

The big concepts paper: A defence of hybridism

Agustín Vicente and Fernando Martínez Manrique

ABSTRACT

The renewed interest in concepts and their role in psychological theorizing is partially motivated by Machery's claim that concepts are so heterogeneous that they have no explanatory role. Against this, pluralism argues that there is multiplicity of different concepts for any given category, while hybridism argues that a concept is constituted by a rich common representation. This paper aims to advance the understanding of the hybrid view of concepts. First, we examine the main arguments against hybrid concepts, and conclude that, even if not successful, they challenge hybridism to find a robust criterion for concept individuation, and to show an explanatory advantage for hybrid concepts. Then we propose such a criterion of individuation, which we will call *functional stable coactivation*. Finally, we examine the prospects of hybridism to understand what is involved in recent approaches to categorization and meaning-extraction.

1. Introduction: the heterogeneity of conceptual representations

The last years have seen a renewed philosophical interest in the nature of concepts and their role in psychological theorizing. In a widely discussed book, Machery ([2009]) proposed that cognitive science can do without concepts. The main basis for this *eliminativist* conclusion was the multiplicity of representational structures that are posited in psychological theories to account for the array of effects found in the last decades of categorization research. Prototypes, exemplars, theory-like networks, and ideals have been proposed as the best explanation for different sorts of findings (see Murphy [2002]; Kruschke [2005]; or Machery [2009]) for reviews of the literature from different perspectives). This heterogeneity, along with certain ideas about what a natural kind

is and how concepts are to be individuated, led Machery to the conclusion that so-called concepts do not share any interesting property to be regarded as a unified natural kind, so there was no point in trying to investigate their nature (Machery [2005]). A number of authors raised objections to this pessimistic conclusion (e.g., see replies to Machery [2010]). Among those that accept the multiplicity of representations disclosed by concept research there are two main reactions. One is to argue for a *pluralist* stance (Piccinini and Scott [2006]; Weiskopf [2009]); the other is to endorse a *hybridist* view of concepts (e.g., Gonnerman and Weinberg [2010]; Keil [2010]; Scarantino [2010]). The main difference between both approaches has to do with the question whether the different kinds of representations posited to explain categorization results are or are not separate from each other. The pluralist thinks that for any given category C there are typically several different representations in long-term memory – a prototype of C, a set of exemplars of C, and so on. These representations are actually different concepts that are stored separately, operate independently, and are usually accessed in different tasks. The hybridist thinks that the different representations are in fact parts of a common representation, which constitutes the concept for C. They are stored in long-term memory as a functional whole and all of them can be accessed in the same task (at least to some degree, more on this below).

The debate connects with recent considerations about the need to posit different kinds of representations to account for categorization results. Initially the crux of the debate on categorization was whether a single representational kind can accommodate the results better than the alternatives. Yet in the last years it is increasingly acknowledged that there are multiple representations involved in categorization processes. To mention but a few, Ashby and O'Brien ([2005]) review the evidence that different memory systems may be involved in different category learning processes; Kruschke ([2005]) thinks that the full complexity of human category learning cannot be accounted for by a single type of representation; Voorspoels *et al.* ([2011]) endorse this conclusion and, in particular, explore the idea that the human conceptual system includes both exemplar representations and abstracted summary representations. Thus the emphasis has moved to

characterize which elements affect differential processing of those representations, and how they actually interact. For instance, Ashby and O'Brien mark as 'questions for future research' whether the multiple memory systems involved in category learning operate independently or cooperate in some way, and whether there is “a consolidation of learning into some common long-term (...) category representation”. To a large extent, these are questions to address empirically. Yet there is also room for a lot of conceptual work on how to characterize notions such as “operate independently” or “common representation”, which have to do with how best to individuate concepts.

In this paper we want to defend the view that hybrid concepts are preferable to the alternatives in terms of their explanatory power. First, we will examine the arguments put forward against hybrid concepts. We will conclude that, even if they are not final against hybridism, they challenge this position to a) find a robust criterion for hybrid concept individuation, i.e., to tell when two distinct representational structures can be said to be constitutive of the same concept; and b) show an explanatory advantage for hybrid concepts. Second, we will propose such a criterion of individuation, which we will call *functional stable coactivation*. In a nutshell, the idea is that different structures can be regarded as constituting a common representation when they are activated concurrently, in a way that is functionally significant for the task at hand, and in patterns that remain substantially stable along different tasks related to the same category. Third, we will argue that hybridism does not just equal eliminativism and pluralism in explanatory power. We will examine two strands in the literature that may be better explained by such hybrid concepts: on the one hand, recent work on categorization shows that conceptual structure has the form of rich, heterogeneous representations; on the other, work on meaning extraction shows that we can explain linguistic understanding if we postulate that lexical entries either consist in, or give access to, a rich but integrated conceptual structure which may well include a theory, a prototype, an exemplar-based computational mechanism, etc.¹

2. Two challenges for hybrid concepts: individuation and explanation

Arguments against hybrid concepts have been most explicitly formulated by Machery ([2009], [2010]), Machery and Seppälä ([2011]), and Weiskopf ([2009]). Machery proposed two criteria for concept individuation that, he argued, are not actually met by hybrid theories of concepts. Weiskopf's strategy against hybridism is different: he provides two arguments that lead to the conclusion that hybrid concepts are either unmanageable or explanatorily superfluous. So Machery challenges hybridism to provide adequate criteria of individuation for hybrid concepts, while Weiskopf challenges hybridism to tell what explanatory advantage it provides. Let us consider both challenges in turn.

2.1. The coordination criterion

Machery's proposal is that a number of representations count as a single concept when they meet two different criteria, which he calls *connection* and *coordination*. According to the connection criterion, two representations count as parts of the same concept when they are connected in such a way that retrieval of one of them typically facilitates retrieval of the other in different cognitive tasks. For instance, if one part is used in, e.g., categorization, then another can be used in induction. According to the coordination criterion “the parts of a given concept do not produce inconsistent outcomes, for instance, inconsistent categorization judgments” ([2009], p. 64). We will concentrate on coordination because it is the criterion that Machery emphasizes (Machery [2009]; Machery and Seppälä, [2011]). As for connection, there seems to be plenty of evidence that we indeed may categorize something as an X in terms of prototypes and then induce that it has some properties based on an X-theory. For instance, we may categorize an animal that we see near the shore as a bird based on its sharing enough features with the bird-prototype and then induce, on the basis of our knowledge of sea-birds, that it probably eats fish (see Murphy [2002], Ch. 5, for discussion).²

Machery contends that hybridism and his own “heterogeneity hypothesis” are just terminological variants of the same position unless the hybridist commits herself to the existence of some mechanism of coordination between parts. He contends that hybrid concept theories entail predictions about categorization judgments that are not actually fulfilled, i.e., there are many cases of inconsistent categorization judgments that give the lie to hybrid views. Even though the first case he discusses is Malt’s ([1994]) findings concerning people’s categorization of water, the bulk of the cases draw from linguistic intuitions. Machery ([2009], p. 71) writes: “According to hybrid theories of concepts, a sentence ‘x is a P’ should typically not be ambiguous. For the parts of a given concept of P are not supposed to be independent ways of thinking about P. By contrast, according to the Heterogeneity Hypothesis... a sentence ‘x is a P’ will sometimes be ambiguous. Linguistic evidence supports the Heterogeneity Hypothesis”. The evidence he presents is that sentences such as ‘Tina Turner is a grandmother’ or ‘tomatoes are vegetables’ are true under one reading and false under another.

However, Machery’s argument is not convincing. Parts of concepts have to be coordinated in some way to be properly considered as parts of a single concept, but this does not mean that they have to be coordinated in such a way that they do not result in inconsistent judgments. That is to say: the no-inconsistency requirement is stronger than the coordination requirement, and Machery does not provide any specific argument for the no-inconsistency requirement. Indeed, most critics have questioned the principle of coordination (see Gonnerman and Weinberg [2010]; Keil [2010]; Scarantino, [2010]). On the other hand, the inconsistency phenomenon that Machery identifies is a well-known phenomenon that suggests that, in very many occasions, the constituents of utterance-meanings are modulations of concepts or parts of them (we will deal with this in section 4.2). This entails that the test of inconsistency does not reveal that people are applying different concepts. And in particular, it means that inconsistency is in fact to be expected if prototypes, theories, etc., are parts of larger conceptual structures.

