LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

EMPIRICAL STUDY

Verbal Symbols Support Concrete but Enable Abstract Concept Formation: Evidence From Brain-Constrained Deep Neural Networks

Fynn R. Dobler[a,b,d] Malte R. Henningsen-Schomers,[a] and Friedemann Pulvermüller[a,b,c,d]

[a]Brain Language Laboratory, Department of Philosophy and Humanities, WE4, Freie Universität Berlin [b]Charité – Universitätsmedizin Berlin, Einstein Center for Neurosciences Berlin [c]Cluster of Excellence 'Matters of Activity. Image Space Material', Humboldt Universität zu Berlin [d]Berlin School of Mind and Brain, Humboldt Universität zu Berlin

Abstract: Concrete symbols (e.g., *sun*, *run*) can be learned in the context of objects and actions, thereby grounding their meaning in the world. However, it is controversial whether a comparable avenue to semantic learning exists for abstract symbols (e.g., *democracy*). When we simulated the putative brain mechanisms of conceptual/semantic grounding using brain-constrained deep neural networks, the learning of instances of concrete concepts outside of language contexts led to robust neural circuits generating substantial and prolonged activations. In contrast, the learning of instances of abstract concepts yielded much reduced and only short-lived activity. Crucially, when

---

CRediT author statement – **Fynn R. Dobler**: investigation; formal analysis; writing – original draft preparation; writing – review and editing. **Malte R. Henningsen-Schomers**: conceptualization; methodology; writing – original draft preparation; writing – review and editing. **Friedemann Pulvermüller**: conceptualization; methodology; writing – original draft preparation; writing – review and editing; funding acquisition.

A one-page Accessible Summary of this article in nontechnical language is freely available in the Supporting Information online and at https://oasis-database.org.

The data that support the findings of this study and the analysis code are available via the Open Science Framework at https://osf.io/m8dg5.

This work was supported by the European Research Council (ERC) through the Advanced Grant "Material constraints enabling human cognition, MatCo" (ERC-2019-ADG 883811) and by the Deutsche Forschungsgemeinschaft under Germany's Excellence Strategy through the Cluster of Excellence "Matters of Activity. Image Space Material" (DFG EXC 2025/1 – 390648296). We would like to thank the high-performance computing service of Freie Universität Berlin and Martin Freyer and Phillip Krause for technical support. The authors declare no conflict of interest.

DOI: 10.1111/lang.12646

conceptual instances were learned in the context of wordforms, circuit activations became robust and long-lasting for both concrete and abstract meanings. These results indicate that, although the neural correlates of concrete conceptual representations can be built from grounding experiences alone, abstract concept formation at the neurobiological level is enabled by and requires the correlated presence of linguistic forms.

Keywords concept formation; grounding; abstract concepts; deep neural networks; neurocomputational modeling

Introduction

The meaning of concrete symbols, such as the words *hammer* or *ball*, can be picked up at least in part from the nonverbal contexts in which they are used. Important in this process are statistical regularities between symbol use and the presence of objects and actions of certain types (M. Tomasello & Kruger, 1992; Vouloumanos & Werker, 2009; Waxman & Markow, 1995). Even though such "grounding" of symbolic meaning in objects and actions may not be sufficient for acquiring all facets of meaning (Barsalou & Wiemer-Hastings, 2005; Kintsch, 1974; Pulvermüller, 2018c), it has been shown to be a necessary component of any semantic learning (Harnad, 1990; Searle, 1980). And even before and independently of language learning, infants and even animals may learn some concepts from, and ground them in, experiences, possibly due to the similarity structure across the instances that fall into a category (Mandler, 2004; Pearce, 2008; Perszyk & Waxman, 2018; Pusch et al., 2023). Therefore, most researchers today agree that at least a significant number of concepts and symbols need to be and are grounded in experiences, and, thus, in the "world." With this "grounding set" of symbols as a foundation, additional concepts and meanings can be learned and grounded indirectly, based on symbol contexts in which they are typically used (Cangelosi et al., 2000; Harnad, 2018). In essence, direct grounding establishes a symbol–world relationship through associations between linguistic symbols and perception- and action-related experiences, which is indispensable for semantics.

Author Twitter information: @bll_berlin

Correspondence concerning this article should be addressed to Friedemann Pulvermüller and Fynn R. Dobler, Brain Language Laboratory, Dept of Philosophy & Humanities, WE4, Freie Universität Berlin, Habelschwerdter Allee 45, 14195 Berlin. Email: friedemann.pulvermuller@fu-berlin.de, fynn.dobler@fu-berlin.de

The handling editor for this manuscript was Guillaume Thierry.

Direct and Indirect Grounding of Symbols and Concepts

Such associative grounding mechanisms may appear feasible for learning aspects of the referential meaning of concrete symbols, but they cannot explain all aspects thereof. For example, for explaining the difference between proper names, which can only be used to speak about one specific entity or person, and category terms applicable to a large set of instances, more is required than simple association (Nguyen et al., 2024). An unsurmountable set of challenges is apparently posed by highly abstract concepts and meanings, which, according to standard views, have correlates not in the world, but rather in mental space, where they relate to equally abstract conceptual and semantic features. Because of their abstract mental nature, it is commonly believed that abstract symbols such as *justice* or *game* pose problems for grounding accounts of semantic meaning.[1] Because these concepts lack an experienceable correlate in the world (Hale, 1988), it is not clear how any symbol–world relationship could possibly form. As a result, theorists have proposed that "hybrid" accounts are necessary to cover both concrete and more abstract concepts (Dove, 2016, 2022; Paivio, 1991). Within such accounts, the simpler and more concrete concepts are subject to grounding, whereas the more complex and abstract meanings are not subject to grounding in the (real) world but rather receive their content from the "mental world" or an amodal symbolic system.

Opinions differ on how exactly to best characterize abstract meaning and the semantic relatedness between different concepts. The most common proposal, adopted from structuralist semantic theory, is that sets of abstract semantic features, some of which may capture hierarchical taxonomic relationships, define a given concept or symbol (e.g., [+ANIMATE], [+BEAUTIFUL]; see, e.g., Katz & Fodor, 1963; Mahon & Caramazza, 2009). An alternative approach is offered by distributional semantics, which defines meaning in terms of multidimensional vectors coding for symbol co-occurrences across texts in large text corpora (Landauer & Dumais, 1997; Lenci et al., 2018). A further approach postulates that some of the apparently abstract features immanent to abstract concepts can be "grounded privately" in the individual, by means of them experiencing the related mental state or emotion and tagging it with a concept or symbol (Barsalou & Wiemer-Hastings, 2005; Borghi & Binkofski, 2014; Vigliocco et al., 2013).

Common to these approaches is the belief that different mechanisms are crucial for semantic grounding in the world, to which individuals have access to similar degrees, and for abstract semantics built on symbol–symbol relationships or private linkage of symbols, with information directly accessible only to the individual, be it emotions, mental states, or, in the extreme case, the innate conceptual features and representations of a universal "language of thought" (Fodor, 1975; Pinker, 2008). According to hybrid models of symbolic and grounded information (Dove, 2016; Paivio, 1990, 1991), the meaning of any symbol can include both "modal" sensorimotor information and "amodal" distributional or only-internally-accessible information, with abstract concepts either relying on amodal information exclusively or more strongly than concrete ones, to which the reverse pattern applies, that is, a relatively stronger (or exclusive) relationship to modal information.

The success of hybrid approaches may be due to the fact that they appear to be compromises between extreme positions, thus integrating different views to a degree. However, compromises may also include the weaknesses of one of the extremes, and in the worst case, of both. If hybrid proposals account for the semantically sophisticated class of abstract concepts, why should a grounding set of words be postulated at all, given that the grounding mechanism depends on associative learning, which is widely believed to be insufficient for semantics? Might it therefore be advisable to revert to classic cognitive theories applicable to all kinds of concepts and entirely based on abstract semantic features or symbol–symbol relationships? But this would ignore the grounding problem, which invalidates ungrounded semantic features, subject-internal private tags, and symbol distributions per se as a genuine mechanism (for discussion, see Harnad, 1990; Searle, 1980). After the grounding problem had apparently been solved for some concrete concepts, a hybrid approach seems to imply giving up on the grounding of abstract terms. Such a position would leave the meaning of symbols half unexplained; therefore, it appears weak and insufficient. Could there be ways to extend grounding-in-the-(real-)world to abstract concepts?

One perspective on grounding abstract symbols is offered by distributional semantics, if high word–word co-occurrence values are taken as a basis of indirect grounding of novel word meanings in semantic referential information provided by already known and grounded ones, following the symbolic theft hypothesis (Andrews et al., 2009; Cangelosi et al., 2000; Cangelosi & Stramandinoli, 2018; Parisi & Cangelosi, 2002). This can be based not only on explicit semantic explanation (*A zebra is like a horse with stripes*, whereby *horse* and *stripes* are in the grounding set; see Harnad, 1990), but also on symbol distribution (*zebra* frequently co-occurring with both *horse* and *stripe*; see, e.g., Andrews et al., 2009), which is applicable to abstract words frequently co-occurring with concrete ones (Cangelosi & Stramandinoli, 2018). However, many abstract symbols, and actually some of the most abstract ones, primarily co-occur with other abstract symbols (Lenci et al., 2018; Naumann et al., 2018). For example, the abstract symbol *justice* preferentially co-occurs with the also highly abstract items *criminal*, *system*, or *law* (data from Davies, 2018). For symbols to which this applies, indirect text-based grounding may fail. Therefore, at least a subset of abstract symbols would need to be directly grounded in object perception and motor action, unless an alternative avenue toward learning their relationship to real-world entities can be offered.

Such an avenue could still be provided by some type of "inner grounding." Using this expression loosely, it may even be applied to the private association between an inborn concept and a novel wordform. Some cognitivist theorists envisage ordered interlinking between symbols used by a language community and the individual's own private knowledge of emotions, mental states, and amodal concepts (see, e.g., Pinker, 2008). However, such an association process appears problematic. Most notably, it is unclear how to guarantee that the correct inner concept, feeling, or representation is activated together with a novel externally provided symbol, rather than a different one mismatching with the symbol's meaning. Likewise, information about the degree of specificity of the novel symbol would need to be available to allow the individual to decide whether the word is specific to one particular inner state or rather more general, comprising an entire class of similar ones instead. In case of evidence for unsuccessful learning, there would need to be criteria for revising and repeating the semantic learning process (for further discussion, see, e.g., Baker & Hacker, 2008; Wittgenstein, 1953), which are, however, missing. All these difficulties are relatively easy to tackle if the entity that the novel word is used to speak about is publicly accessible to learners and teachers—which, arguably, inner or mental states and activities such as anger or thinking are not.

