Talking Cells

The use and epistemic benefits of ‘communication’ in biology and cognitive science

Wiktor Rorot

Institute of Philosophy and Sociology, Polish Academy of Sciences
wiktor.rorot.research@gmail.com

February 6, 2026

Abstract

The concept of “communication” has in recent years been extensively used across a number of fields in life and cognitive sciences to describe a wide-ranging set of phenomena: from coordination in bacterial biofilms and metazoan development, through the operation of the nervous system, to a host of different social behaviours of animals. To date, there is no philosophical appraisal of this concept and its various uses. This paper builds on a large-scale corpus study covering over 1.1 million scientific articles across life and cognitive sciences to analyze the role of “communication” in the conceptual practice of contemporary biology and cognitive science. The results show that different fields use the notion of “communication” to identify similar causal structure in the studied phenomena, indicating a shared meaning of the concept across the different contexts.

Building on this analysis, I explore the epistemic goals “biological communication” serves. The conceptualization of various biological processes in terms of communication allows the researchers to highlight the shared causal organization. In some cases, this feeds directly into their empirical practice. In other cases, this conceptualization enables a novel modelling approach to the phenomena studied. More generally, however, viewing biological processes in terms of “communication”, lowers the complexity of the studied phenomena, allowing the researchers a better grasp of biological systems.

More broadly, conceptual frameworks are usually studied in relation to structured epistemic objects, such as theories or models, yet can also exist independently and function as reliable, though fallible, heuristics and cognitive tools. “Communication” offers insights into how such independent frameworks can directly contribute to the scientific understanding of complex phenomena.

1 INTRODUCTION

The application of the language of the newly-born information theory to biological and psychological research in the 1950s and 1960s has been one of the most significant paradigm shifts in contemporary life sciences (Cartwright et al. 2024; Gatenby and Frieden 2007; Gleick 2011; Kull 2000). Its impact on the empirical and epistemic practices of these disciplines can be likened only to the pervasive molecularization of biological studies (for molecularization, see Chadarevian and Kamminga 2003; Morange 2020). Indeed, the concept of “information” has been extensively discussed by philosophers, across all the major contexts of biological and cognitive research it appears in. Yet, these extensive theoretical accounts have overlooked that alongside "information", biology has adopted the closely related terms of "signalling" and "communication" that soon assumed a life of their own. "Communication" (independently of information) is now routinely invoked in life and cognitive sciences to describe a variety of processes. This includes the coordination in bacterial biofilms (quorum sensing), cell differentiation during metazoan development (bioelectric communication), operation of the nervous system—and its relation to other types of tissues within the organism (neural signalling, gut-brain communication), and a host of different social behaviours of animals—from alarm calls, to the complex games of authority and power in human politics.

This intricate landscape of uses prompts the question: what is communication? Do bacteria and humans really engage in the same process, or is it an (unfortunate) metaphor or analogy? What role does the concept play in the conceptual practice of life and cognitive sciences?

In what follows, I address these questions through a large scale corpus study of scientific literature, encompassing over 1.1 million scientific articles in life and cognitive sciences. The study analyses the meaning of "communication" and the various research practices the concept is involved in. The primary goal of this work is to uncover the epistemic functions of this concept. The structure of the paper is the following: I begin by discussing the data used and methodology of the study (section 2). Next, in section 3, I briefly review existing literature on the concept of "animal communication", which has received philosophical attention. This allows me to construct a set of queries used to find relevant fragments in the corpus. In section 4, I present the results of the corpus analysis, highlighting how the different contexts of the use of "communication" indicate a shared meaning of the concept. Finally, in section 5, I review the evidence looking for markers of the different epistemic functions "communication" serves, highlighting its role both as a description of the phenomena, as an attempt at (mechanistic) explanation, and as an experimentally fruitful abstraction. I conclude in section 6.

2 DATA AND METHODS

The methodology of the study follows the approach of digital philosophy of science (see Lean, Rivelli, and Pence 2023; Miłkowski and Nowakowski 2025; Pence and Ramsey 2018; Pence 2022b, 2025). This novel perspective in the philosophical study of science aims to overcome some known limitations of the dominant methodology of case studies (e.g., Pitt 2001) through the investigation of large corpora of scientific texts (e.g., publications, peer reviews) with computational methods (e.g., Malaterre and Léonard 2024; Miłkowski 2023; Overton 2013; Pence 2022a). Drawing on tools advanced in digital humanities and computational linguistics, proponents of this approach construct mixed-methods or multi-level methodologies (e.g., Murdock et al. 2017; Rorot and Miłkowski 2024), which connect large-scale quantitative analysis of the corpus with qualitative close reading using standard philosophical methodologies (for a detailed discussion of the methodology see Rorot 2025).

