Abstract:

The content of this article amounts to a somewhat controversial terminological proposal: the term ‘concept’ is most fruitfully construed as ‘a mental representations having a lexical correlate’. Such a definition makes it possible to treat ‘concept’ as a technical term across the cognitive sciences, but also preserving most intuitions from a looser use of this word in the literature. The central points consist in a) appreciating the qualitative difference between the mental representations correlated with lexical labels and other mental representations, and b) accepting this difference as an effect of the causal influence of language on cognition. The argument is supported by a review of recent empirical results.

Keywords: concept, mental representation, Cognitive Science, categorisation, compositionality, lexical item, language, cognition, nonconceptual content, mentalism

1. Introduction and definitions

This paper presents an argument for a terminological proposal regarding the use of ‘concept’ as a term in Cognitive Science (CS). This terminological character is worth reemphasising: the text to follow does not advance or support any particular theory of what concepts are (although it does presuppose a particular ontological perspective, i.e. mentalism) but rather aims at showing that a certain way of understanding the term ‘concept’ would lead to its more unified, disciplined, and theoretically fruitful use across the cognitive sciences. Specifically, the claim is that the name ‘concept’ should be reserved for those mental representations that are correlated with lexical units. By no means does this imply a ‘linguistic theory of thought’ whereby all mental activity (or at least higher-level cognition) would be based on wordlike

---

1 Research reported in this paper is based on Chapter 4 of my unpublished doctoral dissertation (Wacewicz 2009).
mental representations. The argument presented here is of conceptual nature but is nevertheless grounded in recent empirical results.

Cognitive Science (CS) as understood in this text is taken to be a much broader category than traditional Cognitive Science. CS is assumed to be any study of the mind/brain with strong commitments to naturalism, internalism (i.e. assumption that an organism’s behaviour and mental activity can at least potentially be exhaustively explained with reference only to its internal states), and interdisciplinary collaboration, and at least some commitment to the validity of the notion of mental representation and the representational level of description\(^2\). Crucially, the perspective of CS presupposes mentalism about concepts, i.e. requires that concepts be understood as existing in the minds of individuals rather than being abstract or ideal entities that exist externally to individual cognitive systems.

Although the term ‘concept’ is frequently used in cognitive-scientific literature, its meaning is usually left unspecified, and there does not seem to be a unified and commonly agreed definition. Minimally, it can be assumed that concepts must be (a) a subtype of mental representations that (b) are employed as basic units in a range of higher-level cognitive operations. Mental representation is defined broadly, as any cognitive structure whose identity through time remains stable enough for it to be consistently deployed and redeployed. As to the list of cognitive functions, the most frequently enumerated are categorisation, inference and reasoning, communication/word meaning, and others include reference determination, representation, learning, understanding, explanation, planning, and prediction. Unless stated otherwise, ‘concepts’ means simple concepts (e.g. DOG), as opposed to complex concepts (e.g. GREY DOG) which arise from the combination of two or more simple ones.

Finally, a short comment is in order regarding the notational convention. Regular spelling is used for denoting objects in the extralinguistic reality (dog), upper case letters for concepts (DOG), single quotation marks for linguistic units, such as words corresponding to concepts and objects (‘dog’); single quotation marks are also used as distancing quotes.

### 2. Concepts as correlates of lexical items

The central proposal of this text regards the terminological linking of concepts to lexical items. The term ‘concept’ should be applied to those mental representations that are correlated with entries in the idealised mental lexicon; these include compounds and longer idiomatic expressions that are largely semantically noncompositional and are stored as single units. Representations that have a nonverbal nature, that is sensorimotor representations such as images, maps, sketches, sensations, motor schemas, proprioceptive schemas, etc. can of course enter conceptual structure in various ways and configurations (or at least are not in principle excluded), but they cannot be concepts, in that without some linguistic component they cannot exhaustively form complete structures of individual concepts. Such a convention, it is argued, manages to capture a cognitively real distinction, i.e. the special status of this type of mental representations stemming from the causal influence of language (specifically: lexical labels) on cognition.