Machery and Seppälä ([2011]) argue that their linguistic data support a “polysemy

hypothesis” concerning prototypes and theories. In turn, they hold that this polysemy hypothesis supports their heterogeneity hypothesis. However, it is simply not the case that the polysemy hypothesis supports heterogeneity. The typical explanations of polysemy resort to different *senses* (or aspects, or facets, or parts) of just *one* concept (see Pustejovsky [1995]; Cruse [2004]); for a proper discussion of polysemy, see section 4.2.2). The archetypical example of polysemy is that of ‘book’, which stands for a concept that has two different aspects: the text/content aspect, and the tome/physical object aspect. To be sure, ‘book’ provokes “in a sense, yes, in a sense, no” truth-evaluations. Suppose you have an exemplar of *Madame Bovary* in your hands and someone else says ‘that book is nice’. She may mean that she likes that particular instance of *MB*, that is, that she likes the book as a physical object, or that she likes *MB*, i.e., the content of the book that Flaubert wrote. Suppose further that, as it happens, the book you are holding in your hands is a paperback pocket old edition of *MB* whose cover is full of coffee stains. In that situation, your friend’s words, if they are not ironic, will mean that she likes the text, not that she likes the tome. So what she says is in a sense true and in another sense false. The book is nice in a sense, and is not nice in another (in one and another sense of BOOK). Yet, as we say, while this can be taken to show that ‘book’ is polysemous –and ‘book’ is indeed polysemous-, it is usually held that there is just one concept at play. In sum, the “polysemy hypothesis” points towards hybridism, rather than towards heterogeneity. In other words: if it is conceded that ‘Tina Turner is a grandmother’ is ambiguous because ‘grandmother’ can express either a stored prototype or a stored theory (or definition), the hybridist can point to the ‘book’ case and say “yes, words can express different parts of a single concept on different occasions”.

So there are powerful reasons to think that the coordination criterion is a red herring. Indeed, Machery ([2010]) eventually admits that it may ultimately fail as a criterion. Still, he challenges the hybrid theorist to provide an alternative characterization of the notion of conceptual parthood. We agree that this is a major task to sustain a theoretically cogent notion of hybrid concepts, and we will devote the next section to such an undertaking, but first we want to engage Weiskopf’s distrust

of hybrids, which poses a different kind of challenge.

2.2. Concepts as constituents of thoughts

Weiskopf makes two points against hybrid concepts. The first is that “big” concepts cannot be the elements of thought. The second, that there is a problem concerning the selection of parts of concepts. Let us quote him at length:

“Concepts then might be thought of as single, structured bundles containing exemplars, prototypes, theoretical information, causal information, category ideals, etc. This representational inflation faces at least two problems. First, structures that include all of a person’s ways of representing a category are too large to serve as units of thought. Working memory, where concepts are assembled on-line into thoughts, has limited capacity, which in effect imposes architectural limits on the size of human concepts. Second, there is no reason to think that *all* category information is retrieved and used each time a person tokens a concept. The experimental data surveyed so far tend to show effects attributable to some relatively well-delimited subset of this information, rather than to the entire bundle. The question then becomes how to selectively retrieve *parts* of this large body of information to serve the immediate goals of the subject better. But again, this begins to collapse into the pluralist position” (Weiskopf [2009], p. 169)

The first argument is flawed. The philosophical tradition (and a good part of the psychological tradition as well) has it that concepts are the constituents of thoughts, such that only what goes into a thought can be properly be called ‘a concept’. The problem for large bodies of information to be part of thoughts, according to Weiskopf, is that concepts are assembled into thoughts in working memory, and working memory cannot work with large bodies of information. However, we think

that there are reasons to doubt the assumption that thoughts are constructed in working memory out of concepts. Indeed, if this assumption were accepted, it would undermine *both* hybridism and Weiskopf's own pluralism. Let us explain.

The mainstream view about concepts, which Weiskopf endorses, is that they are more or less stable bodies of information stored in long-term memory (or, stored in memory, period). Against this more or less orthodox view, Barsalou ([1999]) is notable for holding that concepts are temporary constructs of working memory, on-the-fly constructions which lack stability. The mainstream position has it that Barsalou's "concepts" are constructs that draw from information stored in memory, that is, from concepts in the orthodox sense. However, there is one point that seems well-established, namely, that Barsalou-concepts are what he claims them to be, i.e., unstable representations. The evidence he has amassed is convincing enough. Typically, our categorization is sensitive to background factors such that a change in these background factors entails a change in our categorization. So, Weiskopf's invocation of working memory is counterproductive: working memory does not work with the huge representations posited by hybridism, but most probably it does not work with Weiskopf's concepts either. The representations tokened in working memory are typically not the representations stored in long-term memory. Suppose, for instance, that concepts were prototypes: the unstable prototype used in working memory does not need to be –and will typically not be– the alleged stable prototype stored in memory and constitutive of the concept. Rather, it will be an instance of the prototype that will include only contextually relevant features. That is, it can be argued that the constituents of thoughts are usually slimmer than full-fledged concepts, i.e. bodies of information stored in memory, either because they are constructions derived from them³ or because they are parts of full-fledged, stored concepts. If this is a problem, then it is as problematic for hybrid concepts as it is for Weiskopf's plural concepts, given that any subkind of the plurality will suffer from it.

The second problem for "big concepts" comes from Weiskopf's assumption that to posit a hybrid concept is to posit a representation that is "employed in all circumstances when a concept is

entertained and used”. He argues that this poses a dilemma for hybrids: “either they fall short of accounting for some phenomenon or other, or they attempt to account for *everything* and grow to the point of being unmanageable”. In fact, the evidence shows that only a delimited subset of such a big structure is “retrieved and used” in each cognitive task, so Weiskopf contends that “this begins to collapse into the pluralist position”. The idea is that the pluralist individuates concepts in terms of the functions they serve in different cognitive processes. For instance, if certain tasks involving category C demand prototype processing, and certain other tasks demand exemplar processing, then one can conclude that we have two different concepts of C, respectively prototype-based and exemplar-based. Given that the hybrid theory has to allow that only a subpart of the hybrid concept will be used in each occasion, it will have to concede that different subparts count as different functional units, even if they are somehow linked in long-term memory. But then the hybrid explanation will not be different from the pluralist one, and the greater parsimony of the latter will lead us to dispose of the notion of hybrid as an unnecessary add-on.

In short, Weiskopf is posing a challenge for hybrid theory that goes further than Machery's. The problem is not only to find a robust criterion for hybrid individuation –i.e., one that shows when different representations can be seen as forming a single concept– but also to say which explanatory advantage our theory gains in talking about such a heterogeneous structure instead of talking merely about its putatively component parts. We think that there is a kernel of truth in his claim that, if the hybridist wants to develop a position that is relevantly different from pluralism, she has to assume that the whole representation that constitutes the hybrid concept is “employed in all circumstances when a concept is entertained and used”. However, the notion of “*employing* a concept” cannot mean, as we saw, that the whole concept is retrieved in working memory. There is a weaker and more plausible reading, i.e., that the whole conceptual representation has a role to play. The possibility that the rich, heterogeneous, structure plays a role even if it is not retrieved in working memory will become apparent once one pays attention to an important distinction regarding what it takes to access a concept, an idea that we develop in the next section.

3. Individuating hybrids: functional stable coactivation

As we said above, we concur with Machery in regarding concept individuation as a pressing issue for hybridism. To put it roughly, a concept will be something that is constituted by a number of representational elements, some of them possibly shared by other different concepts, and what matters for concept individuation is to explain what keeps those elements together. Let us begin by seeing how Weiskopf proposes to individuate concepts. Weiskopf contends that concepts are individuated “primarily by the content that they represent and by how they represent it; a cat concept *is* that concept because it represents cats in whatever way that it does” ([2009], p. 147). The first part (the content) is a condition of coreferentiality, i.e., non-coreferential concepts count *ipso facto*, as different concepts⁴. The way Weiskopf fleshes out the “how” condition, in turn, is by typifying concepts in terms of the different structures involved. So if you have, say, a prototype, an exemplar, and a theory associated to the same category, you will have three different concepts of the same thing. Yet there is some link between those different concepts: “What *makes* CAT₁, CAT₂, etc. all cat concepts is that they refer to cats. What enables the thinker to *recognize* this fact is that they are connected to one another via identity links that indicate this co-reference relation to the thinker.” ([2009], p. 166). Weiskopf calls *referential chaining* to the mechanism that establishes the links, and the result is that “concepts are stored, linked together, and retrieved by a set of memory processes that are indifferent to the subkind that they belong to”.