The current study uses the Semantic Scholar Open Research Corpus [S2ORC, version published on August 8, 2024; Kinney et al. (2025); Lo et al. (2020);] which features over 12 million full-text, open access articles and is currently the largest openly available corpus of academic papers. The articles often have extensive annotations, including paragraph segmentation. However, as the corpus is produced through automated means, it includes some noise. S2ORC is a part of the larger Semantic Scholar Open Data Platform which provides metadata, including abstracts, for over 200 million scientific publications, as well as pretrained vector embeddings for full texts of the articles (Singh et al. 2023).[1]

Vector embeddings, on which the majority of the methods used here rely, offer a numerical representation of the semantics of a given document. The approach is motivated by a longstanding linguistic view which claims that the meaning of a word can be approximated by looking at its neighbourhood across texts (see e.g., Firth 1957). This "distributional" assumption has been finessed in recent large language models based on the Transformer architecture (LLM, Devlin et al. 2019; OpenAI 2023; Pennington, Socher, and Manning 2014), which introduce complex mathematical transformations and dimensionality reduction methods to optimize vector representations for a number of computational tasks. These models allow for calculating vector representations at arbitrary granularities—from individual words to whole papers.

The general design of the study is presented in Figure 1, and the individual steps are described in greater detail below.

2.1 *Step 1. Subcorpus selection and preparation*

The full S2ORC corpus was filtered to select English language articles that are assigned the discipline of "Biology" or "Psychology" (cognitive science was not among the disciplinary labels used in S2ORC), using the coarse-grained disciplinary assignments from S2ORC. The goal was to limit the noise (i.e., irrelevant uses of "communication", beyond the biological and cognitive context) in the dataset.[2] This resulted in a subcorpus with approx. 2.8 million articles.

The data was further narrowed down to the work published within selected subdisciplines (see Table 1) by filtering by the journal name, based on the Scimago Journal Ranking (SJR) lists assigning individual journals to more fine-grained subject categories (lists for the year 2023 were used).[3] Table~1 summarizes selected research areas and corresponding SJR categories. The selection of disciplines was intended to filter out applied sciences which are outside the scope of the current study (e.g., agriculture or forestry). The division has been primarily motivated by the spatial scales of biological processes that the areas of biology are concerned with. As such, it should be regarded as an operationalization of the concept of "scales" for the purpose of current research, rather than a claim regarding disciplinary distinctions within life and cognitive science.

The filtered subcorpus was segmented into individual sentences using a rule-based approach. For each sentence, for the purpose of semantic search, a vector representation was calculated using a pretrained, general-purpose LLM (the MiniLMv2 model, see Wang et al. 2020; Wang et al. 2021), selected based on the standard benchmarks, bearing in mind computational restrictions of the available hardware (see the technical details in the online supplement).

1. All results of the study, additional versions of figures presented in the paper, as well as the code used for all the analyses reported here is available in the online supplement on OSF: https://osf.io/z342v/overview?view_only=aca8dfbaa299468fbd3ae35472d74ae2. Due to the size of the full dataset, it was not included in the release, but the S2ORC corpus is openly available through the Allen Institute for Artificial Intelligence, its publisher and maintainer, see https://www.semanticscholar.org/product/api.

2. For example, while artificial intelligence research is commonly considered part of cognitive science, the inclusion of the whole discipline of "Computer Science" would result in including papers discussing e.g., computer network protocols, which make heavy use of the terms central to this study ("communication", "signalling", "connection", etc.) but in a distinct, unrelated context—as uncovered in a preliminary investigation of the corpus.

3. The lists are available online at: https://www.scimagojr.com/journalrank.php?year=2023 (accessed October 8, 2024).

Figure 1: Flowchart summarizing the research design. Individual steps are discussed in detail in the main text.