Three major objections seem to emerge against such a definition. Firstly, the status of ‘concepthood’ would appear to depend on a decision of a lexicographer, which not only looks arbitrary but also refers to an externalised individuation criterion – rather than a mentalistic, system-internal criterion called for in a cognitivist approach. Secondly, the above position is not the standard position in Cognitive Science; in fact, numerous authors assume, openly or implicitly, that “there are many other concepts for which we probably do not have words” (Green et al. 2000 [1996]: 302). For example, it is common for linguists and psycholinguists (such as Steven Pinker, 1995 [1994], or Ray Jackendoff, 1997) to assume that conceptual structure is prior to language and though possibly differing in range and richness, does not differ dramatically in kind between linguistic and nonlinguistic creatures. This gives rise to the third objection, namely that the proposed definition, by stipulating that those who lack language lack concepts, would fail to do justice to the indisputable elaborateness of the cognitive systems of nonlinguistic creatures such as nonhuman animals or prelinguistic children.

In response to the last two points, it should be noted that recognising other concepts in addition to semantic-lexical concepts (i.e. those correlated with words) remains at least as problematic a solution. Semantic-lexical concepts form a coherent, well-defined category with straightforward criteria of concept individuation (i.e. by the corresponding lexical items); and in the actual application of ‘concept’ in the literature it is undoubtedly the semantic-lexical sense of

---

3 Most influentially, a similar position was argued by the philosopher Jerry Fodor (1975), who insisted that a preexisting structure of (innate) concepts was a logical requirement for the process of language acquisition to take place at all.
this term that is the prototypical sense. It is much less clear how to construe ‘other concepts’ or how they could be individuated, and it is difficult to establish a common criterion by which semantic-lexical and ‘other’ concepts would both qualify as types of the same category, i.e. of concepts. A broader understanding of the term ‘concept’ as a generic label for a stable mental structure remains a possibility, but this would make it coextensive with ‘category’ (or ‘category representation’) or ‘mental representation’. Ultimately, however, the case for the definitional link between lexical labels and concepts boils down to whether a cognitively real, qualitative difference can be shown to hold between two groups of mental representations, i.e. those correlated with lexical labels as opposed to other types of mental representations. Objections one and three are dealt with in the next two sections exactly as a consequence of establishing evidence for the cognitive reality of this difference.

2.1. The influence of language on cognition

Summarising – even in the most general terms – the present state of the ‘language and thought’ debate quite obviously exceeds the possibilities of this short article. Therefore, rather than addressing this issue, the present section aims at extracting (and documenting with experimental results) a number of uncontroversial ways in which language use does have a psychologically real impact on the cognitive system. Based on the list below, section 2.2. contains a more specific argument regarding lexical labels.

a) (quite trivially,) in modern societies there exist a range of cultural constructs that can only emerge – and be acquired – within social reality that itself relies on a framework laid out by language (e.g. Pinker and Jackendoff 2005: 206). It follows that certain concepts are in principle unattainable by nonlinguistic organisms. In contrast, language users can draw upon the pool of collective experience of their community in ways that allow them to guide their behaviour more successfully.

b) at least some (but by no means all) of human problem solving is dependent on the presence of ‘inner speech’; it is not only describable by words and sentences of one’s natural language,

---

4 “Vast domains of human understanding, including the supernatural and sacred, the specifics of folk and formal science, human-specific kinship systems (such as the distinction between cross- and parallel cousins), and formal social roles (such as “justice of the peace” and “treasurer”), can be acquired only with the help of language.” (Pinker and Jackendoff 2005: 206)

5 For developed arguments, see in particular Daniel Dennett 1994, and Stevan Harnad 2002.
but is actually rendered in such *stricte* linguistic structures. Whether or not this phenomenon can be implicit, i.e. operating below the threshold of consciousness, is a matter of controversy. Still, there are many familiar examples where problem solvers rely on explicit use of language, such as ‘talking to oneself’.

One widely-quoted experimental example of a situation where being a language user (as opposed to a nonlinguistic creature) predicts performance on a simple spatial cognition task comes from Elisabeth Spelke’s laboratory (Hermer-Vasquez et al. 1999)\(^6\).

c) more generally and fundamentally, language helps ‘discretise’ otherwise continuous cognition: divide it into distinguishable quanta. Abstract entities, relations, events, and processes become objectified\(^7\), thus providing the basic units – corresponding to linguistic units – on which computational processes can operate. These units remain stable in the absence of a co-occurrent perceptual stimulus and are easy for retrieval from long term memory, and maintenance, combination and manipulation in short term memory. This insight, historically reaching back at least to Enlightenment philosophers such as John Locke\(^8\), has been more recently developed by a number of researchers in cognitive science, e.g. Jackendoff (1997)\(^9\).

