Reilly et al. 2016, Shallice & Cooper 2013, Zwaan 2014). These approaches are usually classified as hybrid or pluralist, depending on the relation they postulate between these structures. Hybrid theories claim that individual concepts are a combination of different types of structures whereas pluralist theories claim that each category is represented by different structures and each of them constitutes an independent concept (Machery 2009, Weiskopf 2009). An efficient coding version of a hybrid view could be that individual neural representations in higher cognitive areas are both sparse and selective (and sparseness is the result of a sensory demand whereas selectivity is not). That is, they have a modal and an amodal aspect. In turn, an efficient coding pluralist view could be that some of these structures are selective but not sparse and others are sparse but not selective. These considerations show that the possible predictions about format that can be made by using the efficient coding framework are rich enough to characterize the different positions within the debate.

5. Conclusion

I have argued that a substantial and empirically plausible version of the grounded approach affirms that concepts are structures with modal format situated outside perceptual systems. Elaborating on a suggestion advanced by grounded theorists, I proposed that modal codes are those that were developed or adapted to satisfy specific informational demands imposed by sensory tasks. I showed that the efficient coding framework can be used to identify modal or amodal codes at the neural level, therefore making the debate on format empirically meaningful.

A first approximation to the efficient coding literature reveals that current models and physiological data do not provide conclusive support for or against the grounded view. However, they can be used to articulate predictions that grounded and amodal approaches can make regarding neural coding and specify the experimental and theoretical developments that must be advanced in order to confirm these predictions. Regarding the amodal view, a plausible hypothesis is that non-sensory neural structures dedicated to episodic and short-term memory developed selective codes in order to satisfy demands imposed by learning tasks. I suggested that this proposal is appealing because selective coding seems ubiquitous in brain areas dedicated to higher cognition. However, further experimental research is required in order to show that the proposed PDP models are actually implemented by the relevant neural structures and to determine whether the tasks performed by these structures meet the conditions that make a distributed solution to the superposition problem impossible.

Regarding the grounded view, a promising proposal is that sparse codes in sensory systems evolved to respond to the statistical structure of their input signals. I mentioned that there is solid evidence both for the fact that sensory systems implement sparse codes and that sparse codes optimize the representation of sensory information. In this case, what is required is to determine whether downstream areas dedicated to higher cognition preserve sparseness. A different approach would be to determine whether the property that we know to be shared by perceptual and non-perceptual codes (selectivity) plays a significant role in efficient sensory processing.

The fact that the efficient coding framework can be used to formulate empirically testable hypothesis about modal and amodal format motivates the introduction of the literature on neural coding into the discussion on grounded cognition. Efficient coding constitutes a prolific framework for thinking about how format is implemented at the neural level and how it is related to information processing.

Acknowledgements

I would like to thank the members of the Research Group on Cognition, Language and Perception (CLP) from Buenos Aires (Liza Skidelsky, Mariela Destéfano, Sergio Barberis, Sabrina Haimovici, Nicolás Serrano, Fernanda Velázquez Coccia and Cristial Stábile) for many early discussions of this material. I am also grateful to two anonymous reviewers for very helpful suggestions on crucial aspects of the manuscript. Finally, I am indebted to Julieta Picasso Cazón for ongoing support.

References

Alon, U. (2007a). *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Boca Raton, FL: Chapman & Hall

Alon, U. (2007b). Network Motifs: Theory and Experimental Approaches. *Nature Reviews Genetics* 8:450–61.

Allport, D. A. (1985). Distributed memory, modular subsystems and dysphasia. In: Newman, S. K., Epstein, R., editors. *Current perspectives in dysphasia*. Churchill Livingstone: New York.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychol. Rev.* 61:183–93.

Attwell, D. and Laughlin, S. B. (2001). An Energy Budget for Signaling in the Grey Matter of the Brain. *Journal of Cerebral Blood Flow and Metabolism* 21:1133–1145

Barlow, H. B. (1959). *Symposium on the Mechanization of Thought Processes*. H. M. Stationary, London, No. 10, pp. 535-539

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. In W. A. Rosenblith. (ed.) *Sensory Communication*. Cambridge, MA: MIT Press. pp. 217–34.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–609.