We think that the notion of referential chaining is too weak. What the literature on concept development shows is that typical processes of acquisition *integrate* information.⁵ What the hybridist claims is that such an integration process is better understood as the formation of a rich common representational structure that constitutes the concept. What Weiskopf labels CAT₁ and CAT₂ are but different parts or instances, drawn to working memory under the pressure of task demands, of that rich, common concept. The criterion we propose to tell when different representational

structures can be regarded as constituting a common conceptual representation is *functional stable coactivation*. The idea is that representations belong to the same concept when (i) they are activated concurrently by the same pattern of stimulus, even if only some of them are selected for further processing, (ii) what is activated has some functional significance for the task at hand, even if it is not responsible for the main effects, and (iii) the coactivation patterns remain substantially stable in tasks related to the same category. The three conditions in the criterion play different roles, so let us elaborate a little on each of them.

The coactivation condition (i) captures the idea that when a concept is retrieved all its component parts are retrieved at the same time, given a certain time span.⁶ Which time span is actually relevant is something for empirical research to ascertain. On the one hand, spreading activation models where each node counts as a concept employ very low values –for instance, Anderson’s ACT (Anderson [1983]) posits that it takes about 1 millisecond for each link to be activated. On the other hand, if each concept, as we think it will be the case, is itself realized in a network-like structure then a higher value will probably have to be admitted. At any rate, we think it will have to be a value that makes it unlikely that in cases of priming the concept activated by the prime (say, ‘car’) forms part of the concept activated by the target (say, ‘bus’), given the lag between prime and target.

Now, even if all concept parts are activated concurrently, not all of them need to be selected for further processing. This distinction is relevant to tell the pluralist and hybridist positions apart. First, let us note that a number of concept theorists have warned that it is necessary to distinguish questions of representation from questions of access. For instance, Murphy ([2002]) suggests that the evidence found in research on categorization perhaps does not tell much about concept representation but about access to this representation. This means that claims, say, about exemplars explaining better some categorization effect for a certain concept C are just claims that exemplars are accessed, not that the whole representation of C is constituted by exemplars. Thus the actual representation could consist of a hybrid concept combining prototypes and exemplars, with the

different parts accessed for different tasks. However, the pluralist can always argue that this possibility has no explanatory value given that what is relevant to account for experimental effects is only what is accessed. So the fact that there can be more represented than what is actually accessed does not *per se* show that what is represented constitutes a single concept whose parts are those different structures accessed in each occasion.

To address this issue we need to make a distinction in the notion of 'what is accessed'. Let us begin by noting that whenever a certain stimulus is processed a vast amount of representations can receive some activation. It is reasonable to assume that a certain representation counts as 'accessed' when its activation goes above certain threshold. Most of the representations accessed in this way are not subject to further processing and its activation may simply decay. Yet a few of them –which ones depends on factors such as kind of task, attentional focus, contextual properties, and others– will remain active, will possibly be drawn to working memory, and will be selected for further processing. So there are two dimensions in the notion of 'accessed': to be *activated*, and to be *selected*. The coactivation condition merely points out that all the parts of the same concept are activated concurrently, not that all of them are actually selected to play the leading role in a given task. Still, the pluralist might press the point by arguing that only what is selected is relevant for concept identification given that only what is selected is functionally significant. This leads us to the second element of our criterion, i.e., functionality.

The functionality condition (ii) says that in order to count as part of the same concept the coactivated elements must have some functional significance, even if not all them are responsible of the main effects. Typically, main effects are related to what is selected, which receives further processing and hence plays the prominent role in explanation. Thus if a certain categorization effect is explained by appeal to exemplar concepts, this means that the categorization judgment is carried out selecting and processing exemplars. Yet this does not mean that what is only activated is not functionally significant. We think that examples of this can be found in research showing that having more semantic features (Moss *et al.* [2007]) or more senses (Azuma and van Orden [1997]);

Klepousniotou and Baum [2007]) is beneficial for conceptual processing. As we will explain later, a reasonable explanation is that features and senses are parts of the same concept that are coactivated in a way that has a functional effect in which and how the features/senses relevant for the task are selected.

The functionality condition is meant to discard irrelevant information: when retrieving a particular concept it is likely that there will be representations that are activated at the same time yet are ineffectual for the task at hand. The idea behind functionality is that only representational components that make a positive contribution to select the appropriate tokening of the concept count as part of such a concept. For instance, representations that are coactivated but that compete against each other cannot be regarded as part of the same concept. This helps to see why two homonymous meanings cannot be the same concept. In some cases of homonymy resolution (e.g., when both meanings are equally frequent: see Frisson [2009]) both meanings receive activation, which means that two different representations are coactivated. However, it is customary in the homonymy literature to consider that homonymous meanings are separate representations. The reason is that the process of meaning-extraction in these cases is one of competition: accessed representations compete against each other.

The stability condition (iii) deals with the problem that when retrieving a particular concept it may be the case that, depending on the contextual details of the task, other representations that are not constitutive of the concept will be activated at the same time and even influence the result of the cognitive process. The idea behind (iii) is to prevent those more or less variable but functional representational additions from being regarded as components of the concept. One may wonder whether this may allow too much as constitutive of the concept, given that a word may stably give access to different semantically close representations.⁷ Suppose that a word were capable of activating a whole conceptual field: one would not wish to include the whole field as part of the corresponding concept. Notice, however, two things: First, that the fact that other concepts are activated does not necessarily mean that they will be functional for the task at hand. So there can be

conceptually close information that is accessed and yet –just as in the case of coactivated homonymous words– does not play any functional role, so its activation will typically decline rapidly. Secondly, stability is quite a demanding condition: it does not require that all the representations that constitute a concept be activated in all cases, but it requires that they are activated in most cases. Representations that are only semantically related to a concept may be activated in many contexts, but they are not stably activated across contexts. For instance, the “tome” sense of BOOK seems to be stably activated when we hear ‘book’, even when the sense intended is the “information” sense, as in ‘I have my next book in mind’. This is why we could continue this sentence with ‘it will be heavy’. However, representations related to the “tome” aspect of BOOK will be active mostly only when this “tome” sense comes to the fore, and not when the sense that comes to the fore is the “information” sense.⁸

We want to emphasize that the three conditions work together in such a way that they constrain each other. For instance, when one thinks of the vast amount of information that can be functionally relevant for a given task, one can concentrate only on those pieces that go reliably together at different moments and tasks –i.e., that are stably coactivated– in order to individuate concepts. Yet they cannot be regarded as strict necessary and sufficient conditions for concepthood, so the criterion provides no clear-cut boundaries between concepts. Again, when one thinks of the similar problems faced by any account of concepts that regards them as structured entities –e.g., the problem of individuating prototypes– one may suspect that perhaps there are no such neat boundaries to be found. So we doubt that it will be possible to find concept elements that are *always* functionally and stably coactivated, and we are satisfied with elements that *typically* meet the three conditions, however inherently vague the notion of typical may seem.

4. The explanatory power of hybrid concepts

Of course, it is one thing to offer a criterion for individuating hybrid concepts, and a very different

one to find whether there are instances of such concepts that play an explanatory role in psychological models. Let us summarize when a model will lend support to the hybrid view of concepts in the following two claims:

- 1) *Heterogeneity claim*: The model sustains the idea that there is a common long-term representation that does not belong to a single representational kind.
- 2) *Functional integration claim*: This common long-term representation is functionally integrated.

We think that the most pressing problem for the hybridist comes from the element of functional integration. Models that merely appeal to heterogeneous representations could have a more parsimonious explanation in pluralist terms. To use the toy illustration above: suppose that a model appeals to prototype and exemplar representations to account for different categorization effects but that when the prototypes are selected the exemplars are functionally inert, and vice versa. Then the pluralist can say that there is no gain in hypothesizing a hybrid prototype-cum-exemplar concept. The hybridist needs then to show that the representational components are functionally integrated in the sense that all of them have some impact in the outcome of the subsequent processing, even those that do not seem to be subsequently selected and processed themselves. At the end of the day the hybrid view may be facing a sort of trade-off between heterogeneity and functional integration: on the one hand, a concept will look as more genuinely hybrid inasmuch as it is actually composed by structurally heterogeneous subparts but then the problem is how to avoid the pluralist account in which each subpart counts as a different concept; on the other hand, functional integration prevents the pluralist stance but the problem is that the more functional integration one finds, the more difficult it gets to find structurally heterogeneous subparts, and the less hybrid the concept will look.⁹

In this section we examine two different areas of research –categorization and meaning extraction– to assess the prospects of hybridist explanations in them. We begin with categorization given that it is the phenomenon where eliminativism and pluralism chiefly put their focus,

examining some lines of research that have been overlooked by both positions. Yet we think that processes of meaning-extraction in linguistic comprehension provide a more solid hybridist explanation. As Medin and Rips ([2005], pp. 62-3) point out, sometimes both areas of research seem to be playing with different notions of concept. Yet we agree with them that “it cannot be true that the concepts we use in online sentence understanding are unrelated to the concepts we employ in reasoning and categorizing”. We expect that at some point theories will converge and that some hybrid account of concepts will eventually be in a better explanatory position.