Empirical evaluation is hard to achieve, but relevant experimental data include studies such as that of enculturated chimpanzees faced with the task of grasping the relation “the same”: a prerequisite for success seemed to be the ability to assign a (quasi-)lexical label to this relation (Thompson et al. 1997).

d) language enables or at least enhances metacognition (‘thought about thought’): it makes it possible to form the (explicit) representations of representations (e.g. Jackendoff 1997: 202–

---

\(^6\) The experimental design required the subjects to integrate two kinds of information: geometric (long versus short wall) and colour (blue or white wall) in order to succeed on 100% of trials; relying on only one kind of information ensured success on 50% of trials. The rates of success of both rats and prelinguistic children were close to the latter figure (suggesting no integration), whereas adults were successful on almost all trials (suggesting integration). However, when the adults were engaged in a concurrent verbal shadowing task that placed heavy demands on their linguistic processing, their performance dropped to levels indicating the lack of integration of kinds of information (Hermer-Vasquez et al. 1999).

\(^7\) ‘Objectified’ not in the sense of becoming objective, but in the sense of Szwedek (e.g. 2002), that is in terms of the metaphor of concrete, physical objects.

\(^8\) Locke analyses this aspect of importance of language for mental processes in his *Essay Concerning Human Understanding* (1999 [1690]).

\(^9\) “Language is the only modality of consciousness that makes perceptible the relational (or predicational) form of thought and the abstract elements of thought. Because these elements are present as isolable entities in consciousness, they can serve as the focus of attention, which permits higher-power processing, anchoring, and perhaps most important, retrievable storage of these otherwise nonperceptible elements.” (Jackendoff 1997: 205)
205). This has ramifying consequences, of which the most important is perhaps the link to Theory of Mind: representing the representational states of others. An experimental indication of a strong relation between Theory of Mind and language is the pattern observed in false-belief studies, with language competence predicting performance on false belief tasks rather than the opposite way (Hale and Tager-Flusberg 2003; Lohmann and Tomasello [quoted in Jordan Zlatev, in press]).

e) developmentally, lexical labels facilitate the formation of categories: consistent use of a given word to describe examples of a certain salient aspect of reality helps to highlight commonalities between such elements and leads to the formation of an appropriate category. Without lexical labels acting as catalysts, category formation is delayed or not achieved at all. (This is different from points a) and c) where the ‘referents’ in the world are nonsalient and may be in principle unattainable without language).

For example, Sandra Waxman and Dana Markow (1995) show that in 12-month human children, words serve as inducing stimuli to form categories of simple objects; the observed effect was strongest with nonbasic-level categories. Even more strikingly, a recent study by Gary Lupyan (2006) identifies the same kind of effect in human adults.

f) the scope of at least some categories (the range of perceptual inputs that are categorised together) appears to depend partly on the way in which the corresponding word functions in a given language community. That is, the discontinuities on which categorisation is based depend – in addition to more fundamental perceptual, etc. factors – also on ‘Whorfian’ factors related to the use of a given natural language.

Colour categorisation has traditionally been the area most intensively researched in relation to the Whorfian ‘language influences thought’ hypothesis. While no significant effects have ever been experimentally established, there is convergent evidence from a large number of

---

10 Theory of Mind (ToM) is a somewhat unfortunate but extremely well established name for the socio-cognitive ability of humans and certain higher primates to perceive others as self-governed (non just self-propelling) and rational agents capable of entertaining their own, independent mental states, i.e. beliefs, desires, fears, hopes, etc. Developed Theory of Mind implies the ability to meta-represent, i.e. to have mental states about the mental states of others (e.g. ‘I know that John thinks that beer is in the fridge’). Although in reality it describes an extremely complex and graded set of cognitive abilities, ‘Theory of Mind’ serves as a useful, if simplistic, label that can be ascribed in a binary all-or-none way. In such cases, the traditional litmus test for the presence of ToM in a creature has been the creature’s performance in false-belief tasks. A success in a false-belief experiment requires being able to realise that another individual has an incorrect piece of information, and being able to predict this individual’s behaviour that results from its acting on this incorrect information. For a review of ToM in nonhuman primates see e.g. Heyes 1998.
recent studies that a person’s natural language does in fact exert some influence on their colour categorisation (for instance, Davidoff 2001; Gilbert et al. 2005). At the same time, one must admit that the effects are rather subtle.