Barsalou, L.W., Simmons, W. K., Barbey, A. K.,& Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, 7, 84–91.

Barsalou, L.W. (2016). On staying grounded and avoiding Quixotic dead ends. *Psychonomic Bulletin & Review*, 23, 1122-1142

Binder, J. R., Desai, R. H. (2001). The neurobiology of semantic memory. *Trends in Cognitive Science* 15, 527–536.

Binder, J. R. (2016). In defense of abstract conceptual representations. *Psychonomic Bulletin & Review* 23, 1096-1108

Bell, A. J. & Sejnowski, T. J. (1997). The 'independent components' of natural scenes are edge filters. *Vision Res* 37, 3327-3338.

Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review* 113, 201–233.

Bowers, J. S. (2009). On the biological plausibility of grandmother cells: Implications for neural network theories in psychology and neuroscience. *Psychological Review* 116, 220–251.

Bowers, J. S. (2010). More on grandmother cells and the biological implausibility of PDP models of cognition: A reply to Plaut and McClelland (2010) and Quian Quiroga and Kreiman (2010). *Psychological Review* 117, 300–306.

Bowers, J. S., Vankov, I. I., Damian, M. F., & Davis, C. J. (2014). Neural networks learn highly selective representations in order to overcome the superposition catastrophe. *Psychological Review* 121(2), 248-261.

Bowers, J. S., Vankov, I. I., Damian, M. F., & Davis, C. J. (2016). Why do some neurons in cortex respond to information in a selective manner? Insights from artificial neural networks. *Cognition* 148, 47–63.

Caramazza, A., & Mahon, B. Z. (2006). The organization of conceptual knowledge in the brain: The future's past and some future directions. *Cognitive Neuropsychology* 23, 13–38.

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(1), 51

Chao, L.L. & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *Neuroimage* 12, 478–84.

Chirimuuta, M. (2014). Minimal models and canonical neural computations: The distinctness of computational explanation in neuroscience. *Synthese* 191, 127–153.

Chirimuuta, M. (2017). Explanation in Computational Neuroscience: Causal and Non-causal. *British Journal for the Philosophy of Science*. doi: 10.1093/bjps/axw034

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences* 24, 87–114.

Deen, B., & McCarthy, G. (2010). Reading about the actions of others: Biological motion imagery and action congruency influence brain activity. *Neuropsychologia* 48, 1607–1615.

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences* 21, 355–361.

DeWeese, M. R., Wehr, M., Zador, A. M. (2003). Binary spiking in auditory cortex. *J Neurosci* 23, 7940–7949.

Dove, G. (2009). Beyond perceptual symbols: A call for representational pluralism. *Cognition* 110 (3), 412-431.

Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nat Rev Neurosci.* 2 (11), 820-829.

Eichenbaum H. (2014). Time cells in the hippocampus: a new dimension for mapping memories. *Nat Rev Neurosci.* 15 (11), 732-44.

Eichenbaum, H. (2016). Still searching for the engram. *Learn Behav.* 44 (3), 209-2.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *J Opt Soc Am A* 4, 2379–2394.

Földiák, P. (2002). Sparse coding in the primate cortex, in M. A. Arbib (ed.) *The Handbook of Brain Theory and Neural Networks*, Second Edition, MIT Press. Pp. 1064-1068.

Földiák, P. & Endres, D. (2008). Sparse coding. *Scholarpedia*, 3 (1), 2984.

Földiák, P. (2013). Sparse and explicit neural coding. En *Principles of Neural Coding*, R. Quian Quiroga y S. Panzeri (eds.). Boca Raton, FL: CRC press, pp 379-389.

Freedman, D. J., Riesenhuber, M., Poggio, T. & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* 291, 312–316.

Froudarakis E., Berens P., Ecker A. S., Cotton R. J., Sinz F. H., Yatsenko D., Saggau P., Bethge M., Tolias A.S. (2014). Population code in mouse V1 facilitates readout of natural scenes through increased sparseness. *Nat. Neurosci*. 7, 851–857.

