Intrinsic Lexical Intentionality and the Mathematics of Homomorphism

Hermann Moisl

Abstract

Moisl [1, 2] proposed a model of how the brain implements intrinsic intentionality with respect to lexical and sentence meaning, where 'intrinsic' is understood as 'independent of interpretation by observers external to the cognitive agent'. The discussion in both was mainly philosophical and qualitative; the present paper gives a mathematical account of the distance structure preservation that underlies the proposed mechanism of intrinsic intentionality. The three-layer autoassociative multilayer perceptron (aMLP) architecture with nonlinear hidden and linear output layers is the component in the model which generates representations homomorphic with the environment. The discussion first cites existing work which identifies the aMLP as an implementation architecture for principal component analysis (PCA), and then goes on to argue that the homomorphism characteristic of linear functions like PCA extends to aMLPs with nonlinear activation functions in the hidden layer. The discussion is in two main parts: the first part outlines the model, and the second presents the mathematical account.

1. Introduction

Moisl [1, 2] proposed a model of how the brain implements intrinsic intentionality with respect to lexical and sentence meaning, where 'intrinsic' is understood as 'independent of interpretation by observers external to the cognitive agent'. The discussion in both was mainly philosophical and qualitative; the present paper gives a mathematical account of the distance structure preservation that underlies the proposed mechanism of intrinsic intentionality. The discussion is in two main parts: the first part outlines the model, and the second presents the mathematical account.

2. Outline

2.1 Intentionality

Humans intuitively feel that they possess a head-internal meaningfulness, that is, an awareness of the self and its relationship to the perceived world which is independent of interpretation of one's behaviour by observers. This intuition is captured by the philosophical concept of intentionality [3-5] which is used in present-day philosophy of mind to denote the 'aboutness' of mental states, '*the power of minds and mental states to be about, to represent, or to stand for, things, properties and states of affairs*' [3].

In 1980 and subsequent publications [6] the philosopher John Searle argued that the then-dominant theoretical framework in cognitive science, the Computational Theory of Mind (CTM; [7]), was incapable of explaining how the mind comes to possess this head-internal meaningfulness - what Searle called original intentionality. The essence of his position was that CTM representations lack intrinsic meaning because their meaning is dependent on ascription by an observer. He distinguished two types of intentionality, original and derived. The locus of original intentionality is

the human head, and derived intentionality is that which we attribute to physical mechanisms which we have good reason to believe do not have original intentionality, such as thermostats, whose operation is routinely interpreted by humans as wanting to maintain an even temperature but whose structure is too simple for it to have desires. His argument is that, with respect to intentionality, a computer is like a thermostat. The argument is based on his well known Chinese Room thought experiment. There is a closed room containing Searle and a list of rules in English for manipulating Chinese orthographic symbols. Chinese speakers outside the room put sequences of these symbols into the room and, using the rules available to him, Searle assembles and outputs sequences of Chinese symbols in response. The people outside interpret the input sequences as sentences in Chinese whose meaning they understand and the output sequences as reasonable responses to them, and on the basis of the room's conceptually coherent input-output behaviour conclude that it understands Chinese. Searle himself, however, knows that the room does not understand Chinese because he, the interpreter and constructor of the sequences, does not understand Chinese, but is only following instructions without knowing what the input and output sequences mean.

The Room is, of course, a computer. Searle is the CPU, the list of English instructions is a program, and the input-output sequences are symbol strings; by concluding that the room understands Chinese, its observers have confirmed the Turing Test [8], which says that any device which can by its behaviour convince observers that it has human-level intentionality must be considered to possess it. Searle knows, however, that the room's intentionality is derived, the implication being that physical computer implementations of CTM models, like thermostats, only have derived intentionality. The intentionality of the symbols manipulated by the algorithm of a CTM model is in the heads and only in the heads of their human designers. When physically instantiated, for example by compilation of a CTM model onto a physical computer, this intentionality is lost: the symbols of the interpreted model cease to be symbolic and become physical bit-strings which drive the physical causal dynamics of the machine, but intentionality is not a factor in that dynamics. The behaviour of the machine can be interpreted as intentional, just as the behaviour of the Chinese Room or a thermostat can be, but the semantics is derived because the only locus of intentionality is in the heads of observers. Put simply, a physical computer does not understand what it is doing any more than a vending machine does. It only pushes physical bit-strings around, and humans interpret that activity as intentional [9].

Searle's position remains controversial among philosophers of mind and cognitive scientists more generally after four decades [6] - a clear indication that it is a substantive one worthy of serious consideration. Rather than add to the already-voluminous literature bearing on it, Moisl [1, 2] as well as the present discussion simply accept the validity of Searle's position and, on that assumption, propose a way of modelling intrinsic intentionality at the implementation level.

2.2 The Implementation Level

The black box problem in system identification [10] builds models of physical systems based on observation of responses to system input: given a box whose internal mechanism is hidden but whose input-output behaviour is observable, what mechanism inside the box generates that behaviour? The answer is that there is an arbitrary number of possible different mechanisms for any given input-output behaviour ([11], Ch. 3.2); the only way to know for certain what's in the box

is to look inside.

Applied to black boxes in general, the doctrine of emergence in the philosophy of science [12] addresses the relationship of physics to the 'special' sciences, which study objects, properties, or behaviours that emerge from the physical substrate of the natural world. The standard view is that the sciences are related via levels of description whereby any physical system can be described at an arbitrary number of levels using a theoretical ontology appropriate to each, every level is explanatorily autonomous with respect to the others subject to the constraint of consistency between and among levels, and selection of any particular level is determined by the research question being asked. The principle of supervenience [13] says that descriptions of natural systems constitute a hierarchy where the properties at any given level implement those at the level above. For the physical monist [14], everything in the natural world is physical and therefore describable using the theoretical ontology of physics, but this does not rule out the ontologies of sciences addressing supervenient phenomena or require their reduction to physics [14, 15] on the grounds that different theoretical ontologies are needed to capture different sorts of regularity in nature.

For linguistic meaning the black box is the human head and the input-output behaviour is conversation. The CTM view of what's in the head is that it is a Turing Machine whose program is cognition. When the box is opened, however, one looks in vain for the data structures and algorithms of CTM, and finds instead billions of interconnected neurons. Some have argued that study of the brain by cognitive neuroscience will supplant the theoretical ontology of CTM, but this is not the majority view [16]. The alternative adopted here is nonreductive physicalism [14], which in a cognitive science context says that accounts of the structure and operation of mind and brain are separate and autonomous levels of description. It accepts that human cognition is implemented by and only by the physical brain, but maintains that this does not preclude the mentalistic ontology of CTM or require its reduction to neuroscience. The present discussion focuses on the implementation level - the physical mechanism of intentionality.