Far from being an exhaustive overview, the above is simply a possible classification oriented towards the implicit influence of individual lexical units. An example of an alternative classification of the ways in which language “augments human computation” is the frequently quoted one developed by Andy Clark (1998): memory augmentation, environment simplification (language facilitates categorisation), coordination and the reduction of on-line deliberation (language facilitates explicit planning), taming path-dependent learning (learning from others in addition to from personal experience), attention and resource allocation (written and spoken words as external memory enhancements), data manipulation and representation (explicit manipulation of written text to structure one’s argumentation).

2.2. Concepts versus nonconceptual mental representations

Increasingly many studies in comparative psychology show that numerous aspects of animal cognition must be given interpretations that rely upon complex mental representations. The most prominent examples – alongside ones such as avian navigation or sophisticated principles of food caching, recaching and retrieval by food-storing birds – come from primate social cognition. For instance, the primatologists Robert Seyfarth and Dorothy Cheney (2001) report that highly gregarious baboons can mentally represent and update both the relations of kinship and hierarchical social status. Due to the problem of combinatorial explosion, in groups of as many as eighty individuals this is impossible to achieve by means of one-by-one

---

11 Jules Davidoff reviews, among other experiments, three cross-cultural ones (perceived subjective similarity, category learning, and recognition memory) and concludes (2001: 386): „Put together, these three new cross-cultural studies suggest that categorical perception shows the influence of language on perception. At the very least, our results would indicate that cultural and linguistic training can affect low-level perception... However, more than that, the results uphold the view that the structure of linguistic categories distorts perception by stretching perceptual distances at category boundaries.”

Aubrey Gilbert and co-workers (2005) found that colour discrimination was affected by the difference in the names of contrasted colours when the colours were presented in the right half of the visual field (thus processed by the ‘linguistic’ left hemisphere), but not in the left visual field. This effect was diminished when the subjects were assigned a verbal tasks related to verbal processing, but lasted when the task given engaged non-verbal working memory to a similar extent.
associations, i.e. storing each relation as a separate memory entry; the monkeys must instead employ sophisticated computations defined over ‘concepts’ such as FAMILY or DOMINANCE.

Similarly, human infants are cognitively complex creatures who from their earliest days display advanced understanding of the surrounding world (in ways roughly consistent with the Piagetian account, but at even younger ages). Consequently, infant cognition is routinely described in conceptual terms: it is customary in developmental psychology to speak of a child’s ‘concept’ of objecthood, number, identity, self, animacy, causation, force, etc. (examples are: “infant’s concept of occlusion”, Renée Baillargeon 2001; “infant’s concept of twoness/number”, Karen Wynn, 1992).

It is indeed undeniable that both prelinguistic children and at least some animals do possess advanced mental representations. Still, the common reply to the above examples is to deny that these representations have the status of *bona fide* concepts. The reason behind such a decision is that the presence of a lexical label, especially one frequently used in communication, exerts a profound effect on the functional characteristics of a given mental representation.

Thus, I argue that *bona fide* concepts differ from proto-concepts possessed by animals and infants – as well as from non-lexicalised representations in human adults – in the following ways (some of which are direct results of the correlations with lexical labels):

a) concepts become available for explicit inference (in a way described in 2.1. b),
b) concepts become stabilised, which facilitates their storage, retrieval, and manipulation (in a way described in 2.1. c),
c) concepts become further stabilised by their frequent activations caused by the use of the corresponding lexical items in communication,
d) concepts are even further stabilised by their frequent pre-activations caused by the use of semantically or phonetically related lexemes (as evident in the phenomenon of semantic and phonetic priming; see e.g. Jean Aitchison 1996 [1987]: 109),
e) concepts support compositionality in that they can enter very elaborate complex representations as their building blocks. While it can in principle be imagined that the mental representations in animals and children might have this quality as well, there is no evidence in favour of such a conjecture. On the contrary, what evidence there is points to the opposite. The very first linguistic utterances by children are holophrastic (see e.g. Dennis Bancroft
1995: 64) in a way suggesting limited compositionality of the underlying representations. In contrast, in normal human speech only very rarely do individual words form complete utterances: they almost invariably come as elements of larger compositional structures.

f) most concepts are domain-general in the sense that they can enter reasonings related to any subject. In contrast, the mental representations in infants and animals seem to be domain-specific, i.e. limited to a particular, narrow range of contexts. There is no evidence that the complex computational processes invoked to explain e.g. avian navigation or primate social cognition generalise beyond their proprietary domains. This mirrors the generality constraint (Gareth Evans, 1982)\textsuperscript{12}, which does not seem to be met in the case of infants/animals.

g) concepts are of a palpably different level of generality than some of the mental representations ascribed to animals and prelinguistic children. The latter are more general, lacking the rich inferential content of concepts; on the other hand, they are more closely dependent on sensory imagery. It seems that the mental representations of ‘causality’, ‘objecthood’, ‘occlusion’, etc. in children and animals can be more appropriately explained, not in terms of concepts, but rather in terms of image schemas (for such suggestions concerning infants see Mandler [cited by Jordan Zlatev: in press], and animals – Marc Hauser 1997 [1996]).