Fusi, S., Miller, E. K. & Rigotti, M. (2016) Why neurons mix: high dimensionality for higher cognition. *Curr. Opin. Neurobiol*. 37, 66–74.

Fuster, J. M. (2015). *The Prefrontal Cortex*. Fifth Edition, San Diego: Academic Press.

Gage, N., & Hickok, G. (2005). Multiregional cell assemblies, temporal binding and the representation of conceptual knowledge in cortex: A modern theory by a “classical” neurologist, Carl Wernicke. *Cortex* 41 (6), 823–832.

Gallese, V., Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology* 22, 455–479.

Gill, P. R., Mizumori, S. J. & Smith, D. M. (2011). Hippocampal episode fields develop with learning. *Hippocampus* 21, 1240–1249.

Glenberg, A. M., Gallese, V. (2012) Action-based language: A theory of language acquisition, comprehension, and production. *Cortex* 48: 905–922.

van Hateren, J. H., van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc R Soc Lond B Biol Sci* 265, 359-366.

van Hateren, J. H., Ruderman, D. L. (1998). Independent component analysis of natural image sequences yields spatiotemporal filters similar to simple cells in primary visual cortex. *Proc R Soc Lond B Biol Sci* 265, 2315-2320.

Hauk, O., Johnsrude, I., Pulvermüller, F. (2004). Somatotopic Representation Of Action Words In Human Motor And Premotor Cortex. *Neuron* 41:301–307.

Hauk, O. & Tschentscher, N. (2013). The body of evidence: what can neuroscience tell us about embodied semantics? *Frontiers in Psychology* 4, 1–14.

Hillis, A. E., & Caramazza, A. (1995). Cognitive and neural mechanisms underlying visual and semantic processing: Implications from optic aphasia. *Journal of Cognitive Neuroscience* 7(4), 457–478.

Holtmaat, A., & Caroni, P. (2016). Functional and structural underpinnings of neuronal assembly formation in learning. *Nature Neuroscience* 19, 1553–1562.

Hyvarinen, A., Hoyer, P. O. (2000). Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. *Neural Comput* 12, 1705-1720.

Jamrozik, A., McQuire, M., Cardillo, E.R., & Chatterjee, A. (2016). Metaphor: Bridging embodiment to abstraction. *Psychonomic Bulletin & Review* 23, 1080-1089.

Kiefer, M. & Pulvermüller, F. (2011) Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex* 48, 805–825.

Kosslyn, S.M. (1980). *Image and Mind*. Cambridge, MA: Harvard University Press.

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. New York, NY: Oxford University Press.

Lambon Ralph, M., Sage, K., Jones, R. & Mayberry, E. (2010). Coherent concepts are computed in the anterior temporal lobes. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 2717-2722.

Lambon Ralph, M. A. (2013). Neurocognitive insights on conceptual knowledge and its breakdown. *Philosophical transactions of the royal society* 369, 20120392.

Lehky S. R., Sejnowski, T. J., Desimone, R. (2005). Selectivity and sparseness in the responses of striate complex cells. *Vision Res* 45, 57-73.

Leshinskaya, A., & Caramazza, A. (2016). For a cognitive neuroscience of concepts: Moving beyond the grounding issue. *Psychonomic Bulletin & Review* 23, 991-1001.

Levy, W.B. & Baxter, R.A. (1996). Energy-efficient neural codes. *Neural Computation* 8, 531–543.

Machery, E. (2007). Concept empiricism: A methodological critique. *Cognition* 104 (1), 19-46.

Machery, E. (2009), *Doing without Concepts*, Oxford, Oxford University Press.

Machery, E. (2016). The amodal brain and the offloading hypothesis. *Psychonomic Bulletin & Review* 23 (4), 1090-5.

Mahon, B. Z. (2015). What is embodied about cognition? *Language, Cognition and Neuroscience* 30, 420–429.

Manns, J. R., Howard, M. & Eichenbaum, H. (2007) Gradual changes in hippocampal activity support remembering the order of events. *Neuron* 56, 530–540.

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology* 58, 25–45.

Martin, A. (2009). Circuits in mind: The neural foundations for object concepts. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences*, 4th ed., Cambridge, MA: MIT Press. pp. 1031–1045.