A way to avoid these difficulties is offered by a grounding-in-action perspective (Glenberg & Kaschak, 2002; Pulvermüller, 2005): It is not the internal grounding as a private within-subject process that is relevant; what counts is the accessibility of the "inner" state or emotion in the behavior and actions of the language learner to those engaged in symbolic communication. The regular and natural expression of emotions and inner states in overt bodily actions and movements by the language-learning child is the basis of grounding of some of the symbols in the abstract spectrum (see also Dreyer & Pulvermüller, 2018; Moseley et al., 2012; Moseley & Pulvermüller, 2014; Pulvermüller, 2018b). Such grounding of abstract concepts was first proposed by Wittgenstein (1953), who emphasized that an internal state such as joy or suffering from pain is normally expressed in the behavior of the young learning child by natural oral, facial, and bodily expressions. These action-related manifestations of the inner state function as criteria for the correct application of the appropriate verbal symbols and descriptions by the language-competent adult and, hence, for the teaching of appropriate verbal expressions for the inner states experienced by the child. In this perspective, symbols for supposedly inner, but in fact action-related, emotional and mental states are learned not privately but interactively, whereby the interacting adult is guided by criteria for the behavioral manifestations of these states. In other words, at least some abstract symbols can be, and typically are, grounded in action. Such grounding-in-action can extend even to highly abstract expressions, such as *causation* or *regression to the mean* (Glenberg, 2022; Pulvermüller, 2018a). In line with this research and related modeling work (Henningsen-Schomers et al., 2023; Henningsen-Schomers & Pulvermüller, 2022), we here aim to further develop, describe, and investigate a neuromechanistic model of abstract semantic grounding in object perception and motor action.

The possibility for direct grounding is particularly obvious for symbols and expressions that can be the basis of statements about real-world facts. However, in the same way as one can concretely state that "This action was smooth," one may claim that "This action was democratic." Both assertions can be verified or falsified by applying established criteria (e.g., by assessing the movement trajectory or recounting votes). This shows that, even though abstract concepts have no direct correlate in the world, there are real-world "instances" of the concept, which can be relevant for learning the related symbol. Therefore, by pointing to concrete instantiations of democratic activities, one can teach the meaning of the related symbol, very similarly to the way one can teach concrete meanings by pointing to instances of concrete concepts.

Different Perspectives on Abstract Concepts

When aiming at a mechanistic explanation of concrete and abstract concepts and meanings, a clear explication of the major difference between the two is necessary. To this end, a range of different distinctive features are under discussion (for recent review and discussion, see Borghi et al., 2022). Compared with concrete ones, abstract meanings are viewed as relatively less grounded in action and perception, relatively more related to inner states and social interaction, more reliant on language and communication, and also more variable across texts and individuals. Furthermore, some authors describe structural differences between concrete and abstract concepts and semantics, with instances of concrete concepts sharing more object- and action-related semantic features than abstract ones (Langland-Hassan et al., 2021; Lupyan & Mirman, 2013; Sloutsky, 2010). Buccino et al. (2019) and Löhr (2022) argue that abstract concepts lack shared grounded semantic features entirely.

These approaches, which consider absolute or graded differences in the presence of shared semantic features as the core criterion for the concrete–abstract distinction, are reminiscent of a claim by Wittgenstein (1953), who, however, provided a most precise description of the meaning of some relatively abstract symbols. He argued that concepts such as GAME or ABILITY have instances that indeed do not share a core set of semantic properties but rather show partial similarities, as the faces of the members of a family might do. Family resemblance in Wittgenstein's sense thus means partial sharing of features across only a subset of category instances (Baker & Hacker, 2008; Wittgenstein, 1953).

This structural, feature-based approach led Pulvermüller (2013, 2018a, 2023) to propose that, whereas most concrete concepts and meanings can be characterized by shared category-defining action- and object-related semantic features, abstract entities are typically characterized by a lack of feature sharing but by family resemblance instead; and, indeed, a wide range of abstract terms (*beauty*, *democracy*, *safety*, *causation*, etc.) fulfill the family-resemblance criterion, whereas many concrete ones seem to be characterized by shared features (*hare*, *shirt*, *hammer*). Therefore, we believe that the structural distinction between full feature sharing and family resemblance (i.e., partial feature sharing) captures the difference between most concrete and abstract concepts more precisely than the mere claim about the presence, absence, or prominence of shared semantic features. In particular, the latter approach leaves open what "holds together" an abstract category, whereas the family resemblance idea clarifies this issue.

Operationalizing Concreteness Versus Abstractness as Feature Sharing Versus Family Resemblance

The proposed structural distinction between concrete and abstract concepts or meanings is illustrated in Figure 1. Although two extreme cases are illustrated, this structural distinction implies a continuum between full feature sharing and family resemblance, along which gradually more concrete or abstract concepts are situated. One example of such an intermediate semantic type is that of large categories of objects (e.g., BIRD), which include a majority of instances sharing all semantic features, although some may lack one or more of these otherwise shared properties (OSTRICH, PENGUIN). In our simulations of concrete and abstract conceptual and semantic processing and representation, this structural difference was implemented. Therefore, our results and their interpretation depend on the claimed parallelism between the contrasts of concreteness versus abstractness and of semantic feature sharing versus family resemblance.

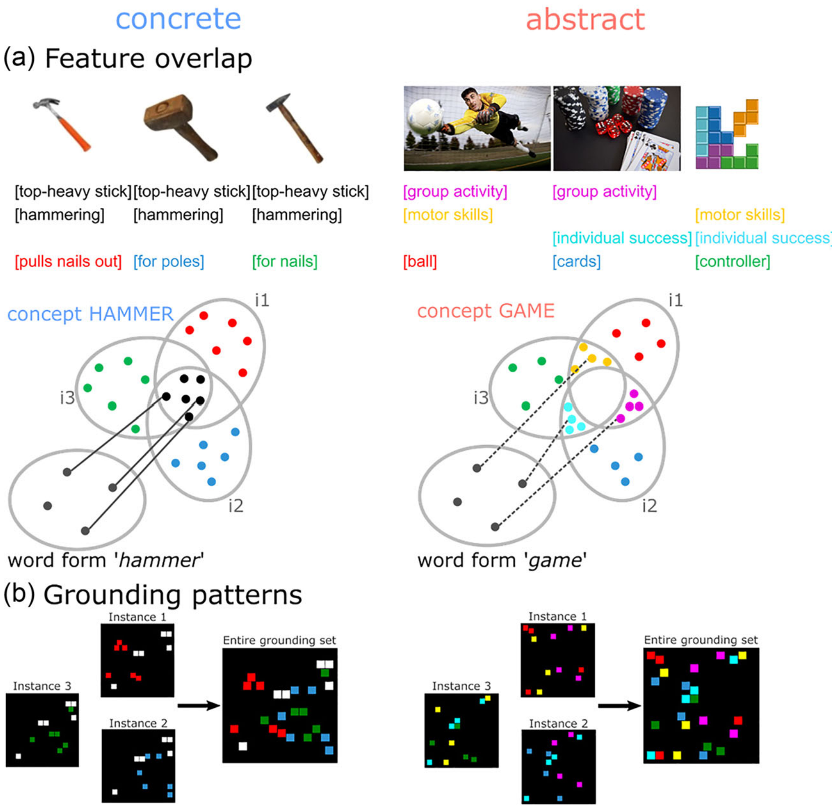

Figure 1 Illustrations of conceptual structure—full semantic feature overlap versus family resemblance—of relatively more concrete and abstract concepts and meanings. Panel A: Examples of overlapping and specific grounded experiential features of instances of concrete (left) and abstract (right) concepts. Below, a schematic illustration of the structural difference (full vs. partial semantic feature overlap) is depicted. Panel B: Specific grounding patterns used in this study. Colors indicate whether a neural element is part of one (blue, red, green), two (yellow, cyan, magenta), or all three instances (white).

We now illustrate the aforementioned structural difference between concrete and abstract semantics in terms of feature sharing versus family resemblance. We speak of "semantics" or "meaning" with reference to (verbal) symbols and of "concepts" when discussing language-independent representations.

The word *hammer* is used to refer to a class of objects. What hold this category together are action-related features (e.g., that all the objects are good for hammering) and also visual and haptic features (that all are top-heavy sticks). Still, hammers can vary widely in shape, size, material, color, or function. For example, one instance may be ideal for hammering nails into wood, a second may be better for advancing larger, heavy objects (e.g., poles), and a third exemplar may have an additional function (e.g., removing nails). These specific object- and action-related features are not general characteristics of the concept and are of little relevance semantically.

Now consider different instances of games, such as football, poker, and Tetris. Poker and football allow several people to partake in them, whereas Tetris is designed to be played by only one person. Success in football and Tetris strongly depends on motor skills, but success in poker does not. In poker and Tetris, the individual plays on their own, whereas football is a team sport. Note that none of the three features mentioned is shared by all instances of the semantic category, although one might rightly argue that they capture some of the "essence" of the meaning of *game* (Wittgenstein, 1953). One may try to defend the idea that the game concept is held together by shared features, for example, by claiming that all games are played for fun. But counterexamples come to mind immediately (professional gamers, billionaire football players, so-called Russian roulette). One may claim that after all, a common feature of games is that they are all governed by rules and conventions, which is correct, but much too unspecific to be helpful in characterizing the concept or word meaning (as it applies to almost all human activities). In essence, relevant experiential semantic features across all (or most) instances are missing; similar examples can be created for other abstract concepts too (BEAUTY, SAFETY, JUSTICE, etc.). Therefore, we claim that concrete concepts are typically characterized by semantic feature sharing and overlap, whereas, for most abstract ones, only partial feature sharing or family resemblance applies.

The structural difference between conceptual categories with feature sharing versus those with family resemblance is schematically illustrated in Figure 1A with the use of Venn diagrams. Each small dot in the Venn diagrams represents one specific perceptual or action-related feature; the overlapping ovals represent sets of features characterizing specific instances of a concept. Features shared between two or more instances of a concept are called semantic or conceptual features. Each instance also includes sensorimotor features not shared with other category members, which we call unique (or "idiosyncratic"). The set of all shared features of a category is called the semantic or conceptual representation.

Figure 1A depicts the full feature sharing of concrete concepts as well as the partial feature sharing of abstract concepts. This depiction can be used for explanation if the small dots are considered as neuronal elements, for example, local clusters of neurons. If these are connected to each other, and neurobiological principles of learning and synaptic modification are applied, predictions about the functionality of conceptual representation can be derived. Frequent activation of different instances (due, for example, to instance perception) will always lead to coactivation of all the shared conceptual neurons of concrete concepts. Thus, processing of the instances of a concept will build representations consisting of strongly interconnected shared neurons for concrete concepts. If a verbal symbol is perceived in conjunction with instance activations, its neuronal correlate will be bound to the conceptual representation, further strengthening it and building a semantic representation of symbol plus concept.

The situation for abstract concepts with family resemblance is different. The partially shared neuronal elements included in any instance representation will activate together only when this same instance is processed; during other instance experiences, these neuronal elements will activate in a disjoint manner, thus weakening their mutual links. A link between the disjoint neuronal subsets (shown in yellow, magenta, and cyan) can be built only through a different representation frequently coactivated with each of the subsets. A symbol regularly used together with different instances can serve this function of linking up with each of the partially shared subsets and thus binding together the abstract concept (for a more detailed explanation, see the Discussion section below and Henningsen-Schomers et al., 2023). This model predicts that concrete concepts can be learned preverbally by experience and are enhanced by label learning, whereas abstract concept formation requires concordant symbol learning. In the Discussion section below, we will relate these predictions to empirical and experimental research.