2.3 Meaning

Proposal of an implementation model implies clarity about what is being implemented. 'Meaning' is understood, and its theoretical characterization is approached, in a variety of ways, for an overview of which see [17]. Speaks [17] distinguishes 'logical' approaches in the tradition of Frege, where meaning is seen as semantic interpretation of symbols in an abstract formal system, and 'foundational' approaches which focus on the mechanism of semantic interpretation; foundational approaches are subcategorized into 'use' theories such as those of Grice, and 'mentalistic' ones which relate linguistic meaning to the structure of cognition. The present discussion takes the mentalistic approach. Specifically, it adopts the tradition in Western thought [18] ranging from Aristotle to theories of mental content in present-day linguistics and cognitive science more generally [19] that the meaning of a word is its signification of a mental concept, and a mental concept is a representation of the mind-external environment causally generated by the cognitive agent's interaction with that environment. In recent times this tradition has continued in attempts to 'naturalize' the mind, that is, to see the mind as an aspect of the natural world and therefore theoretically explicable in terms of the natural sciences [4, 20]. The precursors of naturalism were empiricist philosphers like like Mill (1806-73; [21]) and scientists like von Helmholtz (1821–1894;

[22]) and Mach (1838–1916; [23]). Von Helmholtz stressed the importance of sensory perception of and bodily interaction with the environment in generating a coherent system of mental representation whose structure mirrors that of the environment, and Mach saw human mentality as a teleological dynamical system tending to equilibrium with the environment via sensory and enactive interaction. In the present day, the tradition exists in a variety of disciplines and approaches to the study of mind and language: naturalistic, evolutionary, and teleological epistemology [5, 24, 25], externalist semantics in philosophy of mind [26], evolutionary psychology [27] and embodied cognition in cognitive psychology [28-30], cognitive linguistics [31, 32] and conceptual semantics [33, 34] in generative linguistics.

2.4 The Model

According to Searle, '*intentionality in human beings (and animals) is a product of causal features of the brain',* and '*any attempt literally to create intentionality artificially (strong AI) could not succeed just by designing programs but would have to duplicate the causal powers of the human brain'.* Since the validity of Searle's position was and in the present discussion continues to be assumed, the choice of artificial neural networks (ANN) as the modelling framework was obvious: though radically simplified with respect to the biological brain, they do retain its fundamental architectural characteristics as a collection of massively interconnected processing units that learns to represent environmental inputs via synaptic strength modification. The research question that motivates the present discussion thereby becomes: How can the brain or a physical mechanism analogous to it implement intrinsic intentionality?

The solution proposed for implementation of lexical intentionality was the structure of interconnected ANNs shown in Figure 1.

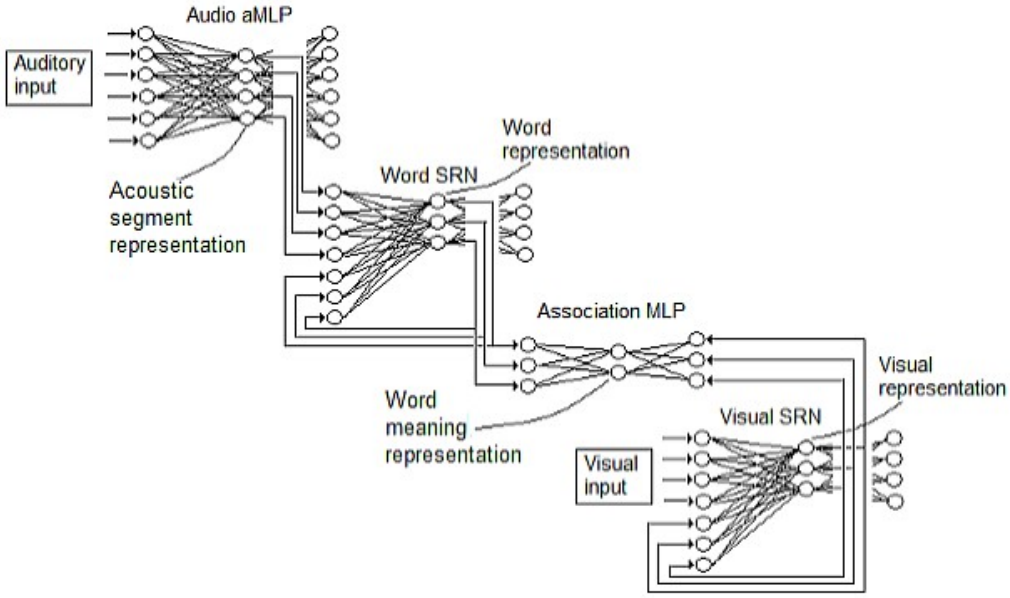


Figure 1. The structure of the lexical intentionality model.

Spoken word and visual inputs from an environment are simultaneously presented to their eponymous subnets, where sequences of representations are generated in their respective hidden

layers; the numbers of units in the various subnets are small for tractability of graphical presentation and would need to be much larger in a practical implementation. These representations are associated in the association subnet, whose hidden layer was argued to be the implementation of lexical intentionality, that is, of the meaning of the word.

Fundamental to the model is the autoassociative multilayer perceptron (aMLP), an example of which is shown in Figure 2.

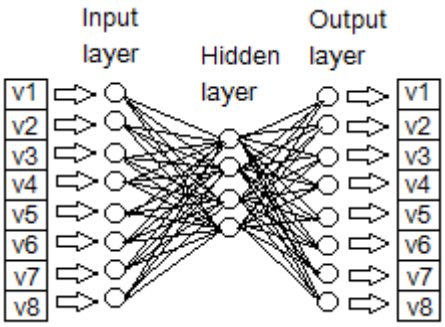


Figure 2. Example of an autoassociative multilayer perceptron.

The aMLP appears as the audio input subnet in Figure 1, and with feedback in the word and visual subnets. The input and target output of an aMLP are identical, so that after training with a collection of inputs I, presentation of any given input I_j generates I_j in the output units. The hidden layer, which contains fewer units, is a compact representation of I_j, and when the representations of all the components of I were cluster analyzed, their structure was found to be similar to that of I; 'similar' as understood here is defined in Section 3. For example, if I is the collection of the 26 letter forms of the Roman alphabet represented as 12 × 12 bitmaps, as shown in Figure 3, an aMLP is able to learn representations of the similarity structure of these bitmaps in its hidden layer, as shown by the cluster analyses in Figure 4.

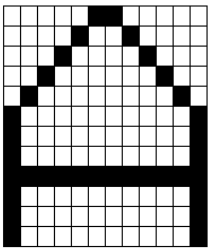


Figure 3. A letter bitmap.