In summary, the overall changes to a mental representation resulting from its becoming correlated with a lexical ‘tag’ have a qualitative dimension and make the ascription of a different status fully legitimate\textsuperscript{13}.

\textsuperscript{12} The generality constraint stipulates that if a subject is in possession of a concept A, then they should be able to meaningfully combine this concept with all other (semantically relevant) concepts in their repertoire; it is a condition of the possession of the concept A that the subject be able to entertain all (sensible) thoughts that are comprised of the concept A together with any other concepts possessed by this subject. It is assumed that the combinations in question are limited to thoughts that have truth conditions, are well formed and not nonsensical, e.g. because of category mistakes. Evans himself concedes that the generality constraint might be merely “…an ideal to which our actual system of thoughts only approximately conforms” (1982: 105), but it remains an evaluative criterion nonetheless.

\textsuperscript{13} Some thinkers, such as Daniel Dennett (1996) and Euan MacPhail (1998), argue for a still much more profound divide between linguistic and language-less creatures. According to them, language is absolutely foundational in the development of the central notion of self. Hence, consciousness as we know it cannot be meaningfully attributed to nonlinguistic creatures. Such a stance meshes with the argument presented above, but does not follow from it as a necessary consequence.
Also, it is important to spell out an additional condition for ‘concepthood’, already implicit in the above argument; namely, that the lexical label correlated with a mental representation be part of a richer communicative system. While even a very simple organism can easily internalise an association between an arbitrary sign and a class of inputs, such a mechanical stimulus-response pairing is something very different from a word of human language. Most significantly, as remarked in point e), lexemes are not autonomous, holistic utterances, but units functioning almost exclusively as parts of larger structures; both the existing paradigmatic and the potential syntagmatic relations between lexemes are meaning-constitutive. In addition, the frequency and stability of the communicative use of a lexical label seems to be necessary to support some of its influences, listed above, on the associated cognitive structure (entrenchment effects).

In short, although one cannot ignore extant research that documents the richness and sophistication of cognition in nonlinguistic creatures, one can still qualify cognition in such creatures as nonconceptual. Very characteristically, this is precisely the stance assumed by many theorists of animal/infant cognition. The chief proponent of the notion of ‘nonconceptual content’, José Luis Bermúdez (e.g. 2003), identifies advanced cognitive operations available to nonlinguistic creatures and to an extent downplays the significance of language for thought. Still, he endorses what he calls “the Priority Principle”, that is the linguistic criterion in the divide between concepts and ‘nonconcepts’. This also serves to show that ‘the lexical criterion’ does not suggest the radical and heavily criticised ‘linguistic model of thought’, whereby all (higher) mental processes are defined over word-like units. Quite the opposite, the very presence of nonconceptual (i.e. non-linguaform) content directly points to the existence of the other kinds of cognitive mechanisms than those based on language.

3. Conclusion

On the present analysis, construing concepts as mental representations with lexical correlates may be considered an optimal solution from a cognitivist perspective. It has the double advantage of, firstly, preserving most of the pre-theoretical intuitions connected to the word ‘concept’, and secondly, of making ‘concept’ a technical term available for use across the

---

14 See especially Terrence Deacon (1997: 65–67), who takes paradigmatic and syntagmatic relations between signs to be constituentive of their symbolic character.
cognitive sciences. Several arguments, grounded in recent empirical data, were adduced for the qualitative difference that is made by mental representations’ being correlated with lexical items of a natural language (both as a direct result of such a correlation, and indirectly through its functioning within a linguistic cognitive system). The apparent problem of the sophistication of the cognitive systems of nonlinguistic creatures is resolved by reference to the notions of mental representation (in the general sense) nonconceptual content. In this context, the process of language acquisition in the child is seen as directly, causally responsible for the formation of fully-fledged concepts on the basis of more primitive proto-conceptual mental representations.

**Bibliography**


http://www.philosophy.ed.ac.uk/staff/clark/pubs/magic.pdf


http://cogprints.org/192/00/rolelang.htm ED 04/2006