Aims and Strategies of the Current Study

The main aim of the present research was to investigate plausible mechanistic correlates of concrete and abstract concept formation in the human brain. To this end, we used explicit mathematically precise and biologically constrained deep neuronal networks. To approximate realistic neuronal mechanisms, we fashioned the networks under study according to known properties of the human brain (see Methods section; Pulvermüller et al., 2021). In particular, the models were governed by neurobiologically founded learning mechanisms and included artificial "neurons" ordered in different sets of model "areas," which correspond to areas relevant for language and concept processing in human cortex (Garagnani & Pulvermüller, 2016; R. Tomasello et al., 2017, 2018). These brain-like models were applied to simulate the learning of instances of concrete and abstract concepts in isolation and in the context of concept-specific symbols. The neuronal circuits that formed in the networks as a consequence of learning were mapped and interpreted, along with their activity dynamics.

A range of previous studies have used neural networks to study concept formation and semantic learning of concepts in linguistic and symbolic context (e.g., Cangelosi et al., 2000; Chen et al., 2017; Elman, 2004; Hoffman et al., 2018; Ito et al., 2022; Johnston & Fusi, 2023; Lupyan, 2012; Rogers & McClelland, 2004; Wermter, 2004; Westermann & Mareschal, 2014). However, these studies did not address the main question of our current research about mechanistic differences between concrete and abstract semantic processing. Some of these earlier neural networks were limited to simple examples of concrete concepts and were structurally quite basic (e.g., Lupyan, 2012), although other research suggests that more complex "deep" networks including several areas or layers are essential for abstract representations (Bengio et al., 2013; Ito et al., 2022). In the present work, we therefore used more complex networks simulating activity in several cortical areas known to be important for concept and language processing. A further reason to favor reasonably complex networks over simple ones was the main aim of our current work, to reveal putative brain mechanisms underlying human cognition. This main aim is achievable only by using networks that closely resemble, both structurally and functionally, relevant parts of the human brain. Note that most previous neural simulation studies used networks quite distant from human brain structure and function, favoring learning efficacy or standard connectionist architectures instead of brain constraints. Previous research with brain-constrained network models has shown them to be well suited for addressing mechanisms of word–meaning mapping (see Constant et al., 2023; Garagnani & Pulvermüller, 2016; R. Tomasello et al., 2017, 2018, 2019).

Two previous studies have already used brain-constrained deep neural networks to simulate the neurobiological mechanisms underlying concrete and abstract conceptual (Henningsen-Schomers & Pulvermüller, 2022) and semantic representations (Henningsen-Schomers et al., 2023). Here, we now focus on the activation dynamics of concept and symbol representations in brain-like networks. We address the main hypotheses (a) that functional representations can be learned preverbally for concrete but not for abstract concepts and (b) that concordant learning of symbols enhances the functionality of concrete conceptual representations but is necessary for building functional abstract concepts. Do the dynamics of networks mimicking relevant aspects of the human brain provide support for these claims?

Method

Following earlier modeling work (Constant et al., 2023; Henningsen-Schomers et al., 2023; Henningsen-Schomers & Pulvermüller, 2022; R. Tomasello et al., 2018, 2019), we modeled neuronal learning and brain activity using brain-constrained deep neural network models (see Pulvermüller et al., 2014, 2021; see Figure 2, Panels A and B, and Appendix S1 in the Supporting Information online for details). The model was implemented on the neural network simulation platform Felix (Wennekers, 2009).

Model Architecture

To make sure that the structure of the network resembles the structural features of the human cortex, a range of areas relevant for language and conceptual processing were implemented. The model consists of a total of 12 model areas intended to mimic inferior and dorsolateral frontal cortices, superior temporal and inferior temporal cortices, as well as occipital cortices. The six perisylvian areas correspond to cortical areas involved in auditory perception (A1, AB, PB) and articulatory planning and control ($M1_i$, PM_i, PF_i). The six extrasylvian areas correspond to the ventral visual stream (V1, TO, AT) and the hand motor system ($M1_L$, PM_L, PF_L;[2] see Figure 2, Panels A and B, for details). This allows for the simulation of nonlinguistic sensorimotor processes as well as the articulation and auditory perception of wordforms.

We also made sure that neuronal structure and connectivity within areas resembled those of cortical areas. Therefore, each model area or layer contains 625 excitatory and 625 inhibitory neurons (called *e*- and *i*-cells, respectively) with within-area connections. Further details about the model, such as how random connectivity or local and global inhibition are implemented, as well as relevant equations, are provided in Appendix S1 of the Supplementary Materials online.

Stimuli

To model the perception of an object (e.g., a hammer) and a typical action performed with the object (e.g., hammering), we activated neurons thought to code for perceptual visual features in area *V1, together with additional neurons in $^*M1_L$ related to hand motor features of the action. Although this type of simulation targets the acquisition of both perception- and action-related information, we speak about "stimulation" using patterns that index real-world instances. To model learning of a verbal wordform used to speak about a given instance (i.e., object and/or action), we activated the neuron sets in *V1 and $^*M1_L$ together with a set of neurons in $^*M1_i$ representing the articulatory

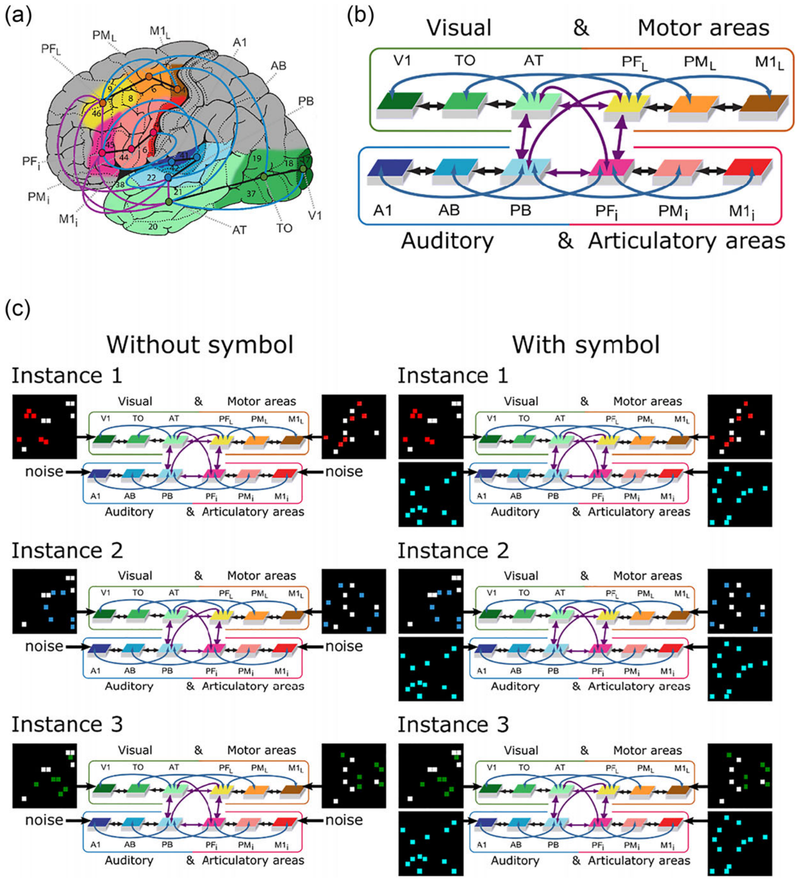

Figure 2 Large-scale network structure and simulation conditions. Panel A: Structure and connectivity of the neural network model. In total, 12 brain areas were modeled, including areas in frontal, temporal, and occipital cortex. Perisylvian areas comprise an inferior frontal articulatory system (red colors) and a superior temporal auditory system (blue colors), and extrasylvian areas comprise a lateral dorsal hand motor system (yellow/brown) and a visual "what" stream of object processing (green). Numbers refer to Brodmann areas, and the colored lines represent long distance cortico-cortical connections as documented by neuroanatomical studies. Panel B: Schematic depiction of the brain areas modeled (using the same coloring for different brain areas as in Panel A), along with their connectivity structure. Arrows indicate between-area connections. Panel C: Training regime used for concrete concepts and symbols. Colored dots indicate whether neural elements are related to instance-specific (blue/green/red), semantic/conceptual (white), or symbol form (cyan) information. For each concept, three instance-related grounding patterns were presented separately to the network in random order. For symbol learning, a wordform pattern was presented together with one of the instance-related patterns of the related concept.

features of a spoken word, plus a set of neurons in *A1 thought to stand for the acoustic phonological features of the same wordform. In all of the simulations, each instance- or wordform-related stimulation pattern included 12 neurons in each relevant primary area (either *V1 and *$M1_L$ or *A1 and *$M1_i$), with each neuron coding specific visuomotor, or phonological features. Examples of specific patterns are shown in Figure 1B, and all patterns used are listed online on the Open Science Framework platform (https://osf.io/m8dg5).

Three instances of a given concept were modeled (see Introduction for a detailed discussion). An instance-related activation pattern includes a pair of neuron stimulation patterns in the sensorimotor extrasylvian primary areas *V1 and *$M1_L$. Multiple instance patterns of the same concept share a subset of their stimulated neurons through either full overlap or pairwise overlap. We implemented the instances of concrete and abstract concepts by way of complete mutual sharing of conceptual feature neurons or pairwise and thus only partial sharing of features across the instance patterns, respectively (Figure 1). We call all neurons shared between two or more instances “conceptual” or “semantic,” as their sharedness is what makes the instances members of the concept or category. We accounted for the discrepancy in the numbers of unique and conceptual neurons between concrete and abstract grounding patterns in the analysis of our results.

In total, 10 concepts each consisting of three grounding patterns (thus overall 30 patterns) containing 12 neurons each were constructed according to the constraints described above. Unique neurons were instance-specific, that is, no unique neuron was used in simulations of instances of more than one concept, to ensure that there was no randomly varying overlap between concepts, which could have confounded the results; instead, we kept between-concept overlap at a constant 0.

Spoken wordforms were implemented as correlated neuron stimulation patterns in the perisylvian primary areas *A1 and *$M1_i$. These wordform-related grounding patterns consisted of 12 neurons each. Ten nonoverlapping wordform patterns were created, corresponding to the 10 concepts to be mapped. In different simulation conditions, abstract and concrete concepts were associated with the same 10 wordform patterns to control for potential wordform-induced confounds.

Two types of models were trained: one with associated verbal symbols and one without (symbol and no-symbol conditions). A model without symbols was thought to imitate pre- or nonverbal concept learning based on the similarity of perceptuo-motor features and therefore received only extrasylvian input during training. In contrast, a model with symbols mimicked the co-occurrence of (a) simultaneously articulating and (b) hearing a correlated wordform while (c) perceiving visually similar stimuli and (d) activating similar action representations (see Figure 2C). Thus, one can interpret the models in the no-symbol condition as being trained by nonlinguistic experiences in the world, whereas the symbol condition implements concordant nonlinguistic and linguistic experiences.