4.1. Categorization

Research on categorization shows that it is certainly difficult for a single theory of concepts to account for all the data, and what counts as categorization can be quite diverse (see Malt and Sloman [2007]). Prototype theories may be good at explaining some typicality effects, but they fail at explaining categorization by essence. Theory theories can explain categorization by essence, but they do not have an explanation for typicality effects, and so on and so forth. As we said, pluralism concludes from all this that for a given category there are different kinds of concepts, while eliminativism claims that there are no such things as concepts. According to hybridism, a better explanation lies in regarding a concept as an integrated heterogeneous structure. Depending on the task at hand, and on background factors, one part or another of this complex structure receives more activation and plays the leading functional role. Taken separately, prototypes, theories, etc. may be not concepts, but components of concepts.

Now, one may worry that the differences between hybridism and its competitors are merely a matter of terminological preference, so all that is at stake is how to label the heterogeneity of representational structures that everybody pretty much agrees are involved in categorization. Indeed, whether a model received the label of ‘hybrid’ is not a guarantee that it cannot be interpreted in pluralist terms. Some models of categorization (e.g., Anderson and Betz [2001];

Erickson and Kruschke [1998]) seem to comply with the heterogeneity claim in the sense that they appeal to more than one kind of representational structure (prototypes plus exemplars, exemplars plus rules, etc.) to be recruited for each different task. These are probably the kinds of models that traditionally are posited as hybrid. Yet it is unclear whether those representations are functionally integrated or are merely co-opted in different contexts. This is why these models are prone to the eliminativist criticism and why a pluralist explanation could still be suitable for them.

Other models deal more aptly with the functional integration claim but the extent to which they posit heterogeneous representations is unclear. We mentioned above proposals that emphasize the idea of information integration, such as (Hayes *et al.* [2003]; Sloutsky [2010]), or (Kiefer and Pulvermüller [2012]), and pointed out that the integration process is better understood as the formation of a rich common representational structure that constitutes the concept. If this interpretation is correct these proposals are closer to the hybridist position. However, some of these models may not be hybrid in the required sense. For instance, Hayes and Kearney ([2010], p. 214) contend that recent models that integrate data-driven and theory-based approaches “are not “hybrid models” in the sense used by Machery” because “they do not assume separate “exemplar” and “knowledge” modules”. Indeed, some of these approaches (as Carmichael and Hayes [2001]) take exemplars as the representational basis that incorporates theory-like knowledge.

As we said, we do not want to turn the issue into a terminological matter so let us state again what the hybridist is and is not committed to. The hybridist is committed to the existence of a rich common representational structure that constitutes the concept. The structure is rich in the sense of having different representational components and it is common in the sense that they are integrated as a functional unit. We admit that models that take, say, only exemplars as their representational basis are not hybrid in the required sense. The hybridist, however, is not committed to the existence of “separate modules” for each representational component. This is an assumption that some current categorization models make, but the notion of having a “rich” representational structure does not by itself involve such a demand. Recall our distinction above between what is selected and what is

represented. The fact that what is selected appears to have properties of a certain representational kind does not necessarily mean that what is stored in long-term memory is precisely such a kind of representation. To put an example, it could be the case that the long-term representation of a concept consisted in a collection of exemplars and that, when the concept is put to use, there were an on-line process of abstraction that produced a summary representation of the exemplars to be processed in working memory. We admit that it is difficult that a model like this could be regarded as endorsing hybrid concepts. However, just as what is selected might be prototype-like even though there were no actual prototypes among the long-term conceptual representation, it might be the case that we had conceptual representations that were not represented inherently *neither* as prototypes *nor* as exemplars, and yet sometimes they *behaved* as a prototype, and others as an exemplar. We contend that models that are based on such conceptual representations are closer to the hybridist position as long as they endorse a rich common functionally integrated representational basis for each concept. We want to explore such an interpretation in two recent approaches: the Varying Abstraction Framework (Verbeemen *et al.* [2007]) and the Conceptual Structure Approach (Moss *et al.* [2007]; Taylor *et al.* [2011]). Neither of them has been posited as hybrid in the way in which the term has been traditionally employed –which is closer to the heterogeneity claim. Yet we think that they open up new possibilities of understanding what a rich common representation is, so this is why we will devote some space to analyzing them.

The Varying Abstraction Framework “allows intermediate levels of abstraction, in which some exemplars are merged together in a single representation and other exemplars are kept apart in separate representations” (Verbeemen *et al.* [2007], p. 549). In categorization tasks the context will determine the level of abstraction that is used in a decision –from “complete abstraction (i.e., a single summary representation) to no abstraction at all (i.e., every exemplar is represented in memory and activated in categorization decisions)” ([2007], p. 549). Now, the question is whether the representations posited in this approach can be regarded as hybrid. We think that there is room for such an interpretation given that: (a) the VAF endorses a notion of representation that is “rich” in

the sense that it assumes that for the same category abstractions and exemplars can coexist, and (b) the representations posited in the VAF are functionally integrated in the sense that exemplars can be merged together in a single abstracted representation, so there will be a functional effect of which exemplars are activated for a particular task. So there is room to interpret the VAF as meeting, to some extent, both the heterogeneity and the functional integration claims.

However, it is possible to challenge this interpretation¹⁰ by arguing that in the VAF concepts are not really hybrid given that it does not “glue together” exemplar and prototype pieces. Rather, it posits coherent, integrated concepts that can *look like* prototype or exemplar-based concepts depending on their content. Moreover, the VAF is based in the generalized context model (Nosofsky, [1984], [1986]), which is an exemplar-based model that computes similarity as a function of relevant stored exemplars. The difference would lie in the fact that the VAF posits pseudo-exemplars instead. To get pseudo-exemplars one has “to first make a partition of the category exemplars for each category involved and then to construct, for every subset of exemplars, a prototype by averaging over all the exemplars in that subset. These prototypes are called the pseudo-exemplars and are used to represent the category” (Verbeemen *et al.* [2007], p. 541). When the partition consists of a single subset then the system behaves as a prototype model, when the number of subsets equals the number of exemplars the system behaves as full-exemplar model, and other in-between behaviors occur for intermediate numbers of subsets. The gist of the objection, thus, is that in the VAF we have one coherent type of concept that can yield a range of behaviours, not a hybrid concept after all.

The question is what sort of representation a pseudo-exemplar is: could it be regarded as a sort of hybrid concept, or is it a single-kind concept that appears as a new competitor to the existing prototype and exemplar models? To address this issue, let us first stress again that the hybridist is not necessarily committed to the idea that the conceptual components are semi-separable in the quasi-modular sense that we mentioned above. It may be the case that prototype-like, exemplar-like or other concept properties were supported by the same rich constitutive basis. Second, it seems to

us that the spirit of the VAF is not to propose a new kind of representation to compete with prototypes and exemplars. In our view, the VAF does not discard that, for any given category, both prototypes and exemplars are present: they simply are supported by the same representational basis. Indeed, as Verbeemen *et al.* ([2007], p. 539) point out, exemplars and prototypes have a “fuzzy status”, so it is not entirely clear when a given representation counts as an exemplar or as a prototype. We think that pseudo-exemplars can be understood as representations that conflate prototype-like and exemplar-like properties, which appear as extreme poles of a continuum and which come up depending on the particular categorization task.

The central claim in the Conceptual Structure Approach (Moss *et al.* [2007]; Taylor *et al.* [2011]) is that “a given concept can be defined in terms of the features that make up its meaning, and that the quantity and quality of these features, as well as featural interrelationships, – its internal structure – determines how a concept is activated during normal language comprehension and production” (Moss *et al.* [2007], p. 219). The way a concept is activated depends particularly on two key feature properties: *distinctiveness* and *co-occurrence*. Distinctiveness is a matter of how many concepts share a given feature: the fewer concepts share a feature, the more distinctive the feature is. Co-occurrence has to do with the strength with which a feature is associated to other features: features that are strongly correlated tend to co-occur while lowly correlated features co-occur infrequently. In our view, properties like these determine what kind of common representation constitutes a concept and how its elements work as a functional whole. On the one hand, we think that the degree of distinctiveness can be related to how much a concept will appear as a prototype or as an exemplar. First, concepts that share many features are unlikely to appear as exemplars. Second, even though a distinctive feature may characterize prototypically a whole category (e.g., *has an udder* as distinctive of cows) at the extreme degree of distinctiveness we find features that belong to an individual exemplar of the category (e.g., the particular udder of my cow Daisy). On the other hand, co-occurrence will be crucial to determine which features are functionally related to which others –and, as research on conceptual structure investigates, what patterns of functional

disruption appear.