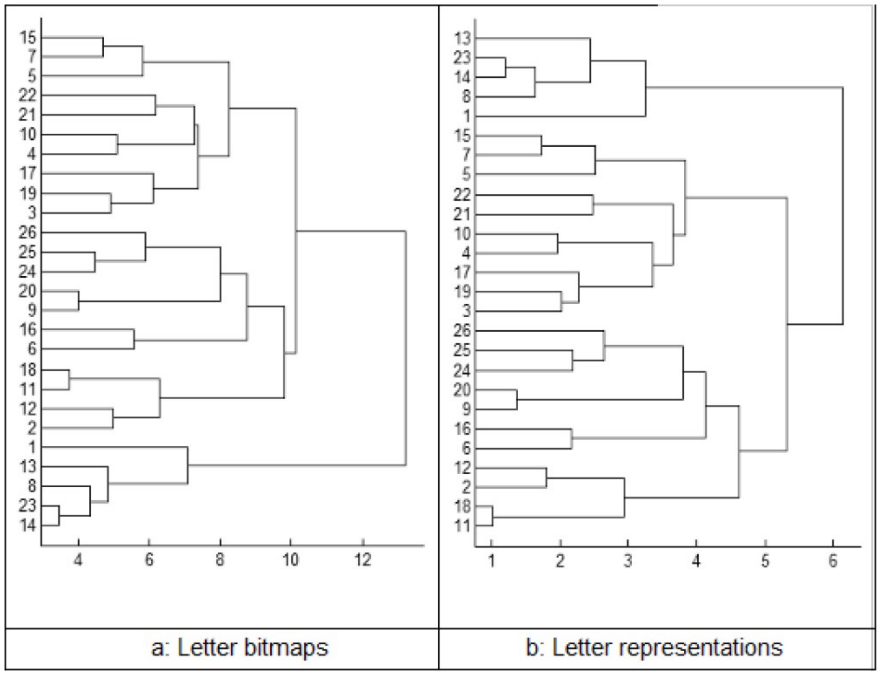


Figure 4. Cluster trees for letter bitmaps and their aMLP representations.

The idea that the structure of head-internal representations generated by a cognitive agent's interaction with an agent-external environment is similar to the spatial and temporal structure of that environment has a long history in cognitive science. It was proposed in Antiquity by philosopers like Aristotle, Augustine, and Boethius [18] and, more recently, by Mach [23] and von Helmholtz [22]; current examples are ([4, 5, 19, 35-53]).

The relevance of similarity-based models to present concerns is that they can be understood as implementation-level models of intrinsic intentionality in biological brains. Their implication is that the formal similarity structure of the neural activations which causally drive brain dynamics reflect the similarity structure of mind-external objects and their interactions, and are thereby 'about' the mind-external world without involvement of a system-external interpreter.

The model in Figure 1

(i.) causally generates its own system-internal representations of external environmental input.
(ii.) The physical form of these representations is determined by that which they represent.
(iii.) For a given environmental domain, the structure of the representations is similar to that of the domain and thereby model it.
(iv.) The representations are causal in the input-output behaviour of the system.

Assuming the validity of the foregoing comments about the relevance of structural similarity to modelling of intentionality, the structure in Figure 1 is a mapping of words to visual states of the world, and thereby a model of a physical system that implements intrinsic lexical intentionality.

Finally, it is freely admitted that Figure 1 is too simple to serve as a general implementation model for lexical meaning. It does not, for example, address the intentionality of what classical antiquity and medieval scholastic philosophy called universals [18] like "truth" and, more prosaically, the abstract category "human", which have no existence in the mind-external world and cannot

therefore generate sensory representations. Nor does it incorporate the extensive neuroscientific work on the integration of language and object recognition reviewed, for example, in Plebe and de la Cruz ([54]: Ch. 6); as one of the reviewers of this paper pointed out with respect to the input/target output identity of an aMLP, '*it is not properly true that -in humans- the same sensory input always generates the same output due to phenomena such as learning, adaptation, and updating*'. Figure 1 is, however, not intended as a general model for lexical meaning. The intention is much more limited and specific: to propose a possible solution to Searle's problem of intrinsic intentionality in physical systems. The research question was and is: '*How can the brain or a physical mechanism analogous to it implement intrinsic intentionality?*'. The proposed answer is: structural similarity of system inputs and their system-internal representations.

3. Mathematics

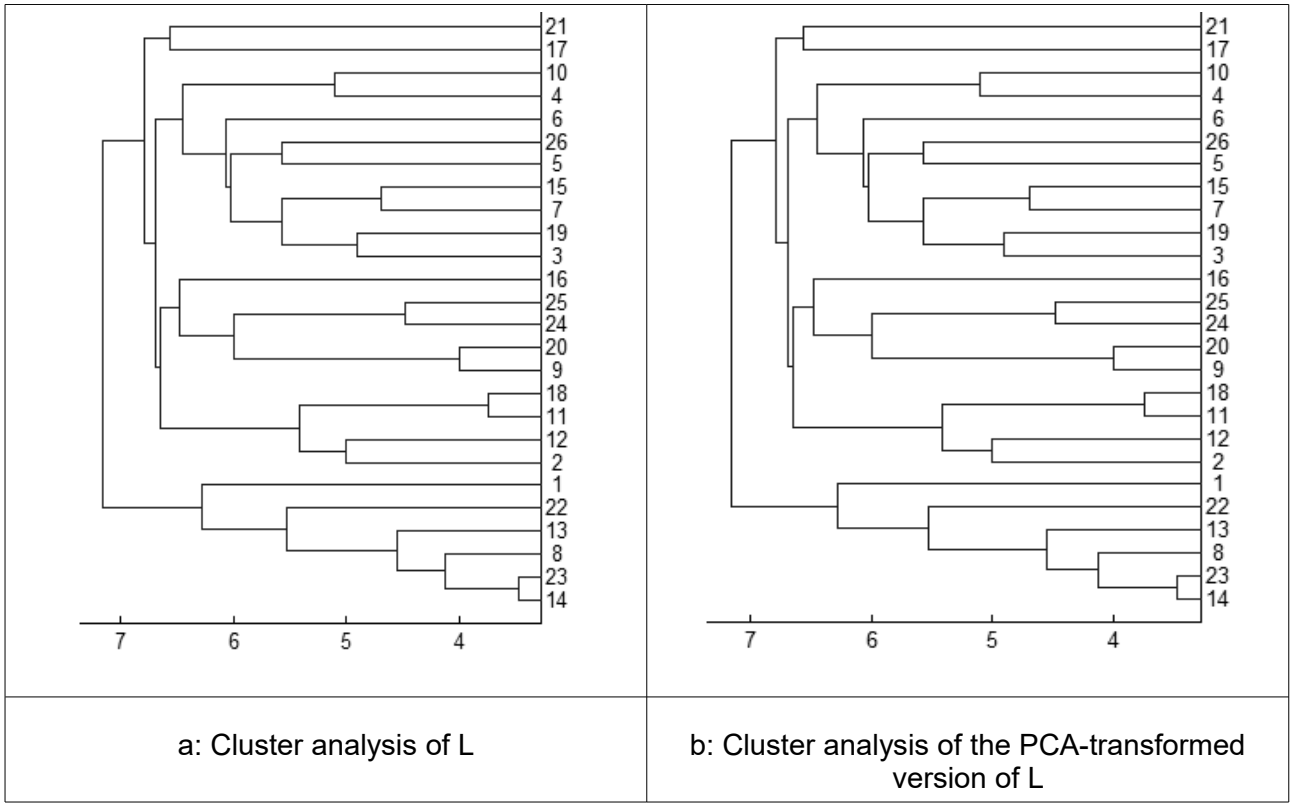
This part of the discussion describes a way of understanding the foregoing preservation of input similarity structure in system-internal representations as mathematical homomorphism; the references for standard mathematical topics used in what follows are Gowers et al [55] and Weisstein [56], and for artificial neural networks the reference is Aggarwal [57].

In current mathematics a space is understood as a pair S = (*Obj, Op*), where *Obj* is a set of mathematical objects of some particular type and *Op* is a set of operations defined on *Obj* such as scalar multiplication defined on vectors. The input, hidden, and output layers of ANNs are mathematically represented as real-valued vectors, so what follows will focus on vector spaces.