Training Procedures

As described in Henningsen-Schomers et al. (2023), training of models was conducted in a 2 × 2 factorial design, with the variables semantic type (concrete vs. abstract grounding patterns) and symbol learning (conceptual grounding patterns copresented with or without a wordform pattern). In each of the four conditions, 12 models, each with different random initiations of connections and weights, were trained on the stimuli; these 12 models were the same across conditions, thus yielding 48 models in total. The 12 base models were used to account for variation that would be present across human participants engaged in concept and symbol learning. The split between conditions was necessary to avoid interference between the conceptual grounding patterns. Note that previous research has already shown that the neural circuits forming for concrete and abstract concepts and symbols differ in their topography and robustness (Henningsen-Schomers & Pulvermüller, 2022), thus making it likely that one may suppress or modify the other during learning. Note furthermore that, in language learning, abstract meanings are acquired later than concrete ones, which provides further motivation for implementing partly separate mechanisms.

For models trained without a symbol, a training trial consisted of presenting a randomly chosen sensorimotor grounding pattern to the extrasylvian primary areas *V1 and $^*M1_L$ for 16 timesteps. The perisylvian primary areas *A1 and $^*M1_i$ experienced uncorrelated white noise during the training time. For models with symbol learning, the perisylvian primary areas instead received the activation pattern of the wordform of the concept to which the sensorimotor grounding pattern belongs (see Figure 2).

Finally, a wordform control model was trained, which received input only to *A1 and $^*M1_i$, and white noise to the extrasylvian areas. A wordform representation learned by this model is not grounded in sensorimotor experience and encompasses all neurons that represent nongrounded wordform features. These wordform control neurons were excluded from data analysis, to prevent an influence of purely wordform-induced activation on the semantic activity patterns, which our analysis targets.

Between stimulus presentations, primary model areas received uncorrelated white noise activity. The duration of the interstimulus interval was determined by the levels of global inhibition in *A1 and *PB, which had to be below a threshold of 0.75 or 0.65, respectively, for the next trial to be initiated. This prevented persistent neuronal activity of a given trial from influencing subsequent ones. In line with previous research, training continued until each instance had been repeated 2,000 times (Henningsen-Schomers & Pulvermüller, 2022).

Testing Procedures

We asked whether the perception of a conceptual instance activates conceptual or semantic representations. To assess this, we imitated instance experience, by stimulating model areas *V1 and *$M1_L$ with the previously learned grounding patterns. During testing, wordform patterns were not directly stimulated so that word-related representations in perisylvian cortex could only become activated indirectly through their related conceptual instances. The synaptic weights of the models were fixed during testing, so that no further learning could occur. Throughout testing, white noise background activity (noise parameter = 8) was provided, but the unstimulated perisylvian primary areas *A1 and *$M1_i$ did not receive additional uncorrelated input. In separate testing trials, each of the 30 trained grounding patterns was presented for two timesteps, followed by 28 poststimulation timesteps, for a total of 30 timesteps. Model activity was quantified as the number of spikes across all neurons of an area per timestep. To establish a baseline of network activity, intrinsic network activity was measured for 10 timesteps prior to stimulation. To reset model activity prior to the next stimulus presentation, the membrane potential of all neurons was set to 0.

To assess the acquired neuronal representation of a conceptual instance, we recorded the activity of all neurons of the network following stimulation of each instance-specific grounding pattern and classified responsive neurons as active if their dynamic estimated firing rate reached 75% of the maximal firing rate in their area for at least two timesteps. In addition, we classified each active neuron according to whether it was activated by only one single instance-related pattern (“instance-specific neurons”) or by two or all of a concept’s instances (“conceptual/semantic neurons”).

For all statistical tests, we first calculated average values of the 30 instances for each of 12 networks, which were thought to simulate processes in different simulated “subjects.” Then, these “subject-specific” results were averaged across networks, and statistical tests were applied (for details, see Appendix S2 in the Supporting Information online).

Results

To unravel the influence of verbal symbols on concept and meaning processing, we assessed two measures of neuronal dynamics. The first is the magnitude of peak activity, which describes the number of neurons within the instance cell assemblies (CAs) that are simultaneously active across all extrasylvian areas at the timestep of maximal activity. We define the point of maximal activity as the "ignition" of the CA (cf. Braitenberg, 1978). Larger values may indicate larger and/or more strongly connected neuronal sets. The second measure is the working memory period or the duration of CA reverberation. It denotes how long significant activity is maintained within the CA after peak activity within a given area. In accordance with previous research (Schomers et al., 2017), it is defined as the number of consecutive timesteps after the activation peak during which the number of CA spikes exceeds the average number of spikes in the prestimulation period by two or more standard deviations.

Magnitude of Cell Assembly Ignition

A 2 × 2 ANOVA on the number of spikes at t_{max} (the timestep of maximal activity) with the variables semantic type (concrete vs. abstract) and model type (no symbol vs. symbol) revealed significant main effects for both variables as well as for the interaction between them: semantic type, $F(1, 11) = 12.11$, $p < .0001$, $\eta_p^2 = .52$, 90% CI [.13,.70]; model type, $F(1, 11) = 631.30$, $p < .0001$, $\eta_p^2 = .98$, 90% CI [.95,.99]; interaction, $F(1, 11) = 183.87$, $p < .0001$, $\eta_p^2 = .94$, 90% CI [.85,.96].

As the interaction was significant, we conducted post hoc paired t tests comparing the effect of semantic type on concepts and symbols, and the effect of associating a symbol by concept type. The results were Bonferroni corrected for four comparisons (critical $p = .0125$).

The difference between concrete (M = 72.41, SD = 0.91) and abstract concepts (M = 70.03, SD = 0.64) was significant ($t(11) = -6.18$, $p < .001$, Cohen's $d = 1.78$, 95% CI [1.38, 3.09]), as was that between concrete (M = 75.70, SD = 1.25) and abstract symbols (M = 81.91, SD = 2.05) ($t(11) = 7.64$, $p < .001$, Cohen's $d = 2.21$, 95% CI [1.42, 5.22]). The effect of associating a wordform with a concept during learning was significant for both concrete concepts (no symbol: M = 72.41, SD = 0.91, symbol: M = 75.70, SD = 1.25; $t(11) = 9.60$, $p < .001$, Cohen's $d = 2.77$, 95% CI [2.02, 5.24]), and abstract concepts (no symbol: M = 70.03, SD = 0.64, symbol: M = 81.91, SD = 2.05; $t(11) = 23.08$, $p < .001$, Cohen's $d = 6.66$, 95% CI [5.37, 11.96]).

All effect sizes, as indicated by Cohen's d, were large. All differences were significant when restricting analyses to unique or to conceptual neurons

($p < .001$). Note, however, that without symbol involvement, opposing effects were present for unique-instance-related neurons (larger numbers for concrete than for abstract concepts) and for shared conceptual ones (smaller numbers for concrete than for abstract concepts; see Figure 3C); congruent effects were seen after symbol-related semantic learning.

Duration of Reverberation: Working Memory Period

We conducted a 2 × 2 ANOVA on the duration of significant CA activity with the variables semantic type (concrete vs. abstract) and model type (no symbol vs. symbol). It revealed a significant main effect of model type, $F(1, 11) = 5{,}297.37$, $p < .0001$, $\eta_p^2 = 1.00$, 90% CI [.99,1.00], and of semantic type, $F(1,11) = 47.49$, $p < .0001$, $\eta_p^2 = .81$, 90% CI [.54,.88]. The interaction Model Type × Semantic Type was also significant, $F(1, 11) = 743.16$, $p < .0001$, $\eta_p^2 = .99$, 90% CI [.96,.99]. We conducted paired post hoc t tests with Bonferroni correction for four multiple comparisons (critical $p = .0125$). The working memory period was substantially longer for concrete (M = 10.63, SD = 0.77) than for abstract concepts (M = 4.92, SD = 0.67), $t(11) = -14.96$, $p < .001$, Cohen's $d = -4.32$, 95% CI [−6.29, −3.59]; and a reverse difference was seen for concrete (M = 11.75, SD = 0.58) and abstract symbols (M = 13.25, SD = 0.40), $t(11) = 5.45$, $p < .001$, Cohen's $d = 1.57$, 95% CI [0.94, 3.09]; which, however, may be due in part to the relatively larger number of conceptual neurons of the latter. The effect of learning a symbol was significant for both concrete concepts (no symbol: M = 10.63, SD = 0.77, symbol: M = 11.75, SD = 0.58; $t(11) = 7.39$, $p < .0001$, Cohen's $d = 2.13$, 95% CI [1.37, 4.69]), and abstract concepts (no symbol: M = 4 .92, SD = 0.67, symbol: M = 13.25, SD = 0.40; $t(11) = 58.63$, $p < .001$, $\beta = 8.33$, 95% CI [8.02, 8.65], Cohen's $d = 16.93$, 95% CI [13.54, 25.30]); however, it was much larger for abstract concepts than for concrete ones (see Figure 3B). All effect sizes were large.

Discussion

We used a brain-constrained network model of human cortical areas essential for processing language and concepts, as well as areas involved in perceptions and actions relevant for conceptual and semantic grounding, to investigate the putative neuronal consequences of experiencing different instances of concrete and abstract concepts in isolation and in the context of verbal symbols. We asked whether both concrete and abstract concepts would thereby form and be grounded in conceptual instances of objects and actions. Concreteness versus abstractness of a category was operationally defined by a structural difference

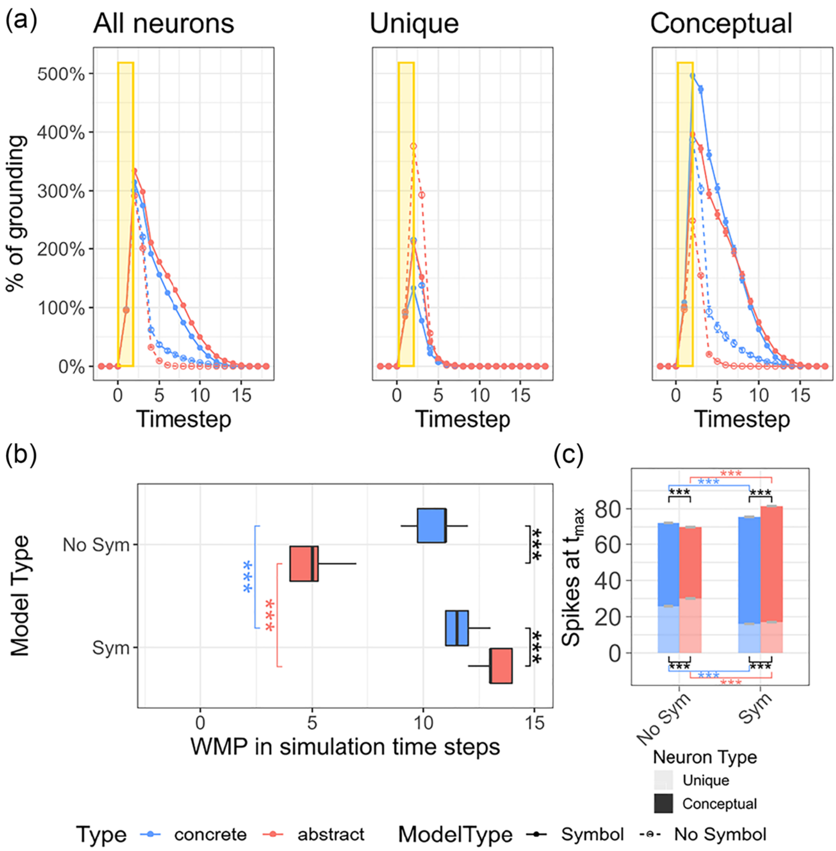

Figure 3 Temporal activity dynamics elicited by instances of concrete and abstract concepts after learning out of and within symbol context. Panel A: Activity dynamics of an instance-related cell assembly (CA), for either all neurons in the CA (left panel), only unique neurons (middle panel), or only conceptual neurons (right panel). The stimulation period is marked in gold. Depicted is the number of spikes per timestep, normalized for the number of neurons in the grounding pattern. Whereas unique neurons contribute to early stages of CA activity, including its ignition, sustained activity is primarily carried by conceptual neurons. Panel B: Reverberation time or working memory period (WMP) for concrete and abstract concepts and symbols. Without a symbol, abstract concepts barely show any prolonged activity or reverberation, whereas concrete concepts do. Abstract concepts benefit the most from learning concepts in context of symbols and become functionally similar to concrete ones. Panel C: The number of spikes at ignition. Through the addition of a symbol, the raw number of neurons involved in ignition increases. Additionally, relatively more conceptual than unique neurons contribute to ignition for both types of symbol.