Martin, A. (2016). GRAPES—Grounding representations in action, perception, and emotion systems: How object properties and categories are represented in the human brain. *Psychonomic Bulletin & Review*. 23: 979-90

Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C.L., Ungerleider, L. G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. *Science* 270 , 102–5.

Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 262 (841), 23–81.

McCaffrey, J., & Machery, E. (2012). Philosophical issues about concepts. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3 (2), 265-279.

McDonald, C. J., Lepage, K. Q., Eden, U. T. & Eichenbaum, H. (2011). Hippocampal "time cells" bridge the gap in memory for discontiguous events. *Neuron* 71, 737–749 (2011).

Mély, D. A. & Serre, T. (2017). Towards a Theory of Computation in the Visual Cortex. In Zhao Q. (ed.) *Computational and Cognitive Neuroscience of Vision. Cognitive Science and Technology*. Springer, Singapore, pp. 59-84.

Meteyard, L., Rodriguez-Cuadrado, S., Bahrami, B., Vigliocco, G. (2012) Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex* 48, 788–804.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science* 298, 824–827.

Muzzio, I. A., Levita, L., Kulkarni, J., Monaco, J., Kentros, C., Stead, M., Abbott, L. F., Kandel, E. R. (2009). Attention Enhances the Retrieval and Stability of Visuospatial and

Olfactory Representations in the Dorsal Hippocampus. *PLoS Biol* 7(6): e1000140. doi:10.1371/journal.pbio.1000140.

Neininger, B., and Pulvermüller, F. (2003). Word-category specific deficits after lesions in the right hemisphere. *Neuropsychologia* 41, 53–70.

Nieder A. (2016). The neuronal code for number. *Nature Reviews Neuroscience* 17,366-82.

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 34, 171–175.

Olshausen, B. A., Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381:607–9.

Olshausen, B. A., & Field, D. J. (2004). Sparse coding of sensory inputs. *Current Opinion in Neurobiology* 14(4), 481-487.

O'Reilly, R. C., & Busby, R. S. (2001). Generalizable relational binding from coarse-coded distributed representations. *Advances in Neural Information Processing Systems* 14, 75–82.

Page, M. P. A. (2000). Connectionist modeling in psychology; A localist manifesto. *Behavioral and Brain Sciences* 23, 443-467.

Palm, G., Knoblauch, A., Hauser, F., Schüz, A. (2014). Cell assemblies in the cerebral cortex. *Biological cybernetics* 108, 559-572.

Pastalkova, E., Itskov, V., Amarasingham, A. & Buzsáki, G. (2008). Internally generated cell assembly sequences in the rat hippocampus. *Science* 321, 1322–1327.

Perez-Orive, J., Mazor, O., Turner, G. C., Cassenaer, S., Wilson, R. I., Laurent, G. (2002). Oscillations and sparsening of odor representations in the mushroom body. *Science* 297, 359–365.

Plaut, D. C., & McClelland, J. L. (2010). Locating object knowledge in the brain: Comment on Bowers's (2009) attempt to revive the grandmother cell hypothesis. *Psychological Review*, 117, 284–288.

Piazza, M., Mechelli, A., Price, C. J., & Butterworth, B. (2006). Exact and approximate judgements of visual and auditory numerosity: An fMRI study. *Brain Research* 1106, 177–88.

Prinz, J. (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge, MA: MIT Press.

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, 22, 253–336.

Pulvermüller, F. (2001). Brain reflections of words and their meaning. *Trends in Cognitive Sciences* 5, 517-524.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nat. Rev. Neurosci*. 6, 576–582.

Pulvermüller, F. (2013). Semantic embodiment, disembodiment or misembodiment? In search of meaning in modules and neuron circuits. *Brain and Languag*, 127, 86–103.

Pylyshyn, Z. (2003). Return of the mental image: Are there really pictures in the brain? *Trends in Cognitive Sciences*, 7, 113–118.

Quian Quiroga R, Panzeri S. (2009). Extracting information from neural populations: Information theory and decoding approaches. *Nat Rev Neurosci* 10:173–185.