A vector space is a set V of vectors and the associated operations are vector addition and multiplication of a vector by a scalar. In what follows, some way of characterizing distances among vectors in V will be required, and these two operations on their own are insufficient for that. What is required is a metric *d*, that is, a function that returns a measure of the distance between two vectors v and w ϵ V, inclusion of which transforms a vector space into a metric space. The most familiar metric space is the *n*-dimensional Euclidian in which the metric is Euclidian distance, the shortest distance between two points. Input, hidden, and output aMLP layers are here interpreted as Euclidian spaces.

A homomorphism is a structure-preserving map between two spaces ([55]), and in the present case between two Euclidean ones. We are interested in determining whether or not the distance structure of the input space of an aMLP is preserved in its hidden layer space, where preservation of distance structure is taken to be systematic preservation of proportionality of Euclidean distance between of any two vectors v and w in two spaces of possibly-different dimensionalities.

Every linear transformation of vectors in a Euclidean space is homomorphic. This is exemplified in principal component analysis (PCA; [58, 59]); the relevance of this example will emerge shortly. PCA transforms a Euclidean input space E_1 into a Euclidean output space E_2 by linear combination of the vectors in E_1, and the distance structure of E_1 is preserved in E_2. For example, Figure 5a shows a cluster analysis of the letter matrix, henceforth L, described in Section 2, and Figure 5b of its PCA transformation:

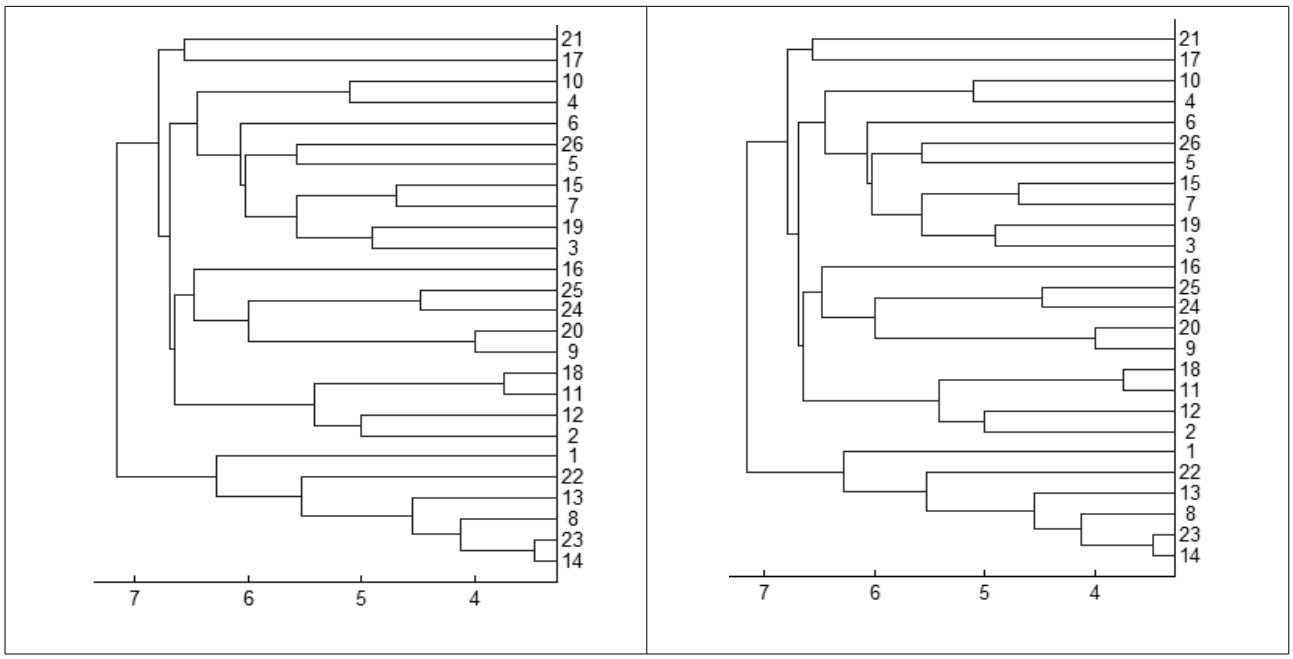


a: Cluster analysis of L

b: Cluster analysis of the PCA-transformed version of L

Figure 5. Cluster analyses of L and its PCA transformation, both 144-dimensional.

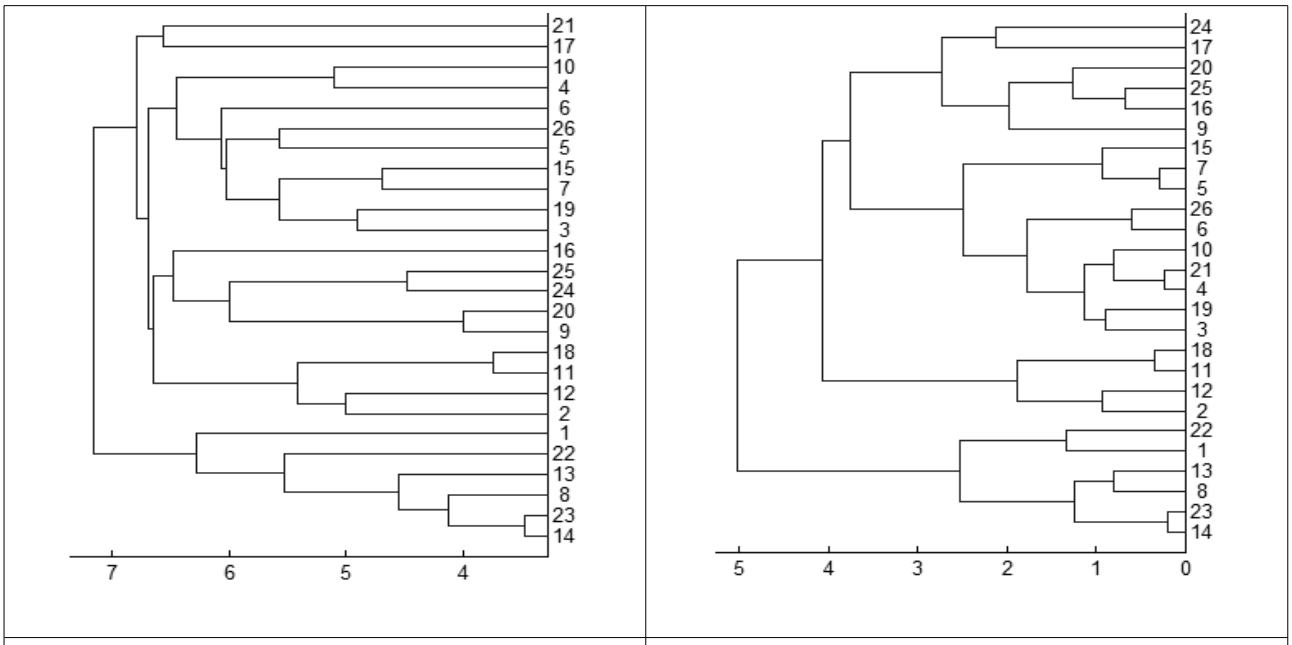
Hierarchical cluster analysis constructs its trees on the basis of relative Euclidean distances between and among vectors; the trees in Figure 5 are identical, so it follows that distance structure has been preserved. This applies, within limits, to projection of a matrix of dimensionality *m* to another of dimensionality *n*, where $n < m$. PCA is extensively used as a linear data dimensionality reduction method. Say, for example, that one wants to reduce the 144-dimensional L to a 50-dimensional one. The result is shown in Figure 6.



a: Cluster analysis of L	b: Cluster analysis of L PCA-reduced to dimensionality 50

Figure 6. Cluster analyses of 144-dimensional L and the PCA-transformed 50-dimensional version.