(see Introduction and Figure 1): The instances of concrete concepts had fully shared experiential features, whereas those of abstract ones were only partially shared between some of their instances, thus showing a pattern of family resemblance. Experiential information about objects and actions was modeled by sets of neurons concordantly activated in visual and dorsal motor areas of the model (Figure 2C, left-hand side). To simulate learning in symbol context, additional articulatory and auditory verbal phonological information was implemented by concordant neuronal activation patterns in articulatory and auditory model areas (Figure 2C, right-hand side).

Biologically plausible unsupervised Hebbian learning was applied. We found that, after learning of conceptual instances, large sets of model neurons (~86 per concrete instance, ~81 per abstract instance) widely distributed across the network reliably activated to each instance's presentation (see Figure S2 in Appendix S2 of the Supporting Information online), and that the dynamics of these neuronal sets were dominated by the "semantic neurons" representing shared features of their conceptual instances. For concrete concepts, these shared semantic neurons produced substantially stronger activity peaks (ignitions) and stayed active substantially longer (reverberation) than those of abstract concepts. Concordant verbal symbol presentation during instance learning substantially increased set sizes (~136 per concrete instance, ~142 per abstract instance), activation peaks, and reverberation times for both concept types. The observed dynamics of reverberation times, a measure related to working memory, suggest that, within a brain-constrained neural architecture, maintainable abstract conceptual engrams (memory traces) are only formed if verbal symbols are available for labeling them.

Simulation Strategy and Evaluation

Concrete and abstract concepts were both modeled by activation patterns thought to relate to concrete real-world instances of concepts. When presenting different instances of each concept, it was assumed that the network would extract the common features of conceptual instances and build conceptual representations based on biological learning of feature correlations (see Pulvermüller, 2002, 2018c; Sloutsky, 2010; Sloutsky & Robinson, 2008).

In line with brain theory and neurocomputational research, the mechanistic neurobiological basis of a representation or engram may be a set of strongly connected neurons that, due to their strong mutual links, function as a unit. In different brain theory frameworks, these are called cell assemblies (CAs), neuronal assemblies, groups or circuits, or cognits (Braitenberg, 1978; Deco et al., 2013; Fuster & Bressler, 2012; Hebb, 1949; Milner, 1957; Palm, 1982;

Pulvermüller & Garagnani, 2014; Zipser et al., 1993). These CAs can be defined structurally, based on their relatively strong internal connections compared with average neuronal connectivity in a network, or functionally, by way of activation criteria. Two functional criteria for assessing CAs are that they strongly activate to stimulation, a process sometimes called "ignition," and that they retain activity for some time after stimulation, due to "reverberation" of neuronal signals between their neurons (Braitenberg, 1978; Hebb, 1949; Palm, 1982). CA ignition can be interpreted as a mechanistic correlate of the recognition of objects, actions, or symbols, whereas CA reverberation at an activity stage significantly above rest is seen as the biological basis of working memory (Fuster & Bressler, 2012; Schomers et al., 2017; Zipser et al., 1993). Ignition and reverberation may also reflect the processing of neuronal conceptual, linguistic, and semantic representations. In the sections below, we follow Henningsen-Schomers and colleagues (2022, 2023) in discussing the structure of neuronal representations that formed in the four conditions—concrete and abstract concept formation without and with concordant symbols. However, we extend their previous work by analyses of dynamic aspects, focusing on measures of ignition and reverberation.

Neuronal Correlates of Concrete and Abstract Concept Formation

After learning of conceptual instances outside symbol context, the simulated experiences of both concrete and abstract conceptual instances led to reliable and reproducible activations, which overlapped between instances. Henningsen-Schomers and Pulvermüller (2022) remarked that most of the activated neurons were shared between two or more instance activations, and that this predominance over instance-specific ones was particularly strong for concrete conceptual instances. The relatively greater dominance of shared semantic neurons suggests more robust representations for concrete than for abstract concepts. These authors described the distribution of concrete neurons across the model's extrasylvian areas as exhibiting a "belly shape"—with more neurons in the central areas (*PF_L, *AT) than in primary ones (*$M1_L$, *V1)— which contrasted with the "slim" figure of the shared-neuron sets of abstract concepts. They also speculated that this structural difference may have functional consequences, as a "big belly" with many neurons in central areas allows for easy activation across a neuronal circuit, whereas a "slim belly" with only few neurons bridging between primary "peripheral" cortices may make such circuit-internal processing more difficult.

Considering the functional activation data obtained from our brain-constrained models, we can resolve these issues: The results for ignition dynamics show clearly that overall activation peaks did not substantially differ between conditions after instance-based concrete and abstract concept learning without symbols (Figure 3A). However, focusing on shared neurons only and therefore the network correlates of conceptual representations, one can see that there are substantially greater ignition amplitudes for the neuronal correlates of concrete concepts than for those of abstract concepts (Figure 3C). There is an opposite pattern for instance-specific neurons, which activate relatively more strongly in the abstract concept condition. This pattern is, once again, consistent with stronger representations for concrete than for abstract concepts learned outside symbolic context. In contrast, individual instance representations seem to be better implemented and more strongly activated after abstract compared with concrete conceptual learning.

Peak activation data suggest, but do not unambiguously prove, that circuit representations of, or CAs for, concepts have formed. The answer is, however, provided by the second functional measure, reverberation times, which are significantly and substantially longer for concrete than for abstract concepts learned outside symbol context (Figure 3, Panels A and B). Numerically, periods of reverberatory activity are more than twice as long on average for concrete as compared with abstract concepts (5.0 vs. 12.5 simulation timesteps). Still, this difference does not strongly argue for or against the presence of CAs for abstract concepts. The key argument is provided if these numbers are related to network structure: Stimulation lasts for two timesteps, and after this, there is free bidirectional activation spreading across the 12 areas of the network. There is a minimum of three synaptic steps between any two primary areas, so that spreading activation accounts for at least five timesteps, three for spreading plus two due to stimulation. Strikingly, the abstract learning conditions did not yield significant reverberant activation after five timesteps. The five timesteps of activation brought about by instances of abstract concepts can therefore be explained without any actively maintained reverberatory activity. After five steps of feedforward activity flow, the conceptual neurons of abstract concepts became inactive, whereas activity persisted for concrete conceptual neurons. As the second functional criterion for CA activation, that is, presence of reverberation, is met after concrete but not abstract conceptual learning, we conclude that the brain-constrained network builds neuronal memory representations, engrams, or CAs for concrete concepts outside language context, but not for abstract ones.

This result is in agreement with empirical results on concept learning in humans: Concrete symbols tend to be acquired faster than abstract ones (Brown, 1957; Gentner, 1982; Schwanenflugel & Akin, 1994) and can be acquired completely nonverbally, not only by human infants (Mandler, 2004; Perszyk & Waxman, 2018), but even by other species, including nonhuman primates and birds (Pearce, 2008; Pusch et al., 2023). Our study also allows us to specify the feature that is critical to an explanation of this difference: The full semantic feature overlap of all (or at least most) members of concrete categories may be necessary for the formation of conceptual representations outside symbol context; partial overlap and family resemblance is not sufficient.

Causal Influence of Language on Concept Formation

Once symbols are learned together with concept instantiations, the picture described above changes substantially. Even though we controlled for the additional neuronal material and activations directly caused by symbol-related coactivation by excluding neurons activated in the wordform-only control condition, the size of activated neuron sets, and in particular that of shared semantic neurons, increased significantly for both types of concepts. Likewise, symbol addition led to more substantial and more prolonged activation, as shown by peak activation and reverberation measures (e.g., Figure 3, Panels A and B). This is clear evidence for a facilitatory causal effect of language on conceptual processing in brain-constrained deep neural networks, which is in line with classic and current "Whorfian" theories of linguistic relativity (Athanasopoulos & Casaponsa, 2020; Bohnemeyer, 2021; Gumperz & Levinson, 1991; Lupyan et al., 2020; Maier & Abdel Rahman, 2019; Majid et al., 2004; Miller et al., 2018; Thierry, 2016; Whorf & Carroll, 1976). Please note that our simulation procedures imitating brain processes of perceiving and experiencing objects and actions in the world remained unchanged across simulations and learning conditions, and that the influence of linguistic-symbolic learning documented in the simulations must therefore be present at the neurocognitive level of stored conceptual and symbolic representations that emerged in the brain-constrained networks as a result of correlation-driven learning.

The causal effect of language on concept formation was particularly pronounced for abstract concepts. After instantiations of abstract concepts had been copresented with symbols, they elicited reliable, strong, and prolonged activations comparable with those of concrete concept instantiations bound to a symbol. Note, furthermore, that activation magnitude and duration were carried by shared semantic neurons for both concept types, and that the neurons specific to instances did not contribute to prolonged reverberation (see Figure 3, Panel B).

In summary, learning a symbol benefits the ignition magnitude and reverberation time of both concrete and abstract concepts in a brain-constrained neural network. This effect is stronger for abstract than for concrete concepts, as, with learning outside symbol contexts, there is no clear evidence for abstract concept representations having formed. As the only difference between concept and symbolic semantic learning is the absence or presence of a co-occurring wordform, we conclude that this additional linguistic information is causal for the changes in neuronal dynamics. It is notable that this causal effect is intrinsically linked to higher within-concept similarity of instance representations and lower between-concept similarity, implying stronger "Whorfian" effects for abstract than for concrete concepts (Henningsen-Schomers et al., 2023). Our present work now shows that there are not only gradual symbol-learning effects of different sizes on the formation of concrete and abstract semantic mechanisms. In our simulations, the networks did not build abstract conceptual representations in the absence of symbol information; copresentation of symbols with family-resemblant instances was necessary to enable the formation of abstract conceptual-semantic representations. This is qualitatively different from the symbol-related solidification of concrete concepts, as abstract concepts could only emerge in the presence of language. These results constitute neurocomputational support for strong versions of linguistic relativity according to which language is not only a weak facilitator of thought but a conditio sine qua non for some (abstract) cognitive processes (for discussion, see Henningsen-Schomers & Pulvermüller, 2022; Pulvermüller, 2013, 2023).