We think that this approach is capable of meeting the heterogeneity and functional integration claims of the hybridist view. First, the CSA endorses a rich common representation where both prototype-like and exemplar-like properties appear supported by the same distributed componential model, in which a concept is distributed among a number of features and its properties depend on the sorts of features (shared/distinctive, perceptual/functional, and so on) relevant for their content and context of processing. Second, the approach regards features of the same concept as typically interrelated parts that are activated together when the concept is retrieved and are functionally integrated in the sense that co-activated parts of the representation can influence the process without being actually selected. To illustrate the latter point we can consider the evidence pointed out by Moss *et al.* ([2007]) that having more semantic features may indeed be beneficial to conceptual and lexical processing. In our view, this piece of evidence would also show that all the activated features are functionally involved even if only a selection of them makes it to working memory. The rationale for this conclusion is the following: If the selection of features were not influenced by the features with which they are connected, then there should not be differences between concepts with a larger or a fewer number of them. And if the selection process were merely a process of letting the ineffectual features' activation decay, or of merely “going through” the different features until the relevant ones were selected, then one should find that having more features is disadvantageous. So the fact that there is a processing advantage in having more features shows that they are integrated in a functional whole, just as the hybrid concept theory contends.

One may raise about features the same kind of worry as the one raised about pseudo-exemplars: could not one say that the CSA poses a single kind of representation –namely, features– as the basis for all concepts? If so, then concepts cannot be regarded as hybrid after all. We think that such an objection focuses on the wrong level of analysis. Notice that features *are not* concepts; they simply provide the fine-grain structure of concepts. The level that matters to distinguish concepts is the level at which we find a functional unit, and this unit can be found in a collection of

interrelated features with both prototype and exemplar properties. Once again, the fact that prototypes and exemplars are not stored as separable components –e.g., as supported by entirely different sets of features– is not decisive against the hybrid interpretation.

With all their differences, we think that models based on the VAF or on the conceptual structure approach share with the hybrid view important commitments that challenge the explanations put forward by either the pluralist or eliminativist views. Nevertheless, we accept that our interpretations of the VAF or the conceptual structure approaches are contentious. As we said above, there is a trade-off between heterogeneity and functional integration. Perhaps these models lean towards the integration pole and so look less hybrid in the way in which this notion traditionally was understood. Yet we also think that our analysis opens up new ways of understanding what it takes to have hybrid concepts, and that, at the very least, this leaves the hybrid view in a better position than its competitors.

4.2. Meaning-extraction

One often cited –and then forgotten– use of concepts is meaning-extraction. The process of linguistic comprehension involves retrieving the thoughts that the speaker wants to communicate, and this process involves accessing and using concepts. Now, the three approaches to concepts that we are addressing in this paper –eliminativism, pluralism, and hybridism– entail different claims about what is accessed. The eliminativist emphasizes the heterogeneity of representations and thus entails the claim that the representation that is accessed varies from occasion to occasion, probably depending on specific demands of the communicative situation. So the representations accessed each time that the word 'cat' is processed do not have anything in common to be regarded as part of the concept CAT. The pluralist constrains the heterogeneity of concepts appealing to a number of kinds of representational structures that are employed for different purposes. So each time that 'cat' is processed a particular kind of representation –e.g., a prototype, an exemplar– will be accessed,

and each of those kinds will count as a different concept. The hybridist contends that all those heterogeneous representational structures are associated in a rich common representation that constitutes the concept. So each time that 'cat' is processed a single rich concept CAT is accessed, even if only some parts of it come to be selected.

We think that meaning-extraction offers a better case for an explanation in terms of conceptual structures that are, on the one hand, pretty heterogeneous and, on the other hand, functionally integrated enough so as to allow for linguistic comprehension. In this section we want to present two arguments in support of the hybridist view that lean on the idea that different conceptual structures are stored as a single representation. The first argument is that the claim seems to be supported by widely accepted models of linguistic comprehension. The second argument is that prototypes, exemplars, etc., behave as senses of polysemous expressions, and that polysemy is best treated as a phenomenon that involves a single representation (with some qualifications that we will see below). Let us consider each argument.

4.2.1. Linguistic comprehension and rich lexical entries

Linguistic comprehension involves accessing concepts and combining them. The selection of concepts is not a simple task of de-codification: hearers have to select concepts depending on contextual factors, and, usually, they have to *modulate* (we borrow the term from Recanati [2004]) the selected concepts before, and also in parallel, with the composition process. So, meaning extraction consists precisely in these processes of selection and modulation.

Now, a plausible claim is that meaning extraction involves accessing quite a lot of information. Take the following examples:

(3) That's a real book

(4) Tina Turner is no grandmother

- (5) Tina Turner is a grandmother
- (6) Chihuahuas are dogs
- (7) France is hexagonal
- (8) My job is a jail

It is plausible to think that to understand (3) the hearer has to retrieve an ideal of book. In order to interpret (4) as a true utterance, the hearer must activate the prototype of grandmothers, whereas (5) demands accessing to its definition. The concept (or part thereof) of dog relevant in (6) is theory-like. Examples (7-8), in turn, are metaphors or loose uses which demand modulations of stored concepts, which plausibly involves retrieving different structures: for instance, (7) probably involves retrieving an exemplar of a hexagon with which the shape of France is compared –given that it does not make much sense to speak of prototypes of geometrical figures; on the other hand, (8) would require to retrieve the prototype of life in jail so as to see that someone’s job shares some relevant features of this prototype.

Of course, the details of what is accessed in each of those specific examples can be questioned. Yet the general point is that different kinds of structures can be accessed in different utterances of a word. Now, we take it that a plausible (rough) explanation of what goes on when we understand what these utterances communicate involves a process whereby the hearer, constrained by contextual information, has to select the type of structure (prototype, theory, etc.) that is going to be relevant to get the speaker’s communicated thought. This idea has not been directly explored by any model of communication, as, in general, no model of communication until now has been preoccupied with the distinction between conceptual structures. However, we take it that some of them are congenial to what we are claiming. To mention an influential model, let us consider Relevance Theory (RT). According to RT (see Carston [2002]; Wilson and Carston [2007]), the process of understanding consists in accessing an encoded atomic concept and then narrowing or widening its extension in order to satisfy the presumption of relevance. We have argued elsewhere

(Vicente and Martínez-Manrique [2010]) that its commitment to atomism is not justified and it actually poses problems to understand how the processes of narrowing and widening are to be understood. What matters for our present purposes is that narrowing and widening require accessing the encyclopaedic information associated with the encoded concept, and selecting the relevant information: “lexical comprehension involves a process of ad hoc concept construction, based on information readily accessible from the encyclopaedic entries of the encoded concepts and constrained by expectations of relevance (...) In each case, the comprehension process works in the same way, by selection of an appropriate set of encyclopaedic assumptions to act as premises for the derivation of the expected contextual implications”. (Wilson and Carston [2007], p. 245).

So, according to RT, the comprehension process involves a process of selection that draws on accessible information. Now, relevance theorists are not too specific about the nature of encyclopaedic information. It is coherent with the theory that the encyclopaedic information is formed by different structures (a prototype, an exemplar, a theory, etc.). In fact, the examples above point towards that view about the encyclopaedic entry: in some cases, the ad hoc concept which satisfies the demands of relevance is constructed out of prototype information; in some other cases, the information used is world information, etc. So we think that it is utterly congenial to the RT model to claim that the selection process driven by the expectations of relevance involves accessing different structures located in a given lexical entry. What we want to note, in any case, is that hearers have to go through a process of *selection* among the accessed structures every time they have to understand an utterance. In our view, this means that every time a person hears a certain word, she has to select between the different structures associated to it as a first step. Then, she may also have to modulate one of them or construct an ad hoc concept using the information stored in one of these structures.

Models like this allow for the possibility of having different structures associated under a single label, e.g., a lexical entry.¹¹ Still, the claim that the different structures belong to a single lexical entry is compatible with three possibilities: the hybrid view that they all form a concept,

which is *the* concept associated with the word; the pluralist view that, though they all belong to the same lexical entry, they are still different concepts; and even the eliminativist view that thinks that they do not have any interesting property in common to be regarded as a concept. For instance, according to the pluralist view, a lexical entry for ‘grandmother’ could have separate conceptual structures such as: PROTOTYPE <Typical features: OLD, SWEET, WOMAN, GREY OR WHITE HAIR, LOVES CHILDREN, ...>; DEFINITION <MOTHER OF A MOTHER>; EXEMPLARS <MY GRANDMA; LOONEY TUNES GRANNY>. Each of these structures would count as a different concept and contextual information would make hearers select which of them has to be used in retrieving the speaker’s meaning. In contrast, the hybrid view holds that all those elements are related as a single concept. We want to offer two arguments in favor of the hybrid position. The first one is an application of the criterion of functional stable coactivation. The second one, which we leave for the next subsection, relies on considerations from polysemy.