Again, the trees are identical. But, as noted, there is a limit. PCA is often used to reduce high-dimensionality matrices to dimensionality 2 or 3 for graphical display; reduction to dimensionality 3 is shown in Figure 7.



a: Cluster analysis of L	b: Cluster analysis of L PCA-reduced to dimensionality 3

Figure 7. Cluster analyses of 144-dimensional L and the PCA-transformed 3-dimensional version.

There is a family resemblance between the trees, but they also differ substantially. What has happened to distance preservation is that dimensionality has been reduced to a value lower than the intrinsic dimensionality of the letter data matrix, where intrinsic dimensionality is the minimum number of variables required to represent a given data matrix without significant loss of information ([60], Ch. 3).

PCA provides a criterion for identifying intrinsic dimensionality. One of the ways of calculating PCA is by matrix eigendecomposition which, briefly, works as follows. Given a mean-centred $m \times n$ matrix M, PCA creates a covariance matrix $C = M^T M/(m-1)$ and then abstracts two matrices from C: E_{vect}, whose columns are the eigenvectors of C and constitute the orthogonal basis of a new vector space into which M will be projected, and E_{val}, which is a diagonal matrix containing the eigenvalues of C in descending order of magnitude and which represent the lengths of the new basis vectors, that is, the amount of variance in M that each of the basis vectors represents. It often happens that data is redundant in the sense that the variance of its variables overlaps. The

eigenvalues in E_{val} make it possible to identify such redundancy: the largest eigenvalue and the corresponding eigenvector represent the largest direction of variance in M, the second-largest eigenvalue and the corresponding eigenvector represent the second largest direction of variance in M, and so on to *n*. Plotting the eigenvalues provides an indication of intrinsic dimensionality, that is, of how many mutually orthogonal variables are required to represent the variability in M without significant loss of information. The eigenvalue plot for the covariance matrix abstracted from L is shown in Figure 8.

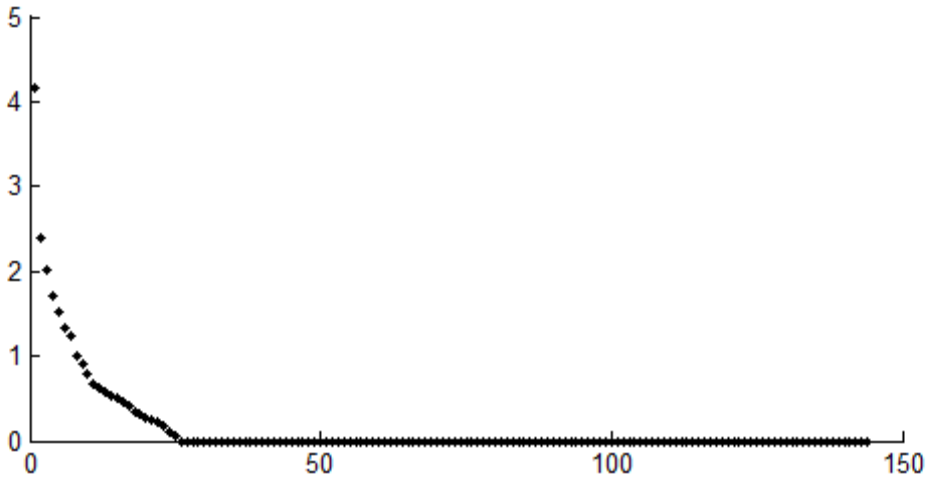


Figure 8. Distribution of E_{val}.

The intrinsic dimensionality is about 25; going below that compromises distance preservation.

The reason for going into all this is that, since Baldi & Hornik made the connection explicit in 1989, the aMLP architecture has been recognized as an implementation of PCA (for example, [57, 61-63]). The significance of this for present purposes is that the homomorphism characteristic of PCA thereby applies also to aMLPs.

An alternative to the eigenvalue decomposition method for calculating PCA is singular value decomposition (SVD; [58, 59, 64]). Explanation of how an aMLP implements PCA is simpler via SVD, and this follows.

SVD is a theorem in linear algebra which says that any real-valued matrix D with *m* rows and *n* columns can be represented as the product of three matrices:

$$D_{m,n} = U_{m,m}S_{m,n}V^{T}_{n,n}$$

where

- U, S, and V are the matrices whose product gives D.
- The column vectors of U are the eigenvectors of the square matrix which results from the multiplication of D by a transposition of itself, that is, DD^{T}, and these constitute an orthonormal basis for the column vectors of D.

- The column vectors of V are the eigenvectors of the square matrix which results from the multiplication of D^T by D, that is, D^TD, and these constitute an orthonormal basis for the row vectors of D.
- S is a diagonal matrix, that is, a matrix having non-negative real values on its main diagonal. These values are the singular values of D in descending order of magnitude, and are the square roots of the eigenvalues of DD^T or D^TD.

When D is a covariance or correlation matrix SVD and PCA are identical. SVA is more general than PCA because it can be applied to matrices of arbitrary dimensions with unrestricted numerical values whereas PCA is restricted to square matrices containing covariances and correlations, but in practice it is a straightforward matter to calculate a covariance or correlation matrix for whatever matrix one wants to analyze, so the choice between SVD and PCA is a matter of preference.

Both eigenvalue decomposition and SVD simply restate the given matrix D in a new vector space having an orthogonal basis and the same dimensionality as D. But one of the main uses of PCA is dimensionality reduction. With the eigendecomposition approach this is achieved by selecting the first *k* largest eigenvalues from E_{val} and deleting all the columns *k*+1...*n* from E_{vect} prior to multiplication by M, that is, ME_{vect}, thereby projecting the original *m*-dimensional matrix into the *k*-dimensional space. The corresponding SVD operation is to select the the first *k* columns from S, yielding S_k, and then to multiply US_k, which results in an $m \times k$ matrix consisting of the largest *k* principal components.

How does all this relate to aMLP architecture? Given D, an aMLP with hidden layer dimensionality *k* approximates the US_k product and V matrices of SVD by using a gradient descent method, back propagation, to optimize the standard mean squared error objective function, which minimizes the difference between the target output and the actual output of the aMLP with respect to D. Once trained, each of the *m* row vectors of D generates a *k*-dimensional hidden layer activation vector, and all *m* hidden layer vectors constitute an $m \times k$ matrix H which is an approximation to the US_k matrix of SVD; for details see Aggarwal ([57]).