The neurocomputational model of concrete and abstract symbol processing can be related to experimental findings. In regard to neuroimaging results for concrete word processing, many studies have shown particularly strong activation of sensory and dorsal motor areas processing perception- and action-related features that are relevant for semantics. For example, action-, sound- or vision-related words (e.g., *grasp*, *thunder*, *zebra*) respectively produce particularly strong activation in lateral sensorimotor, auditory, and visual areas (Binder & Desai, 2011; Kiefer & Pulvermüller, 2012; Pulvermüller, 2018b). Such category-preferential activation patterns were successfully simulated with the 12-area semantic model (see, e.g., Garagnani & Pulvermüller, 2016; R. Tomasello et al., 2017, 2018). The current simulations did not distinguish between semantic categories of concrete words, but simulated concrete symbols as semantically related to both action and perception, thus allowing for predictions on brain activity elicited by concrete words generally. Compared with model activity for these concrete items, abstract words led to relatively stronger activation, particularly in connector hub areas, including inferior and lateral prefrontal areas (see Appendix S2 in the Supporting Information online, Figure S4A, panels AT and PF_L). This finding is consistent with, and provides a putative explanation for, the enhanced prefrontal activation of abstract as compared with concrete words observed in previous neuroimaging work (Binder et al., 2005). However, the match between these experimental and simulation results is not perfect, as anterior temporal activation differences predicted by the model are not reported in the aforementioned experimental work. Future modeling work may fruitfully explore further perspectives on explaining preexistent and predicting future experimental findings.

Neuronal Mechanisms of Concept Formation

A reason why it is difficult to form stable representations for abstract concepts lies in the correlation strength across semantic neurons (Pulvermüller, 2013). Let us reconsider the specific case of our present simulation parameters to illustrate this: For concrete concepts, all six of the shared conceptual neurons coactivate whenever an instance is experienced; these neurons' activations are therefore highly correlated. Through Hebbian learning, these neurons strengthen their mutual synaptic links, thus forming sets of neurons that are likely to act as a conceptual CA. As activity propagates from primary to more central areas of the model, these frequently active shared semantic neurons will be most efficient in recruiting additional neurons with similar activation specificity. This accounts for the development of a "belly shape" across areas (cf. Henningsen-Schomers & Pulvermüller, 2022). For abstract concepts, however, only two thirds of conceptual neurons (in our example, eight out of 12) activate to a given instance. If the other instances appear, half of these neurons (four out of eight) will activate independently of the other half, so that the three subsets of four semantic neurons (shown in cyan, magenta, and yellow in Figure 1) exhibit quite a weak correlation; as a consequence, no CA representation of an abstract concept forms outside of a symbol context (see Introduction, prediction (a)).

This situation changes significantly due to association with a symbol, with consequences for the neuronal representations of concrete and abstract symbols. For concrete concepts, this symbol copresentation with conceptual instances increases the set of frequently coactivated neurons and, therefore, the size of the CA of the shared neuronal set and conceptual-semantic representations. This results in moderate increases in CA size, ignition amplitude, and reverberation time. Even after removing the neurons directly activated by learned meaningless wordforms, this difference persisted; the increase may therefore be due to additionally recruited neurons, not by ones directly activated by the symbol. In contrast, during copresentation of a symbol with instances of an abstract concept, a more substantial change is brought about. Each of the (in our example, three) subsets of semantic neurons is now active in a larger fraction (here two thirds) of the cases of symbol-instance coprocessing, which enables the formation of an abstract semantic representation binding symbol-related and shared-instance-related information. In this perspective, the reason why abstract concepts can form in symbol context but not outside it lies in the increase in correlation between the partaking neurons, which is provided by copresent language units (for further explanation, see Pulvermüller, 2023).

In summary, the coactivation of neurons in the perisylvian cortex with shared semantic neurons in extrasylvian space provides increased correlation values that help in binding together the disparate subsets of partially shared semantic neurons of abstract concepts with family resemblance structure. The low-correlation conceptual neurons active to abstract conceptual instances can bind to the "mediating" linguistic neurons because these coactivate more frequently than the subsets of partially shared neurons. This enables the formation of semantic representations for abstract symbols-plus-concepts, which resemble those of concrete concepts. These effects emerged from the implementation of neurobiological and neuroanatomical principles in a brain-constrained neural network model (see prediction (b)). As these are general principles of Hebbian learning, they do not apply only to category formation through semantic similarity and category terms, but might also underlie behavioral effects of category terms versus proper names on attention modulation and memory performance (see, e.g., Althaus & Mareschal, 2014; LaTourrette & Waxman, 2020; Nguyen et al., 2024).

Limitations and Future Directions

One may argue that the present simulation study is limited as it relies on the assumption that conceptual and semantic learning start without any explicit preprogrammed knowledge, almost at a tabula rasa state, and that both concepts and their related verbal symbols are learned from scratch. Therefore, one may suggest that our approach cannot capture nativist theories of abstract conceptual representations assuming presence of concepts a priori, before and independent of any learning (Fodor, 1975, 2008; Pinker, 2008). However, as we here show how the emergence of representations of abstract concepts can be explained by neurobiological principles in a neural architecture with relevant similarity to the human brain and based on the similarity structure of perceptions and actions along with their symbol contexts, the extreme position of conceptual nativism may be seen to fall victim to Occam's razor. The present results suggest that the strong assumption of a priori concepts, which is immanent to nativism, is neither necessary nor motivated.

We note that quite a bit of preprogrammed knowledge is implemented in the structure of the brain-constrained networks that we applied, for example in the connectivity structure resembling links between human cortical areas. Schomers et al. (2017) have shown how such inborn structural information may determine cognition and language, most importantly the specifically human capacity for verbal working memory and for building large vocabularies. As there is a clear link between structural anatomical information immanent to our brains and the knowledge we can acquire, the impression of a tabula rasa model must be revised.

One may argue that the model we use here is rather complex, and that similar results may be obtainable with much simpler architectures and models. In this context, some related earlier models may be considered, ranging from models composed of McCulloch–Pitts neurons to ones implementing areas as "hidden layers" of a fully distributed parallel distributed processing architecture or as Kohonen maps (Chen et al., 2017; Pulvermüller & Preißl, 1991; Wermter, 2004). However, as these models come with assumptions with questionable biological foundation (e.g., learning depending on error-gradients, winner-takes-all dynamics, lack of connections within each area, nondiscreteness of representations), it may be difficult to find, within these models, clear candidates for biological counterparts of concrete and abstract concepts and in particular their distinctive neurocomputational features. We argue that brain-constrained modeling is necessary to achieve this goal (Pulvermüller et al., 2021). As already mentioned in the Introduction, our main reasons for choosing a relatively complex model are twofold. First, earlier work suggests that relatively complex multilayer networks are most suitable for representing abstract concepts or features (Bengio et al., 2013; Ito et al., 2022). Second, we aim at achieving insights into the workings of the brain. To mimic its processes in a variety of areas relevant for concepts and language, it is necessary to implement at least a representative subset of these areas and of their connectivity, internal structure, dynamics, and learning processes. This is impossible without a degree of complexity.

A major caveat concerning the present simulations is the simple and to a degree artificial nature of the learning examples. We created overlapping sets thought to stand for individual instances of a concept, with activated neuronal units representing specific experiential features. Only three instances per concept and only two sets of 12 neuronal units per conceptual instance pattern were chosen, with emergent conceptual and semantic representations including about 80–140 artificial neurons. It is desirable to complement this work with more realistic examples, taking into account specific concepts, symbols, and features relevant for semantic grounding. This will require addressing more complex feature compositions of concepts, which could include prototypes of given categories, variability of feature overlap patterns, and example categories in between fully concrete and extremely abstract ones (e.g., the aforementioned large-category terms). Implementing the continuum of fully shared and pairwise shared features (Rosch & Lloyd, 1978) could be the target of fruitful work in the future.

However, we also wish to mention an advantage of the relatively abstract and structurally oriented approach chosen here: By using a straightforward structural difference between conceptual types controlled for number of neurons and features and by choosing a number of neurons still open to meticulous analysis, it becomes possible to trace neurofunctional and neurostructural differences between representations and to uniquely attribute them to the specific cognitive-structural differences implemented.

Additionally, one may argue that the training regime used in this study is rather primitive compared with human developmental trajectories of concept acquisition. We address this concern in Appendix S3 in the Supporting Information online.

One may also claim that there are relevant features of abstract concepts (see Banks & Connell, 2023; Borghi et al., 2022; Dove, 2021; Pexman et al., 2023; Villani et al., 2021) that our model does not capture. For example, that abstract concepts relate more to emotions, mental states, or complex interaction schemes than concrete ones is not implemented (but see Introduction for discussion). A useful strategy to overcome this limitation is to explicitly model a wider range of subtypes of abstract concepts and symbols semantically related, for example, relatively more to singular or group actions, to objects and actions of different types, or to emotions or mental states (see Dreyer & Pulvermüller, 2018; Harpaintner et al., 2020).

The structural difference between concrete and abstract concepts implemented here, feature sharing versus family resemblance, may help explain several of the observations reported to distinguish between these categories. That abstract symbols, in contrast to concrete ones, appear in relatively more variable contexts and carry more different meanings in these variable contexts (Barsalou & Wiemer-Hastings, 2005; Borghi et al., 2022) can be seen as a direct manifestation of their family resemblance property, which implies that they are used to speak about quite different "things." The greater variability across contexts comes with a relatively low probability of a symbol occurring in each specific context, so that the concept is less associated with each of its contexts. The on-average later age of acquisition of abstract words is also naturally explained by the lower feature correlations and therefore relatively longer time required for building semantic links across the family resemblance patterns of abstract symbols. Likewise, the lower "perceptual strength" of real-world associations fits well with the feature variability immanent to family resemblance. And finally, of course, the present distinction captures the earlier claim that abstract concepts share fewer features than concrete ones (Langland-Hassan et al., 2021; Löhr, 2022; Sloutsky, 2010)—with the additions that family resemblance structure specifies (a) in what way reduced or absent feature sharing is realized and (b) what underlies the coherence of abstract concepts, providing a greater degree of precision than a mere statement about presence, dominance, or absence of shared features. It appears that the feature sharing versus family resemblance distinction sits well with a range of known differences between concrete and abstract concepts previously highlighted in the cognitive and linguistic literature and may in fact provide an explicit operational avenue toward explaining or capturing several of these.

Conclusion

We conducted a brain-constrained neurocomputation simulation study to explore the putative brain mechanisms of concept formation and form–meaning mapping for concrete and abstract symbols, as well as their impact on conceptual and semantic processing. The formation of conceptual representations in the brain-like networks was manifest in the formation of cell assembly circuits that ignited and reverberated. We found that the model's processing of familiar concrete and abstract concepts was modulated by whether they had been acquired in the context of an associated verbal symbol. Without a symbol, the model failed to form reliable, strong, and stable representations for abstract concepts, but succeeded for concrete ones. With a symbol, the model succeeded for both types of concepts, which demonstrates a causal effect of language context on concept formation, as claimed by strong theories of linguistic relativity. While, in the neurocomputational model, a conceptual-symbolic link was beneficial for both conceptual types, abstract concepts required an associated symbol to develop stable conceptual representations. Thus, the learning of linguistic symbols might be crucial for the human ability to learn concepts such as JUSTICE, PEACE, GAME, or CAUSALITY.