To apply the criterion of functional stable coactivation one has to show that the different structures within the lexical entry are stably coactivated, that the coactivated material is functional with respect to the task at hand, i.e., meaning extraction, and that the process of meaning-extraction does not consist in meaning competition. In the general account of meaning-extraction we sketched, in order to reach the meaning intended by the speaker, the hearer has to be able to access the different structures that can be expressed, or that can be used in modulation processes. This means that: (a) they all have to be activated; and (b) that they are all functional. It may well be that in some cases the hearer does not go through all the different structures, and so that not all of them are actually *used* in understanding what is communicated. If we adopt the model proposed by RT, it can be that the expectations of relevance are satisfied without consulting everything that is stored in the lexical entry. According to RT, we follow a path of least effort, such that we test interpretive hypotheses in order of accessibility (Wilson and Sperber [2004]). So, the process of selection can stop at an early stage. However, this does not mean that not all the information stored is functional. The information has to be available to the mechanism of selection because any part of it can be

relevant. So, it seems that if lexical entries store different structures, then those different structures are all active and functional in meaning-extraction. Finally, this selection process, unlike that of homonymy resolution, does not involve competition. Searching into the space of possible meanings of a word that a lexical entry gives access to is quite unlike choosing between *this* meaning or *that* meaning of a homonymous term¹².

It seems that there are two options available for pluralists and eliminativists. One of them is to hold that the different representations or kinds of representations are stored under separate lexical entries. This implies treating the different conceptual structures as homonymous meanings, which seems rather implausible (see below). The other option is to hold that prototypes, theories, and the like are not *selected*, but *triggered*. That is, in meaning extraction, hearers do not have to activate the whole conceptual information associated to a term, and then select a part; rather, linguistic and extra-linguistic context directly trigger one structure at a time. However, we think that this “triggering model” is extremely dubious. The fact that a quick process of selection (for instance, one in which the presumption of relevance is quickly met) may *look like* a case where the context simply triggers one interpretation does not mean that there is no underlying process that makes a fast decision about the appropriateness of the interpretation. This is more patent in cases in which meaning extraction is not so fast and smooth, which cannot be accounted by a triggering model.

4.2.2. Polysemy and hybrid concepts

We have mentioned that some authors hold that the process of selecting a meaning within the conceptual information that a lexical entry gives access to is akin to the process of polysemy resolution. The similarity is strengthened if we think about just the process of selecting a conceptual structure. In fact, even Machery proposes to think about this process in terms of polysemy resolution, to the extent that he advances what he calls ‘the polysemy hypothesis’ (see Machery, [2009]; Machery and Seppälä [2011]). If we look at Machery’s linguistic evidence, we see that there

is a common pattern between such evidence and what happens with many polysemous expressions. Machery provides null contexts for his input, i.e., contextual information leaves open which conceptual structure has to be recovered. Now, if we provide a null context for a homonymous term, we do not get the kind of reactions that Machery obtains. That is, we do not get “in a sense yes, in a sense no” responses. Any referential expression combined with ‘is a good bank’ does not elicit a “in a sense yes, in a sense no” state of mind provoked by the two homonymous meanings of ‘bank’ (river bank vs. financial institution). However, if you combine a referential expression (say, a demonstrative, or a definite description such as ‘the book on the top shelf’) with ‘is a good book’, you may provoke that state of mind. The hearer, if forced to consider truth-conditions, flips perspectives, and sees that the book may be good in one sense, say, as a text, and not good in another (as a tome).

The different conceptual structures associated to a word behave like the polysemous senses of a term in another respect. Frisson ([2009]) holds that the representation of the meaning of a polysemous word may be underspecific. He distinguishes between *meanings* and *senses*. A homonymous word has at least two different meanings, while a polysemous word has several different senses (although each of the different meanings of a homonymous word may also have several different senses). The representation of the meaning of a polysemous word is typically underspecific in the sense that tokening the representation does not imply tokening one of its specific senses –senses being the semantic material that goes into propositional contents. If you hear in a neutral context ‘all banks around are terrible’, it seems that you activate the two homonymous meanings of ‘bank’. However, if you hear in a neutral context ‘all books around are terrible’, in a first step you activate an underspecific representation; only later, and only if it is required, you determine which of the senses of ‘book’ is the one intended. That is, you retrieve the thought that was communicated at a later step, and only if you find it necessary. Otherwise, you go on reading or listening without resolving the issue of what sense of ‘book’ the writer or reader had in mind. In contrast, homonymy resolution is mandatory: the hearer is *forced* to select which of the available

meanings is the one intended by the speaker. Now, it seems that if you hear something like ‘there is a group of grandmothers over there’, you may well not resolve the issue of whether the speaker is telling you about prototypes or about actual grandmothers. If the context does not give you more information, you may be content with entertaining an underspecific representation.

What we are saying so far is suggestive of the thesis that prototypes, theories and the like are different senses of polysemous expressions. Now, our claim is that, if this is the case, then it seems that we are in a particularly good position to argue that they are all unified under a single conceptual representation. The reason is twofold. First: influential models of polysemy tend to treat it as a phenomenon of expression of different *senses*, and it is plausible that that they are different senses of a single concept. Second: polysemy resolution seems to be a clear case where we find stable functional coactivation of different pieces of information.

Frisson’s distinction between senses and meanings suggests that while meanings can be thought of as concepts, senses are better thought of as facets or parts of concepts (or, perhaps, as instantiations of concepts). Frisson’s notion of underspecific representations is open to two interpretations: representations can be underspecific either because they contain almost no information, or because they are very rich in information. For instance, the representation of the meaning of ‘school’ may consist in a simple pointer to the constellation of senses that ‘school’ can express, or it can consist in the whole network of related senses. Data collected by Klepousniotou and colleagues (see Klepousniotou *et al.* [2012]) support the rich information hypothesis, such that the different senses of a polysemous word would form part of a single representation. It makes sense to think about the single representation as a concept, and about senses as parts of that concept.

On the other hand, recent research in polysemy seems to be well exemplified by Pustejovsky’s ([1995]) “generative lexicon” model –which Frisson ([2009]) believes compatible with his underspecification approach–where lexical entries give access to one big structured concept. For instance, entries for nouns give access to conceptual information that includes information about origins, function, and constitutive features of the noun’s referent (Pustejovsky’s

“qualia structure”). Linguistic and extra-linguistic context is then used to select which part of the concept is going to enter into the truth-conditional meaning of the utterance. Coming from another tradition (Cognitive Linguistics), Cruse ([2004]) explains a good number of cases of polysemy as expressions of different facets and perspectives of a single concept.

To be sure, polysemy is a multifarious phenomenon, and not all cases of polysemy are accountable in terms of one concept/different senses. However, many cases do seem to fit this kind of explanation. So, we think it is reasonable to hold that, if what is at issue is how many concepts there are in a given lexical entry, claiming that the allegedly different concepts display polysemy should induce us to believe that there is only one.

As we say, polysemy resolution is typically a process where we find stable functional coactivation of different pieces of information. The evidence reveals that activation of one of the senses of polysemous expressions facilitates activation of the other senses as well (see Klepousniotou, Titone, and Romero [2008]). In contrast, homonymy resolution involves competition, rather than facilitation. Once a meaning is selected, access to the other meaning is blocked. It is plausible to think that there is no sharp distinction between homonymy and polysemy: some polysemes behave as homonyms (see Foraker and Murphy [2012]). It seems that the more related the senses, the more facilitating effects we find. In cases like the “meat grinding” rule, which accounts for the general senses “rabbit/animal” and “rabbit/meat of the animal”, each sense facilitates the other, while there may be no facilitating effects between the two senses of paper in ‘liberal paper’ and ‘shredded paper’ (see Klein and Murphy [2001]); although it has to be noted that Pylkkänen *et al.*, ([2006]), working with the same pairs and a different methodology obtained different results).

Polysemes whose senses are closely related show a processing advantage: words with multiple, and related senses, tend to be responded to faster than words with fewer senses (see, e.g., Azuma and van Orden [1997]); Rodd *et al.*, [2002]; Klepousniotou and Baum [2007]). These results are parallel to those we mentioned in section 4.1 about the beneficial effect of having more

semantic features, and we take that the explanation points in the same direction: the conceptual information in the lexical entry can constrain the selection process because it is related so as to form a common functional whole. If this were not the case, then having more senses should be either disadvantageous for the speed of response (if one assumes that the hearer simply “goes through” the different senses), or they should have no effect at all (if one assumes that the relevant sense is simply triggered). It seems, then, that the information in the lexical entry that is not actually selected (i.e., what is merely activated) does exert influence on the selected.