There is, however, a caveat. Homomorphism is a concept from linear algebra, and it applies to linear functions of which PCA is one. When the unit activation functions of an aMLP are uniformly linear the homomorphism property of PCA transfers directly [65]. But restricting aMLP architecture in this way confines it to implementation of linear identity functions, and in real-world applications much if not most input data will be nonlinear to some degree (see for example [66]), so aMLP architecture since the time of Baldi & Hornik [65] has incorporated nonlinearity in the form of nonlinear unit activation functions in the hidden layer. The question is: does homomorphism still apply?

Bourlard & Kamp [67] argued that '*for auto-association with linear output units, the optimal weight values can be derived by standard linear algebra, consisting essentially in singular value decomposition (SVD) and making thus the nonlinear functions at the hidden layer completely unnecessary*'; see also Cottrell and Munro [68] to much the same effect. In other words, the presence of nonlinear activation in the hidden layer of an aMLP makes no difference - the aMLP remains a linear system, and as such homomorphism still applies. This seems implausible in the light of proofs by Cybenko [69] and Hornik & Stinchcombe [70] that MLPs with nonlinear hidden

units are universal function approximators - in Cybenko's words, '*that arbitrary decision regions can be arbitrarily well approximated by continuous feedforward neural networks with only a single internal, hidden layer and any continuous sigmoidal nonlinearity*'. Given that data derived from observation of the real world is typically nonlinear to some degree, as noted, an aMLP with nonlinear hidden units will be able to implement the identity function with respect to such data whereas one with a linear hidden layer will not. The implication is that Bourlard & Kamp's claim applies only where the input data is linear or falls within the linear interval of the sigmoid nonlinearity. This is in fact what Japkowicz et al [71] argued and supported with experimental evidence in a paper entitled '*Nonlinear Autoassociation Is Not Equivalent to PCA*'. The question therefore remains: does homomorphism still apply when the activation function of the hidden layer of an aMLP is nonlinear?

The obvious way to test this in the present application is to train an aMLP with a nonlinear hidden layer and a linear output layer using the letter bitmap matrix L, and then to compare the cluster tree for the matrix of hidden layer vectors H generated by the trained net with that for L. The approach needs to be more nuanced than that, however. The number of hidden units *k* in an aMLP, and in artificial neural networks generally, is known to have a strong effect on convergence to whatever function is being implemented, and, traditionally, choice of that number has been heuristic - a heuristic choice of *k* with a negative result with respect to homomorphism doesn't necessarily mean that nonlinear aMLPs fail to preserve homomorphism, because a different choice of *k* might give a positive result. One approach is simply to try a series of random *k*. A more systematic approach, taken here, is as follows:

1. Train the net using a range of hidden layer sizes, say n = 1...200, and then generate the hidden layer matrix H as above.
2. Calculate a matrix D_{hidden} containing the Euclidean distances between all pairs of rows in H.
3. Calculate a matrix D_{input} containing the Euclidean distances between all pairs of rows in L.
4. Row-wise concatenate the values below the main diagonals of D_{hidden} and D_{input} to yield two vectors dv_{hidden} and dv_{input} respectively.
5. Pearson-correlate dv_{hidden} and dv_{input} and save the correlation value in a vector v_{corr}.
6. Plot v_{corr}.

The underlying intuition is that the correlation vector captures the degree of distance structure similarity between L and H. Figure 9 is a plot of correlations from the above sequence applied to L for *k* = 1...200 using the standard sigmoid nonlinearity with range 0...1.

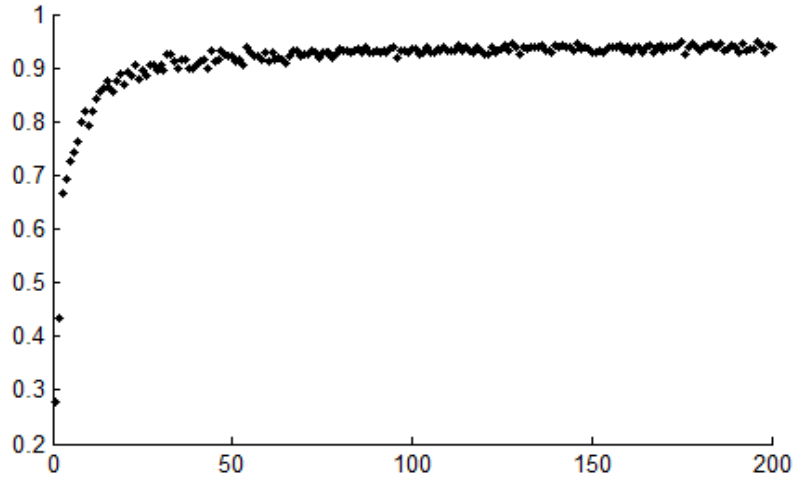
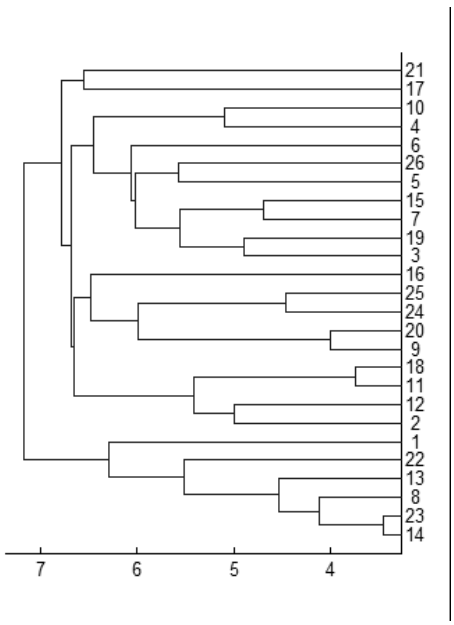
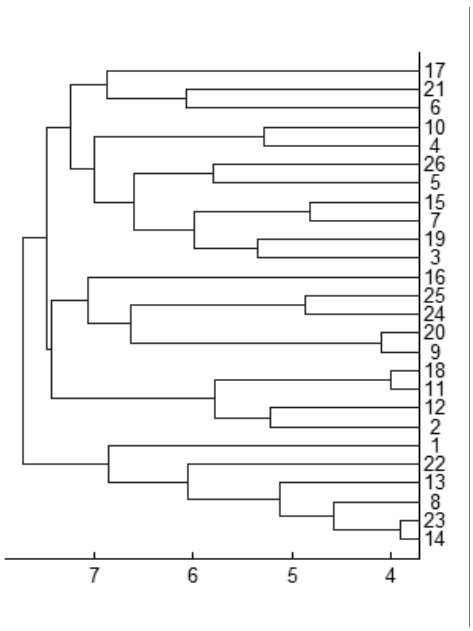


Figure 9. Correlations of input and hidden layer distance vectors for hidden layer sizes 1...200.

Correlation is low for small hidden layer sizes but grows rapidly as the size increases and eventually flatlines at a correlation that fluctuates in the range 0.90...0.95; Spearman correlation gave the same distribution shape with the same numerical range. There is a strong correlation for distance vectors of hidden layer size c.25 onwards, and so the conclusion is that the distance relations in the input data are preserved in the hidden layer for those values of *k*, that is, the aMLP generates a good approximation to homomorphism with respect to preservation of input distance structure in the present application. Figure 10 shows a comparison of the cluster trees for L and H for *k* = 200, and they are virtually identical.



a: L



b: H

Figure 10. Cluster trees (average linkage) for L and H.