Final revised version accepted 21 March 2024

Acknowledgments

Open access funding enabled and organized by Projekt DEAL.

Open Research Badges

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data are available at https://osf.io/m8dg5.

Notes

1 In this work, we tend to use the term "symbol" or "wordform" instead of "(verbal) label." In the context of semantics, the word "symbol" implies a semantic link between a wordform and its semantics. Learners induce such semantic links from copresented information about wordforms and possible referents. The term "label" presupposes that the labeled entity exists before the labeling process and can then be given a verbal tag. However, this assumption may be questioned for concepts, especially for abstract ones, where the formation of the conceptual representation may in part depend on concordant language (for arguments for this latter claim, see the Discussion section).

2 In accordance with the nomenclature in previous work, model areas will be referred to as *A1, *AB, and so on. The asterisk is used to distinguish model areas from their human cortical area counterparts.

References

Althaus, N., & Mareschal, D. (2014). Labels direct infants' attention to commonalities during novel category learning. *PLoS ONE*, *9*(7), Article e99670. https://doi.org/10.1371/journal.pone.0099670

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, *116*(3), 463–498. https://doi.org/10.1037/a0016261

Athanasopoulos, P., & Casaponsa, A. (2020). The Whorfian brain: Neuroscientific approaches to linguistic relativity. *Cognitive Neuropsychology*, *37*(5–6), 393–412. https://doi.org/10.1080/02643294.2020.1769050

Baker, G. P., & Hacker, P. M. S. (2008). *An analytical commentary on the Philosophical Investigations: Vol. 1. Wittgenstein: Understanding and meaning*. Blackwell Publishing. https://doi.org/10.1002/9780470752807

Banks, B., & Connell, L. (2023). Multi-dimensional sensorimotor grounding of concrete and abstract categories. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *378*(1870), Article 20210366. https://doi.org/10.1098/rstb.2021.0366

Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and* *action in memory, language, and thinking* (pp. 129–163). Cambridge University Press. https://doi.org/10.1017/CBO9780511499968.007

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(8), 1798–1828. https://doi.org/10.1109/TPAMI.2013.50

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, *15*(11), 527–536. https://doi.org/10.1016/j.tics.2011.10.001

Binder, J. R., Westbury, C. F., McKiernan, K. A., Possing, E. T., & Medler, D. A. (2005). Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuroscience*, *17*(6), 905–917. https://doi.org/10.1162/0898929054021102

Bohnemeyer, J. (2021). Linguistic relativity. In D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann, T. E. Zimmerman, & D. Voloshina (Eds.), *The Wiley Blackwell companion to semantics* (pp. 1–33). Wiley Blackwell. https://doi.org/10.1002/9781118788516.sem013

Borghi, A. M., & Binkofski, F. (2014). *Words as social tools: An embodied view applied to abstract concepts*. Springer. https://doi.org/10.1007/978-1-4614-9539-0

Borghi, A. M., Shaki, S., & Fischer, M. H. (2022). Abstract concepts: External influences, internal constraints, and methodological issues. *Psychological Research*, *86*(8), 2370–2388. https://doi.org/10.1007/s00426-022-01698-4

Braitenberg, V. (1978). Cell assemblies in the cerebral cortex. In S. Levin, R. Heim, & G. Palm (Eds.), *Theoretical approaches to complex systems* (pp. 171–188). Springer. https://doi.org/10.1007/978-3-642-93083-6_9

Brown, R. W. (1957). Linguistic determinism and the part of speech. *Journal of Abnormal Psychology*, *55*(1), 1–5. https://doi.org/10.1037/h0041199

Buccino, G., Colagè, I., Silipo, F., & D'Ambrosio, P. (2019). The concreteness of abstract language: An ancient issue and a new perspective. *Brain Structure & Function*, *224*(4), 1385–1401. https://doi.org/10.1007/s00429-019-01851-7

Cangelosi, A., Greco, A., & Harnad, S. (2000). From robotic toil to symbolic theft: Grounding transfer from entry-level to higher-level categories. *Connection Science*, *12*(2), 143–162. https://doi.org/10.1080/09540090050129763

Cangelosi, A., & Stramandinoli, F. (2018). A review of abstract concept learning in embodied agents and robots. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *373*(1752). https://doi.org/10.1098/rstb.2017.0131

Chen, L., Lambon Ralph, M. A., & Rogers, T. T. (2017). A unified model of human semantic knowledge and its disorders. *Nature Human Behaviour*, *1*(3). https://doi.org/10.1038/s41562-016-0039

Constant, M., Pulvermüller, F., & Tomasello, R. (2023). Brain-constrained neural modeling explains fast mapping of words to meaning. *Cerebral Cortex*, *33*(11), 6872–6890. https://doi.org/10.1093/cercor/bhad007

Davies, M. (2018). *The iWeb Corpus* [Data set]. https://www.english-corpora.org/iWeb/

Deco, G., Rolls, E. T., Albantakis, L., & Romo, R. (2013). Brain mechanisms for perceptual and reward-related decision-making. *Progress in Neurobiology*, *103*, 194–213. https://doi.org/10.1016/j.pneurobio.2012.01.010

Dove, G. (2016). Three symbol ungrounding problems: Abstract concepts and the future of embodied cognition. *Psychonomic Bulletin & Review*, *23*(4), 1109–1121. https://doi.org/10.3758/s13423-015-0825-4

Dove, G. (2021). The challenges of abstract concepts. In Robinson (Ed.), *Handbook of embodied psychology* (pp. 171–195). Springer International Publishing. https://doi.org/10.1007/978-3-030-78471-3_8

Dove, G. (2022). *Abstract concepts and the embodied mind: Rethinking grounded cognition*. Oxford University Press. https://doi.org/10.1093/oso/9780190061975.001.0001

Dreyer, F. R., & Pulvermüller, F. (2018). Abstract semantics in the motor system? An event-related fMRI study on passive reading of semantic word categories carrying abstract emotional and mental meaning. *Cortex*, *100*, 52–70. https://doi.org/10.1016/j.cortex.2017.10.021

Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences*, *8*(7), 301–306. https://doi.org/10.1016/j.tics.2004.05.003

Fodor, J. A. (1975). *The language of thought*. Harvard University Press. https://doi.org/10.2307/2184356

Fodor, J. A. (2008). *LOT 2: The language of thought revisited*. Clarendon Press. https://doi.org/10.1093/acprof:oso/9780199548774.001.0001

Fuster, J. M., & Bressler, S. L. (2012). Cognit activation: A mechanism enabling temporal integration in working memory. *Trends in Cognitive Sciences*, *16*(4), 207–218. https://doi.org/10.1016/j.tics.2012.03.005

Garagnani, M., & Pulvermüller, F. (2016). Conceptual grounding of language in action and perception: A neurocomputational model of the emergence of category specificity and semantic hubs. *The European Journal of Neuroscience*, *43*(6), 721–737. https://doi.org/10.1111/ejn.13145

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. *Language*, *2*, 301–334.

Glenberg, A. M. (2022). Embodiment and learning of abstract concepts (such as algebraic topology and regression to the mean). *Psychological Research*, *86*(8), 2398. https://doi.org/10.1007/s00426-021-01576-5

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558–565. https://doi.org/10.3758/bf03196313

Gumperz, J. J., & Levinson, S. C. (1991). Rethinking linguistic relativity. *Current Anthropology*, *32*(5), 613–623. https://doi.org/10.1086/204009

Hale, S. C. (1988). Spacetime and the abstract/concrete distinction. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *53*(1), 85–102. https://doi.org/10.1007/BF00355677

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1–3), 335–346. https://doi.org/10.1016/0167-2789(90)90087-6

Harnad, S. (2018). Grounding symbolic capacity in robotic capacity. In L. Steels & R. Brooks (Eds.), *The artificial life route to artificial intelligence* (pp. 277–286). Routledge. https://doi.org/10.4324/9781351001885-11

Harpaintner, M., Sim, E.J., Trumpp, N. M., Ulrich, M., & Kiefer, M. (2020). The grounding of abstract concepts in the motor and visual system: An fMRI study. *Cortex*, *124*, 1–22. https://doi.org/10.1016/j.cortex.2019.10.014

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. John Wiley & Sons.

Henningsen-Schomers, M. R., Garagnani, M., & Pulvermüller, F. (2023). Influence of language on perception and concept formation in a brain-constrained deep neural network model. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *378*(1870), Article 20210373. https://doi.org/10.1098/rstb.2021.0373

Henningsen-Schomers, M. R., & Pulvermüller, F. (2022). Modelling concrete and abstract concepts using brain-constrained deep neural networks. *Psychological Research*, *86*(8), 2533–2559. https://doi.org/10.1007/s00426-021-01591-6

Hoffman, P., McClelland, J. L., & Lambon Ralph, M. A. (2018). Concepts, control, and context: A connectionist account of normal and disordered semantic cognition. *Psychological Review*, *125*(3), 293–328. https://doi.org/10.1037/rev0000094

Ito, T., Klinger, T., Schultz, D. H., Murray, J. D., Cole, M. W., & Rigotti, M. (2022). *Compositional generalization through abstract representations in human and artificial neural networks*. arXiv. https://doi.org/10.48550/arXiv.2209.07431

Johnston, W. J., & Fusi, S. (2023). Abstract representations emerge naturally in neural networks trained to perform multiple tasks. *Nature Communications*, *14*(1), 1040. https://doi.org/10.1038/s41467-023-36583-0

Katz, J. J., & Fodor, J. A. (1963). The structure of a semantic theory. *Language*, *39*(2), 170. https://doi.org/10.2307/411200

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex*, *48*(7), 805–825. https://doi.org/10.1016/j.cortex.2011.04.006

Kintsch, W. (1974). *The representation of meaning in memory*. Lawrence Erlbaum. https://doi.org/10.4324/9781315794563

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

Langland-Hassan, P., Faries, F. R., Gatyas, M., Dietz, A., & Richardson, M. J. (2021). Assessing abstract thought and its relation to language with a new nonverbal paradigm: Evidence from aphasia. *Cognition*, *211*, Article 104622. https://doi.org/10.1016/j.cognition.2021.104622

LaTourrette, A. S., & Waxman, S. R. (2020). Naming guides how 12-month-old infants encode and remember objects. *Proceedings of the National Academy of Sciences*, *117*(35), 21230–21234. https://doi.org/10.1073/pnas.2006608117

Lenci, A., Lebani, G. E., & Passaro, L. C. (2018). The emotions of abstract words: A distributional semantic analysis. *Topics in Cognitive Science*, *10*(3), 550–572. https://doi.org/10.1111/tops.12335

Löhr, G. (2022). What are abstract concepts? On lexical ambiguity and concreteness ratings. *Review of Philosophy and Psychology*, *13*(3), 549–566. https://doi.org/10.1007/s13164-021-00542-9