Quian Quiroga, R., Panzeri, S. (2013). *Principles of Neural Coding*, CRC press, Boca Raton, FL.

Quiroga, R., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature* 435, 1102–7.

Reilly, J., Peelle, J., Garcia, A. & Crutch, S. (2016). Linking somatic and symbolic representation in semantic memory: the dynamic multilevel reactivation framework. *Psychonomic Bulletin & Review*, 23, 1002-1014.

Reynolds, J. (2009). Canonical Neural Computation: A Summary and a Roadmap. A Workshop at Villa La Pietra, Florence, 17-19 April 2009.

Rigotti, M., Barak, O., Warden, M. R., Wang, X., Daw, N. D., Miller, E. K. & Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature* 497, 585–590.

Rogers, T. T. & McClelland, J. L. (2014). Parallel Distributed Processing at 25: Further Explorations in the Microstructure of Cognition. *Cogn Sci* 38, 1024–1077.

Saygin, A. P., McCullough, S., Alac, M., & Emmorey, K. (2010). Modulation of BOLD response in motion-sensitive lateral temporal cortex by real and fictive motion sentences. *Journal of Cognitive Neuroscience*, 22, 2480–2490.

Shallice, T. & Cooper, R. (2013). Is there a semantic system for abstract words? *Frontiers in Human Neuroscience*, 7, 1-10.

Shen-Orr, S. S., Milo, R., Mangan, S., & Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. *Nature Genetics* 31, 64–68.

Smith, F.W.,& Goodale,M. A. (2015). Decoding visual object categories in early somatosensory cortex. *Cerebral Cortex* 25, 1020–1031

Sokoloff, L. (1989). Circulation and energy metabolism of the brain. In G. J. Siegel,. W. Agranoff, R. W. Albers, and P. B. Molinoff, eds., *Basic Neurochemistry: Molecular, Cellular, and Medical Aspects,* 4th ed., pp. 565-590. Raven Press, New York.

Solomon, K. O., & Barsalou, L. W. (2001). Representing properties locally. *Cognitive Psychology*, 43, 129–169.

Stokes, M.G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., and Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron* 78, 364–375.

Stopfer M. (2007). Olfactory processing: massive convergence onto sparse codes. *Curr Biol* 17: R363–R364.

Swindale, N. V. (2008). Visual map. *Scholarpedia*, 3(6):4607.

Thorpe, S. (1995). Localized versus distributed representations. In Arbib, M. A. (ed.) *The handbook of brain theory and neural networks*. MIT Press.

Ustione, A. & Piston, D. W. (2011): A simple introduction to multiphoton microscopy. *J Microsc.* 243, 221-6.

Vergara, J., Rivera, N., Rossi-Pool, R., and Romo, R. (2016). A neural parametric code for storing information of more than one sensory modality in working memory. *Neuron* 89, 54–62.

Von der Malsburg, C. (1986). Am I thinking assemblies? In G. Palm & A. Aertsen (Eds.), *Brain theory*. Berlin: Springer.

Warrington, E. K. (1975). The Selective impairment of semantic memory. *The Quarterly Journal of Experimental Psychology* 27, 635–657.

Weiskopf, D. (2009). The plurality of concepts, *Synthese* 169, pp. 145-173.

Willmore, B. & Tolhurst, D. J. (2001). Characterizing the sparseness of neural codes. *Network* 12, 255–270.

Willmore, B. D., Mazer, J. A., and Gallant, J. L. (2011). Sparse coding in striate and extrastriate visual cortex. *J. Neurophysiol*. 105, 2907–2919.

Wittgenstein, L. (1919). *Tractatus Logico-philosophicus*. D. Pears and B. McGuinness, trans. London: Routledge.

Woolgar, A., Hampshire, A., Thompson, R., and Duncan, J. (2011). Adaptive coding of task-relevant information in human frontoparietal cortex. *J. Neurosci.* 31, 14592–14599.

Wu, L. L. (1995). Perceptual representation in conceptual combination. Doctoral dissertation. University of Chicago.

Zwaan, R. (2014) Embodiment and language comprehension: Reframing the discussion. *Trends in Cognitive Sciences* 18, 229–234.