Now, what interests us most from this debate is the idea that in some cases polysemy resolution involves the coactivation of different senses that facilitate each other. The interesting cases in question are those where the senses are closely related. It is still not clear what close-relatedness amounts to. Some authors (e.g. Klepousniotou, Titone and Romero [2008]) take it that two senses are closely related if they share a high number of features. Thus, for instance, according to Klepousniotou *et al.*, the animal-sense of ‘rabbit’ (as in ‘a rabbit is running’) shares with its meat-sense (as in ‘the rabbit was delicious’) the features [+animate, +farm animal, +edible, +meat]. Their idea is that the senses of a polyseme typically overlap, which explains why we are more able to keep all of them in mind, and so why they facilitate each other and why words with many senses are easier to process. Foraker and Murphy ([2012]) (rightly) dispute this approach: it is not the case that ‘rabbit’ retains the four features above in its two senses. In the animal-sense, the case of a running rabbit, the rabbit is not edible, and it is not meat. In the meat-sense, the rabbit is not animate. The same seems to hold for many other instances of closely related senses of polysemous expressions, such as instances of the container-for-content polysemy (‘The DVD is one-hour long’, ‘the DVD is shiny’), author-for-work of author polysemy (‘My father was a friend of Picasso’, ‘I like that Picasso’), etc., or the numerous different senses of ‘school’ (see Frisson [2009]). However, we think that, if the “polysemy hypothesis” is conceded, the prototypes, theories, etc. associated with a certain category should count as closely related senses under any construal of “sense relatedness”. If what matters, after all, is sharing a high number of features, prototypes, theories etc. would pass the

test. But even if sense relatedness consists in something else, we would expect that a prototype and a theory of a certain category would be closely related. But of course, the issue is not whether prototypes, theories, etc. count as closely related senses under one rendering of the notion or another. The issue is whether they do behave like “rabbit–the-animal” and “rabbit-the-meat”. To us, it would be surprising if they turned out to pattern together with homonyms and “distant” polysemes, rather than with ‘rabbit’, but this is an empirical question that has not been addressed yet¹³.

In sum, the “polysemy hypothesis”, i.e., the hypothesis that prototypes, theories, etc., are senses of polysemous expressions, and so that meaning-extraction in these cases consists in polysemy resolution adds ammunition to the “one big concept” idea, as (i) polysemy resolution is in many cases theoretically approached as selecting a sense within one concept, and (ii) interesting instances of polysemy resolution involves stable functional coactivation of senses. Taken altogether, these considerations render more plausibility to the hybrid view of concepts with respect to the pluralist and eliminativist alternatives.

5. Conclusion

Despite the criticisms from eliminativism and pluralism, hybrid concepts are a live alternative in the theory of concepts. Against eliminativism, it is possible to come up with criteria to tell hybrid concepts apart, such as the notion of functional stable coactivation. Against pluralism, there are models of categorization and of meaning-extraction that are more easily interpretable in hybrid terms. Hybridism thus holds the promise of offering a view of conceptual representations that play a unified role in a diverse range of phenomena.

Funding

Spanish Ministry of Economy and Competitiveness (FFI2010-15717 to A.V., FFI2011-30074-C02-01&2, to F.M. and A.V. respectively).

Acknowledgments

An earlier version of this paper was presented at the 2012 ESPP Conference. We are grateful to the audience there for valuable comments. Special thanks to Asier Lafarga and Eric Margolis, with whom we had very interesting discussions that helped us to shape our then vague ideas. *BJPS* was able to recruit three reviewers who made extensive and very good comments that have, without a doubt, improved the paper. Thank you to all three reviewers, as well as to the editors of this journal. This work is fully collaborative. Order of authorship is arbitrary.

Agustín Vicente

1. *Ikerbasque: Basque Foundation for Science, Alameda Urquijo, 36-5, 48011, Bilbao, Spain &*

2. *University of the Basque Country. Linguistics and Basque Studies Department, Universidad Ibilbidea, 5, 01006 Vitoria, Spain.*

agustin.vicente@ehu.es

Fernando Martínez-Manrique

Departamento de Filosofía I, Edificio de Psicología,

Universidad de Granada,

18071 Granada, Spain

fmmanriq@ugr.es

References

- Anderson, J. R. [1983]: *The Architecture of Cognition*, Cambridge, MA: Harvard University Press.
- Anderson, J. R. and Betz, J. [2001]: ‘A Hybrid Model of Categorization’, *Psychonomic Bulletin and Review*, **8**, pp. 629–47.
- Ashby, F. G. and O'Brien, J. B. [2005]: ‘Category Learning and Multiple Memory Systems’, *Trends in Cognitive Science*, **9**, pp. 83–9.
- Azuma, T., and G. C. Van Orden [1997]: ‘Why Safe Is Better than Fast: the Relatedness of a Word’s Meanings Affects Lexical Decision Times’, *Journal of Memory and Language*, **36**, pp. 484–504.
- Barsalou, L. W. [1999]: ‘Perceptual symbol systems’, *Behavioral and Brain Sciences*, **22**, pp. 577–609.
- Carmichael, C. and Hayes, B. K. [2001]: ‘Prior knowledge and exemplar encoding in children’s concept acquisition’, *Child Development*, **72**, pp. 1071–90.
- Carston, R. [2002]: *Thoughts and Utterances*, London: Blackwell.
- Carston R. [unpublished]: ‘Polysemy, Underspecified Meanings, and the Lexicon’.
- Cruse, A. [2004]: *Meaning in Language: An Introduction to Semantics and Pragmatics*, Oxford: Oxford University Press.
- Erickson, M. A. and Kruschke, J. K. [1998]: ‘Rules and exemplars in category learning’, *Journal of Experimental Psychology: General*, **127**, pp. 107–40.
- Falkum, I. L. [2011]: *The Semantics and Pragmatics of Polysemy: A Relevance-Theoretic Account*, PhD Thesis. University College London.
- Foraker, S. And Murphy, G. L. [2012]: ‘Polysemy in Sentence Comprehension: Effects of Meaning Dominance’, *Journal of Memory and Language*, **67**, pp. 407–25.
- Frisson, S. [2009]: ‘Semantic underspecification in language processing’, *Linguistics Compass*, **3**, pp. 111–27.
- Gonnerman, C. and Weinberg, J. M. [2010] ‘Two uneliminated uses for “concepts”’: Hybrids and

- guides for inquiry', *Behavioral and Brain Sciences*, **33**, pp. 211–12.
- Hayes, B. K. [2006]: 'Knowledge, development, and category learning', *Psychology of Learning and Motivation*, **46**, pp. 37–77.
- Hayes, B. K., Foster, K., and Gadd, N. [2003]: 'Prior knowledge and subtyping effects in children's category learning', *Cognition* **88**, pp. 171–99.
- Hayes, B. K. and Kearney, L. [2010]: 'Defending the concept of “concepts”', *Behavioral and Brain Sciences*, **33**, p. 214.
- Keil, F. C. [2010]: 'Hybrid vigor and conceptual structure', *Behavioral and Brain Sciences*, **33**, pp. 215–16.
- Kiefer, M., and Pulvermüller, F. [2012]: 'Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions', *Cortex*, **48**, pp. 805–25.
- Klein, D. E. and Murphy, G. L. [2001]: 'The representation of polysemous words', *Journal of Memory and Language*, **45**, pp. 259–82.
- Klepousniotou, E., and Baum, S. R. [2007]: 'Disambiguating the ambiguity advantage effect in word recognition: an advantage for polysemous but not homonymous words', *Journal of Neurolinguistics*, **20**, pp. 1–24.
- Klepousniotou, E., Titone D., and Romero C. [2008]: 'Making sense of word senses: the comprehension of polysemy depends on sense overlap', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **34**, pp. 1534–43.
- Klepousniotou, E., Pike G. B., Steinhauer, K., and Gracco V. [2012]: 'Not all ambiguous words are created equal: An EEG investigation of homonymy and polysemy', *Brain and Language*, **123**, pp. 11–21.
- Kruschke, J. K. [2005]: 'Category Learning', in K. Lamberts and R. L. Goldstone (eds), *The Handbook of Cognition*, London: Sage, pp. 183–201.
- Machery, E. [2005]: 'Concepts are not a natural kind', *Philosophy of Science*, **72**, pp. 444–67.
- Machery, E. [2009]: *Doing without Concepts*, Oxford: Oxford University Press.