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology*, *3*, Article 54. https://doi.org/10.3389/fpsyg.2012.00054

Lupyan, G., Abdel Rahman, R., Boroditsky, L., & Clark, A. (2020). Effects of language on visual perception. *Trends in Cognitive Sciences*, *24*(11), 930–944. https://doi.org/10.1016/j.tics.2020.08.005

Lupyan, G., & Mirman, D. (2013). Linking language and categorization: Evidence from aphasia. *Cortex*, *49*(5), 1187–1194. https://doi.org/10.1016/j.cortex.2012.06.006

Mahon, B. Z., & Caramazza, A. (2009). Concepts and categories: A cognitive neuropsychological perspective. *Annual Review of Psychology*, *60*, 27–51. https://doi.org/10.1146/annurev.psych.60.110707.163532

Maier, M., & Abdel Rahman, R. (2019). No matter how: Top-down effects of verbal and semantic category knowledge on early visual perception. *Cognitive, Affective & Behavioral Neuroscience*, *19*(4), 859–876. https://doi.org/10.3758/s13415-018-00679-8

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences*, *8*(3), 108–114. https://doi.org/10.1016/j.tics.2004.01.003

Mandler, J. M. (2004). Thought before language. *Trends in Cognitive Sciences*, *8*(11), 508–513. https://doi.org/10.1016/j.tics.2004.09.004

Miller, T. M., Schmidt, T. T., Blankenburg, F., & Pulvermüller, F. (2018). Verbal labels facilitate tactile perception. *Cognition*, *171*, 172–179. https://doi.org/10.1016/j.cognition.2017.10.010

Milner, P. M. (1957). The cell assembly: Mark II. *Psychological Review*, *64*(4), 242–252. https://doi.org/10.1037/h0042287

Moseley, R. L., Carota, F., Hauk, O., Mohr, B., & Pulvermüller, F. (2012). A role for the motor system in binding abstract emotional meaning. *Cerebral Cortex*, *22*(7), 1634–1647. https://doi.org/10.1093/cercor/bhr238

Moseley, R. L., & Pulvermüller, F. (2014). Nouns, verbs, objects, actions, and abstractions: Local fMRI activity indexes semantics, not lexical categories. *Brain and Language*, *132*, 28–42. https://doi.org/10.1016/j.bandl.2014.03.001

Naumann, D., Frassinelli, D., & Im Schulte Walde, S. (2018). Quantitative semantic variation in the contexts of concrete and abstract words. In M. Nissim, J. Berant, & A. Lenci (Eds.), *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics* (pp. 76–85). Association for Computational Linguistics. https://doi.org/10.18653/v1/S18-2008

Nguyen, P. T. U., Henningsen-Schomers, M. R., & Pulvermüller, F. (2024). Causal influence of linguistic learning on perceptual and conceptual processing: A brain-constrained deep neural network study of proper names and category terms. *Journal of Neuroscience*, 44(9). https://doi.org/10.1523/JNEUROSCI.1048-23.2023

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195066661.001.0001

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue Canadienne De Psychologie*, *45*(3), 255–287. https://doi.org/10.1037/h0084295

Palm, G. (1982). *Neural assemblies*. Springer.

Parisi, D., & Cangelosi, A. (2002). *Simulating the evolution of language*. Springer. https://doi.org/10.1007/978-1-4471-0663-0

Pearce, J. M. (2008). *Animal learning & cognition: An introduction* (3rd ed.). Psychology Press. https://doi.org/10.4324/9781315782911

Perszyk, D. R., & Waxman, S. R. (2018). Linking language and cognition in infancy. *Annual Review of Psychology*, *69*, 231–250. https://doi.org/10.1146/annurev-psych-122216-011701

Pexman, P. M., Diveica, V., & Binney, R. J. (2023). Social semantics: The organization and grounding of abstract concepts. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *378*(1870), Article 20210363. https://doi.org/10.1098/rstb.2021.0363

Pinker, S. (2008). *The stuff of thought: Language as a window into human nature*. Penguin.

Pulvermüller, F. (Ed.). (2002). *The neuroscience of language: On brain circuits of words and serial order*. Cambridge University Press. https://doi.org/10.1017/CBO9780511615528

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews. Neuroscience*, *6*(7), 576–582. https://doi.org/10.1038/nrn1706

Pulvermüller, F. (2013). How neurons make meaning: Brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences*, *17*(9), 458–470. https://doi.org/10.1016/j.tics.2013.06.004

Pulvermüller, F. (2018a). The case of CAUSE: Neurobiological mechanisms for grounding an abstract concept. *Philosophical Transactions of the Royal Society of*

London. Series B, Biological Sciences, *373*(1752). https://doi.org/10.1098/rstb.2017.0129

Pulvermüller, F. (2018b). Neural reuse of action perception circuits for language, concepts and communication. *Progress in Neurobiology*, *160*, 1–44. https://doi.org/10.1016/j.pneurobio.2017.07.001

Pulvermüller, F. (2018c). Neurobiological mechanisms for semantic feature extraction and conceptual flexibility. *Topics in Cognitive Science*, *10*(3), 590–620. https://doi.org/10.1111/tops.12367

Pulvermüller, F. (2023). Neurobiological mechanisms for language, symbols and concepts: Clues from brain-constrained deep neural networks. *Progress in Neurobiology*, Article 102511. https://doi.org/10.1016/j.pneurobio.2023.102511

Pulvermüller, F., & Garagnani, M. (2014). From sensorimotor learning to memory cells in prefrontal and temporal association cortex: A neurocomputational study of disembodiment. *Cortex*, *57*, 1–21. https://doi.org/10.1016/j.cortex.2014.02.015

Pulvermüller, F., Garagnani, M., & Wennekers, T. (2014). Thinking in circuits: Toward neurobiological explanation in cognitive neuroscience. *Biological Cybernetics*, *108*(5), 573–593. https://doi.org/10.1007/s00422-014-0603-9

Pulvermüller, F., & Preißl, H. (1991). A cell assembly model of language. *Network: Computation in Neural Systems*, *2*(4), 455–468. https://doi.org/10.1088/0954-898X_2_4_008

Pulvermüller, F., Tomasello, R., Henningsen-Schomers, M. R., & Wennekers, T. (2021). Biological constraints on neural network models of cognitive function. *Nature Reviews. Neuroscience*, *22*(8), 488–502. https://doi.org/10.1038/s41583-021-00473-5

Pusch, R., Clark, W., Rose, J., & Güntürkün, O. (2023). Visual categories and concepts in the avian brain. *Animal Cognition*, *26*(1), 153–173. https://doi.org/10.1007/s10071-022-01711-8

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition*. The MIT Press. https://doi.org/10.7551/mitpress/6161.001.0001

Rosch, E., & Lloyd, B. (1978). *Principles of categorization*. Lawrence Erlbaum. https://doi.org/10.1016/B978-1-4832-1446-7.50028-5

Schomers, M. R., Garagnani, M., & Pulvermüller, F. (2017). Neurocomputational consequences of evolutionary connectivity changes in perisylvian language cortex. *The Journal of Neuroscience*, *37*(11), 3045–3055. https://doi.org/10.1523/JNEUROSCI.2693-16.2017

Schwanenflugel, P. J., & Akin, C. E. (1994). Developmental Trends in Lexical Decisions for Abstract and Concrete Words. *Reading Research Quarterly*, *29*(3), 251–264. https://doi.org/10.2307/747876

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417–424. https://doi.org/10.1017/S0140525X00005756

Sloutsky, V. M. (2010). From perceptual categories to concepts: What develops? *Cognitive Science*, *34*(7), 1244–1286. https://doi.org/10.1111/j.1551-6709.2010.01129.x

Sloutsky, V. M., & Robinson, C. W. (2008). The role of words and sounds in infants' visual processing: From overshadowing to attentional tuning. *Cognitive Science*, *32*(2), 342–365. https://doi.org/10.1080/03640210701863495

Thierry, G. (2016). Neurolinguistic relativity: How language flexes human perception and cognition. *Language Learning*, *66*(3), 690–713. https://doi.org/10.1111/lang.12186

Tomasello, M., & Kruger, A. C. (1992). Joint attention on actions: Acquiring verbs in ostensive and non-ostensive contexts. *Journal of Child Language*, *19*(2), 311–333. https://doi.org/10.1017/s0305000900011430

Tomasello, R., Garagnani, M., Wennekers, T., & Pulvermüller, F. (2017). Brain connections of words, perceptions and actions: A neurobiological model of spatio-temporal semantic activation in the human cortex. *Neuropsychologia*, *98*, 111–129. https://doi.org/10.1016/j.neuropsychologia.2016.07.004

Tomasello, R., Garagnani, M., Wennekers, T., & Pulvermüller, F. (2018). A neurobiologically constrained cortex model of semantic grounding with spiking neurons and brain-like connectivity. *Frontiers in Computational Neuroscience*, *12*, Article 88. https://doi.org/10.3389/fncom.2018.00088

Tomasello, R., Wennekers, T., Garagnani, M., & Pulvermüller, F. (2019). Visual cortex recruitment during language processing in blind individuals is explained by Hebbian learning. *Scientific Reports*, *9*(1), Article 3579. https://doi.org/10.1038/s41598-019-39864-1

Vigliocco, G., Kousta, S.-T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (2013). The neural representation of abstract words: The role of emotion. *Cerebral Cortex*, *24*(7), 1767–1777. https://doi.org/10.1093/cercor/bht025

Villani, C., Lugli, L., Liuzza, M. T., Nicoletti, R., & Borghi, A. M. (2021). Sensorimotor and interoceptive dimensions in concrete and abstract concepts. *Journal of Memory and Language*, *116*, Article 104173. https://doi.org/10.1016/j.jml.2020.104173

Vouloumanos, A., & Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. *Developmental Psychology*, *45*(6), 1611–1617. https://doi.org/10.1037/a0016134

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology*, *29*(3), 257–302. https://doi.org/10.1006/cogp.1995.1016

Wennekers, T. (2009). *Felix: A simulation-tool for neural networks (and dynamical systems). User guide*. PEARL repository, University of Plymouth. http://hdl.handle.net/10026.1/15219

Wermter, S. (2004). Towards multimodal neural robot learning. *Robotics and Autonomous Systems*, *47*(2-3), 171–175. https://doi.org/10.1016/S0921-8890(04)00047-8

Westermann, G., & Mareschal, D. (2014). From perceptual to language-mediated categorization. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1634), Article 20120391. https://doi.org/10.1098/rstb.2012.0391

Whorf, B. L., & Carroll, J. B. (Eds.). (1976). *Language, thought, and reality*. MIT Press.

Wittgenstein, L. (1953). *Philosophical investigations. (Anscombe, G. E. M., Trans.*). Macmillan.

Zipser, D., Kehoe, B., Littlewort, G., & Fuster, J. M. (1993). A spiking network model of short-term active memory. *The Journal of Neuroscience*, *13*(8), 3406–3420. https://doi.org/10.1523/JNEUROSCI.13-08-03406.1993

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Accessible Summary
Appendix S1. Structure and Function of the Spiking Neuron Model.
Appendix S2. Details on Data Analysis and Additional Results.
Appendix S3. Extended Discussion on Training Paradigms.