Further empirical results indicate that the distance structure preservation of nonlinear single hidden layer aMLPs generalizes. Experiments using the foregoing methodology were conducted using randomly generated binary input matrices with various combinations of the numbers of rows and columns in the range 12-48, keeping these quantities small for tractability. The shape of the correlations was always very similar to that in Figure 9, with maximum correlations in range 0.90...0.98.

That said, it is the case that generalizations based on inductive inference from evidence cannot constitute proof [72]. The results just cited are indicative, but a secure theoretical basis would be useful.

Finally, a single layer aMLP with sigmoid hidden and linear output layer has been shown to be a universal function approximator, as noted, given a sufficient number of hidden units, so this architecture should be all that's required to generate homomorphic representations of environmental input for the model which is the focus of this paper. In practice, however, '*a sufficient number of hidden units*' in any given application may well be found to be very large, and this raises the problem of overfitting, where the neural network learns the input data accurately but generalizes poorly to unseen data from the same input distribution [57]. The response in the neural network research community has been so-called deep learning, whereby multiple hidden layers of relatively small dimensionality replace the single and potentially large single hidden layer in the aMLP, as pioneered by Kramer [73]; see also, for example, Hsieh [74], Scholz et al [62], Goodfellow et al [75], Aggarwal [57]. It remains to be seen whether the hidden layer representations of such deep learning networks are also homomorphic with their inputs.

4. Conclusion

The aim of the foregoing discussion was to show how the homomorphism between the environment and its system-internal representation in a model of intrinsic linguistic intentionality, as implemented by a three-layer autoassociative multilayer perceptron with nonlinear hidden and linear output layers, can be understood mathematically. Existing work which sees such aMLPs as implementations of principal component analysis was cited, the implication of which is that the homomorphism characteristic of linear functions in general applies also to aMLPs. To extend the range of identity functions that can be implemented by an aMLP, however, nonlinear activation functions can be, and in the model in question are, used in the hidden layer; because homomorphism is currently understood as a characteristic of specifically linear functions, its preservation in a nonlinear MLP is not guaranteed. Experimental results were used to show that it is preserved in example applictions, but it was noted that generalizations based on inductive inference from evidence cannot be proof, and that a secure theoretical basis would be useful. The discussion also noted that use of a multilayer 'deep learning' aMLP would address the potential problem of overfitting when a single hidden layer aMLP is used, but that homomorphism in such network would need to be demonstrated.

References

1. Moisl H. Implementation of intrinsic natural language lexical intentionality. Acad Lett. 2021.

Doi: 10.20935/AL117.
2. Moisl H. Dynamical systems implementation of intrinsic sentence meaning. Minds Mach. 2022; 32: 627-653.
3. Jacob P. Intentionality. Stanford: Stanford encyclopedia of philosophy; 2019. Available from: https://plato.stanford.edu/archives/spr2019/entries/intentionality/.
4. Morgan A, Piccinini G. Towards a cognitive neuroscience of intentionality. Minds Mach. 2018; 28: 119-139.
5. Neander K. A mark of the mental: In defense of informational teleosemantics. Cambridge: MIT Press; 2017.
6. Cole D. The Chinese room argument [Internet]. Stanford: Stanford encyclopedia of philosophy; 2020. Available from: https://plato.stanford.edu/archives/spr2020/entries/chinese-room/.
7. Rescorla M. The computational theory of mind [Internet]. Stanford: Stanford encyclopedia of philosophy; 2020. Available from: https://plato.stanford.edu/archives/spr2020/entries/computational-mind/.
8. Oppy G, Dowe D. The Turing test [Internet]. Stanford: Stanford encyclopedia of philosophy; 2016. Available from: https://plato.stanford.edu/archives/spr2019/entries/turing-test/.
9. Piccinini G. The computational theory of cognition. In: Fundamental issues of artificial intelligence. Cham: Springer; 2016. pp. 203-221.
10. Tangirala AK. Principles of system identification: Theory and practice. Boca Raton: CRC Press; 2014.
11. Arbib MA. Brains, machines, and mathematics. 2nd ed. Berlin: Springer Science & Business Media; 1897.
12. O'Connor T. Emergent properties [Internet]. Stanford: Stanford encyclopedia of philosophy; 2020. Available from: https://plato.stanford.edu/archives/fall2020/entries/properties-emergent/.
13. McLaughlin B, Bennett K. Supervenience [Internet]. Stanford: Stanford encyclopedia of philosophy; 2018. Available from: https://plato.stanford.edu/archives/win2018/entries/supervenience/.
14. Stoljar D. Physicalism [Internet]. Stanford: Stanford encyclopedia of philosophy; 2015. Available from: https://plato.stanford.edu/archives/win2017/entries/physicalism/.
15. van Riel R, van Gulick R. Scientific reduction. [Internet]. Stanford: Stanford encyclopedia of philosophy; 2019. Available from: https://plato.stanford.edu/archives/spr2019/entries/scientific-reduction/.
16. Ramsey W. Eliminative materialism [Internet]. Stanford: Stanford encyclopedia of philosophy; 2019. Available from: https://plato.stanford.edu/archives/sum2020/entries/materialism-eliminative/.
17. Speaks J Theories of meaning [Internet]. Stanford: Stanford encyclopedia of philosophy; 2021. Available from: https://plato.stanford.edu/archives/spr2021/entries/meaning/.
18. Moisl H. Intrinsic intentionality and linguistic meaning: An historical outline. In: Words and Numbers--In Memory of Peter Grzybek (1957-2019). Lüdenscheid: RAM-Verlag; 2020. pp. 148-166.
19. Adams F, Aizawa K. Causal theories of mental content [Internet]. Stanford: Stanford encyclopedia of philosophy; 2017. Available from: https://plato.stanford.edu/archives/sum2017/entries/content-causal/.
20. Papineau D. Naturalism [Internet]. Stanford: Stanford encyclopedia of philosophy; 2020. Available from: https://plato.stanford.edu/archives/sum2020/entries/naturalism/.