- Machery, E. [2010]: 'Précis of *Doing without Concepts*', *Behavioral and Brain Sciences*, **33**, pp. 195–244.
- Machery, E. and Seppälä, S. [2011]: 'Against hybrid theories of concepts', *Anthropology and Philosophy*, **10**, pp. 99-126.
- Malt, B. C. [1994]: 'Water is not H₂O', *Cognitive Psychology*, **27**, pp. 41–70.
- Malt, B. C. and Sloman, S. A. [2007]: 'Artifact Categorization: The Good, the Bad, and the Ugly', in E. Margolis and S. Laurence (eds), *Creations of the Mind: Theories of Artifacts and Their Representation*, Oxford: Oxford University Press, pp. 85–123.
- Margolis, E. and Laurence, S. [2010]: 'Concepts and Theoretical Unification', *Behavioral and Brain Sciences*, **33**, pp. 219-20.
- Medin, D. L. and Rips, L. J. [2005]: 'Concepts and Categories: Memory, Meaning and Metaphysics', in K. J. Holyoak and R. G. Morrison (eds), *The Cambridge Handbook of Thinking and Reasoning*, Cambridge: Cambridge University Press, pp. 37–72.
- Moss, H.E., Tyler, L.K. and Taylor, K.I. [2007]: 'Conceptual structure', in G. Gaskell (ed), *Oxford Handbook of Psycholinguistics*, Oxford: Oxford University Press, pp. 217–34.
- Murphy, G. L. [2002]: *The Big Book of Concepts*, Cambridge, MA: MIT Press.
- Nosofsky, R. M. [1984]: 'Choice, similarity, and the context model of classification', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **10**, pp. 104–14.
- Nosofsky, R. M. [1986]: 'Attention, similarity, and the identification-categorization relationship', *Journal of Experimental Psychology: General*, **115**, pp. 39–57.
- Piccinini, G. and Scott, S. [2006]: 'Splitting Concepts', *Philosophy of Science*, **73**, pp. 390–409.
- Pylkkänen, L., Llinás, R., and Murphy, G. L. [2006]: 'The representation of polysemy: MEG evidence', *Journal of Cognitive Neuroscience*, **18**, pp. 97–109.
- Pustejovsky, J. [1995]: *The Generative Lexicon*, Cambridge, Massachusetts: MIT Press.
- Rayo, A. [2011]: 'A plea for semantic localism', *Noûs*. On-line first. DOI: 10.1111/j.1468-0068.2011.00846.x

- Recanati, F. [2004]: *Literal Meaning*, Cambridge: Cambridge University Press.
- Rodd, J., Gaskell, G., and Marslen-Wilson, W. [2002]: ‘Making sense of semantic ambiguity: semantic competition in lexical access’, *Journal of Memory and Language*, **46**, pp. 245–66.
- Scarantino, A. [2010]: ‘Evidence of coordination as a cure for concept eliminativism’, *Behavioral and Brain Sciences*, **33**, pp. 223–24.
- Sloutsky, V. M. [2010]: ‘From Perceptual Categories to Concepts: What Develops?’, *Cognitive Science*, **34**, pp. 1244–86.
- Taylor, K. I., Devereux, B. J., and Tyler, L. K. [2011]: ‘Conceptual structure: Towards an integrate neurocognitive account’, *Language and Cognitive Processes*, **26**, pp. 1368–401.
- Verbeemen, T., Vanpaemel, W., Pattyn, S., Storms, G., and Verguts, T. [2007]: ‘Beyond exemplars and prototypes as memory representations of natural concepts: A clustering approach’, *Journal of Memory and Language*, **56**, 537–54.
- Vicente, A., and Martínez-Manrique, F. [2010]: ‘On Relevance Theory’s Atomistic Commitments’, in B. Soria and E. Romero (eds), *Explicit Communication: Essays on Robyn Carston’s Pragmatics*, Basingstoke: Palgrave MacMillan, pp. 42–57.
- Voorspoels, W., Storms, G., Vanpaemel, W. [2011]: ‘Representation at different levels in a conceptual hierarchy’, *Acta Psychologica*, **138**, pp. 11–8.
- Weiskopf, D. [2009]: ‘The Plurality of Concepts’, *Synthese*, **169**, pp. 145–73.
- Wilson, D. and Sperber, D. [2004]: ‘Relevance Theory’, in L.R. Horn, and G. Ward (eds), *The Handbook of Pragmatics*, Oxford: Blackwell, pp. 607–32.
- Wilson, D. and Carston, R. [2007]: ‘A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts’, in N. Burton-Roberts (ed), *Pragmatics*. Houndmills: Palgrave, pp. 230–60.

¹ We want to defend the plausibility of hybridism. Hybridism as such is not committed to the idea that hybrid concepts are shared, such that every subject who possesses the CAT concept possesses

exactly the same hybrid structure. That is, we do not find it necessary to dispute the rejection of what Weiskopf has called ‘the singularity assumption’. Thanks to an anonymous referee for raising this point.

² However, we think that something like the connection principle is on the right track to individuate hybrids. As a referee pointed out, our own criterion, which we develop in section 3, can be regarded as an elaboration of that principle, especially our point that hybrid parts are functionally relevant for each other.

³ This is the case, for instance, of the *ad hoc* concepts posited by Relevance Theory (Wilson and Carston [2007]).

⁴ Note, however, that different conceptual structures are typically non-coreferential. The prototype of cat does not have the same reference as the cat-theory (see Margolis and Laurence [2010]). We could say that the different conceptual structures associated with a category have similar referential aims –an idea which is difficult to make precise-, or something of that sort, but not that they are coreferential.

⁵ For instance, Hayes *et al.* ([2003]) (see also Hayes [2006]) show the existence of interactions between prior theories and exemplar learning that support the idea that both kinds of information are integrated in the course of learning; Sloutsky ([2010]) emphasizes the role of the ability of integrating cross-modal information in the development of concepts; Kiefer and Pulvermüller ([2012]) support two different levels of integration of distributed features, first, in a representation drawn from different modality-specific systems and, second, in a supra-modal higher-level representation.

⁶ As a referee pointed out, too much emphasis on the idea of simultaneity seems to rule out the possibility of spreading activation within a concept. Of course, we do not wish to rule out such a possibility, but it must be noticed that only spreading activation occurring within certain time limits ought to be allowed for the activated elements to count as part of the same entity –otherwise activation is a process that can go on indefinitely.

⁷ We are rephrasing here an objection raised by an anonymous referee.

⁸ Going to the case of prototypes, theories, etc. It seems plausible to hold that when we hear the word ‘German’, we activate both the definition of ‘German’ (person born in Germany) and its associated stereotype. Both in ‘he is German’ (definitional reading intended) and ‘he is very German’ (stereotypical reading intended), both kinds of representations become active (of course, this is a speculation at this stage, though we think is related to Machery and Seppälä’s ([2011]) results and our interpretation of them). However, if we hear ‘he is German’ we may activate certain representations (say, about geography) that we do not activate when we hear ‘he is very German’.

⁹ This is also a reason why Margolis and Laurence ([2010]) regard their own account of concepts as hybrid only in the sense that different representational parts are “bound to the same mental symbols” but not in the sense that those parts activate each other for a given task. In our terms, those concepts meet the heterogeneity claim but are not functionally integrated.

¹⁰ We owe the following objection to an anonymous referee.

¹¹ Another model that illustrates our approach is Rayo’s “grab-bag” model, in which lexical entries give access to heterogeneous information in heterogeneous formats and meaning extraction consists basically in a process of selection (Rayo [2011]). Recanati’s ([2004]) spread of access model is another congenial approach.

¹² In fact, some authors in RT (Falkum [2011], Carston [unpublished]) suggest that there is no essential difference between the process of meaning-extraction we are focusing on and polysemy resolution (see below).

¹³ It is suggestive to think that, e.g., the polysemy of ‘grandmother’ is similar to the polysemy of ‘book’. A particularity of this latter polysemy is that the two senses or facets (“text” and “tome”) of ‘book’ receive so much activation that it is possible to bind a non-selected sense, as, e.g., in ‘I could not finish this book. Please put it back on the shelf’. In the ‘grandmother’ case, it is possible to find effects that call for a similar explanation: A: ‘You know? Tina Turner is a grandmother’; B: ‘Indeed she is!’. In B’s response ‘grandmother’ seems to stand for *both* the definition/theory and the

prototype (her response would be false either if Tina Turner is not a biological grandmother or if she does not have some central features of the prototype of GRANDMOTHER). This suggests that both senses of 'grandmother' receive a high activation, and that, of course, they do not compete one against each other.