21. Macleod C. John Stuart Mill [Internet]. Stanford: Stanford encyclopedia of philosophy; 2016. Available from: https://plato.stanford.edu/archives/sum2020/entries/mill/.
22. Patton L. Hermann von Helmholtz [Internet]. Stanford: Stanford encyclopedia of philosophy; 2018. Available from: https://plato.stanford.edu/archives/win2018/entries/hermann-helmholtz/.
23. Pojman P. Ernst Mach [Internet]. Stanford: Stanford encyclopedia of philosophy; 2019. Available from: https://plato.stanford.edu/archives/spr2019/entries/ernst-mach/.
24. Rysiew P. Naturalism in epistemology [Internet]. Stanford: Stanford encyclopedia of philosophy; 2020. Available from: https://plato.stanford.edu/archives/fall2020/entries/epistemology-naturalized/.
25. Bradie M, Harms W. Evolutionary epistemology [Internet]. Stanford: Stanford encyclopedia of philosophy; 2020. Available from: https://plato.stanford.edu/archives/spr2020/entries/epistemology-evolutionary/.
26. Lau J, Deutsch M. Externalism about mental content [Internet]. Stanford: Stanford encyclopedia of philosophy; 2014. Available from: https://plato.stanford.edu/archives/fall2019/entries/content-externalism/.
27. Downes S. Evolutionary psychology [Internet]. Stanford: Stanford encyclopedia of philosophy; 2018. Available from: https://plato.stanford.edu/archives/spr2020/entries/evolutionary-psychology/.
28. Anderson ML. Embodied cognition: A field guide. Artif Intell. 2003; 149: 91-130.
29. Barsalou LW. Grounded cognition: Past, present, and future. Top Cogn Sci. 2010; 2: 716-724.
30. Wilson R, Foglia L. Embodied cognition [Internet]. Stanford: Stanford encyclopedia of philosophy; 2015. Available from: https://plato.stanford.edu/archives/spr2017/entries/embodied-cognition/.
31. Gärdenfors P. Geometry of meaning. Semantics based on conceptual spaces. Cambridge: MIT Press; 2014.
32. Geeraerts D, Cuyckens H. Introducing cognitive linguistics. Oxford: Oxford University Press; 2012.
33. Jackendoff R. Foundations of language: Brain, meaning, grammar, evolution. Oxford: Oxford University Press; 2002.
34. Jackendoff R. A user's guide to thought and meaning. Oxford: Oxford University Press; 2012.
35. Bartels A. Defending the structural concept of representation. Theoria. 2006; 55: 7-19.
36. Churchland PM. Plato's camera: How the physical brain captures a landscape of abstract universals. Cambridge: MIT press; 2012.
37. Gallistel CR. Representations in animal cognition: An introduction. Cognition. 1990; 37: 1-22.
38. Gallistel CR. Learning and representation. In: Learning and memory: A comprehensive reference. New York: Elsevier; 2008. pp. 227-242.
39. Gallistel CR, King AP. Memory and the computational brain: Why cognitive science will transform neuroscience. Chichester: John Wiley & Sons; 2009.
40. Gładziejewski P, Miłkowski M. Structural representations: Causally relevant and different from detectors. Biol Philos. 2017; 32: 337-355.
41. Garagnani M, Pulvermüller F. Conceptual grounding of language in action and perception: A neurocomputational model of the emergence of category specificity and semantic hubs. Eur J Neurosci. 2016; 43: 721-737.
42. Isaac AM. Objective similarity and mental representation. Australas J Educ Technol. 2013; 91: 683-704.

43. Matheson HE, Barsalou LW. Embodiment and grounding in cognitive neuroscience. In: Stevens' handbook of experimental psychology and cognitive neuroscience. 4th ed. New York: John Wiley & Sons; 2018. pp. 1-27.
44. Piccinini G. Computation in physical systems [Internet]. Stanford: Stanford encyclopedia of philosophy; 2017. Available from: https://plato.stanford.edu/archives/sum2017/entries/computation-physicalsystems/.
45. Piccinini G, Bahar S. Neural computation and the computational theory of cognition. Cogn Sci. 2013; 37: 453-488.
46. Piccinini G, Scarantino A. Information processing, computation, and cognition. J Biol Phys. 2011; 37: 1-38.
47. Rescorla M. Cognitive maps and the language of thought. Br J Philos Sci. 2009; 60: 377-407.
48. Rupert RD. Causal theories of mental content. Philos Compass. 2008; 3: 353-380.
49. Shagrir O. The brain as an input–Output model of the world. Minds Mach. 2018; 28: 53-75.
50. Shea N. Consumers need information: Supplementing teleosemantics with an input condition. Philos Phenomenol Res. 2007; 75: 404-435.
51. Shea N. VI—Exploitable isomorphism and structural representation. Proc Aristot Soc. 2014; 114: 123-144.
52. Shea N. Representation in cognitive science. Oxford: Oxford University Press; 2018.
53. Thomson E, Piccinini G. Neural representations observed. Minds Mach. 2018; 28: 191-235.
54. Plebe A, Vivian M. Neurosemantics: Neural processes and the construction of linguistic meaning. Cham: Springer; 2016.
55. Gowers T, Barrow-Green J, Leader I. The Princeton companion to mathematics. Princeton: Princeton University Press; 2008.
56. Weisstein EW. CRC concise encyclopedia of mathematics. 3rd ed. Boca Raton: Chapman and Hall/CRC; 2009.
57. Aggarwal CC. Neural networks and deep learning. Cham: Springer; 2018.
58. Jollife IT. Principal component analysis. New York: Springer; 2002.
59. Jackson JE. A user's guide to principal components. Hoboken: John Wiley & Sons; 2003.
60. Lee JA, Verleysen M. Nonlinear dimensionality reduction. New York: Springer; 2007.
61. Diamantaras KI, Kung SY. Principal component neural networks: Theory and applications. Hoboken: John Wiley & Sons, Inc.; 1996.
62. Scholz M, Fraunholz M, Selbig J. Nonlinear principal component analysis: Neural network models and applications. In: Principal manifolds for data visualization and dimension reduction. Heidelberg: Springer; 2008. pp. 44-67.
63. Qiu J, Wang H, Lu J, Zhang B, Du KL. Neural network implementations for PCA and its extensions. Int Scholarly Res Not. 2012; 2012: 847305.
64. Wall ME, Rechtsteiner A, Rocha LM. Singular value decomposition and principal component analysis. In: A practical approach to microarray data analysis. Boston: Springer; 2003. pp. 91-109.
65. Baldi P, Hornik K. Neural networks and principal component analysis: Learning from examples without local minima. Neural Netw. 1989; 2: 53-58.
66. Bertuglia CS, Vaio F. Nonlinearity, chaos, and complexity: The dynamics of natural and social systems. Oxford: Oxford University Press; 2005.
67. Bourlard H, Kamp Y. Auto-association by multilayer perceptrons and singular value decomposition. Biol Cybern. 1988; 59: 291-294.
68. Cottrell GW, Munro P. Principal components analysis of images via back propagation. In:

Visual Communications and Image Processing'88: Third in a Series. Bellingham: SPIE; 1988. pp. 1070-1077.
69. Cybenko G. Approximation by superpositions of a sigmoidal function. Math Control Signals Syst. 1989; 2: 303-314.
70. Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Netw. 1989; 2: 359-366.
71. Japkowicz N, Hanson SJ, Gluck MA. Nonlinear autoassociation is not equivalent to PCA. Neural Comput. 2000; 12: 531-545.
72. Henderson L. The problem of induction [Internet]. Stanford: Stanford encyclopedia of philosophy; 2020. Available from: https://plato.stanford.edu/archives/spr2020/entries/induction-problem/.
73. Kramer MA. Nonlinear principal component analysis using autoassociative neural networks. AIChE J. 1991; 37: 233-243.
74. Hsieh WW. Nonlinear principal component analysis by neural networks. Tellus A Dyn Meteorol Oceanogr. 2001; 53: 599-615.
75. Bengio Y, Goodfellow I, Courville A. Deep learning. Cambridge: MIT press; 2017.