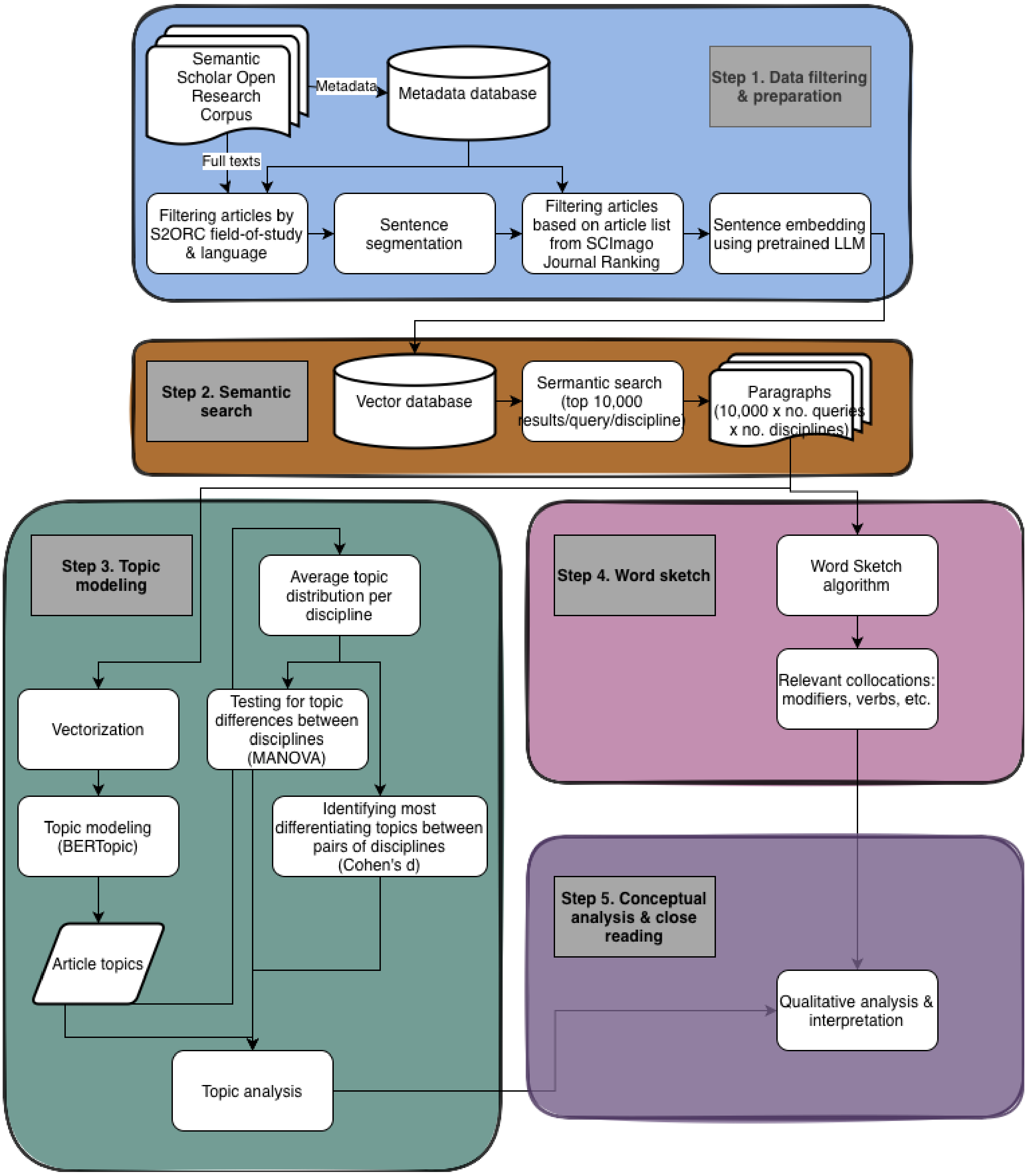


Table 1: Areas within life and cognitive sciences analysed in the study, with the corresponding Scimago Journal Ranking (SJR) labels included in each area. The total number of articles that belongs to at least one of the identified areas is 1,099,571. Note that SJR lists allow for assigning multiple disciplines for a journal, so the article counts do not sum to the total of analysed articles. On average articles have approximately 1.40 discipline assigned, with a median of 1 and a max of 6. Areas and labels are ordered alphabetically.

Area label	Article count (inclusive)	SJR subject categories included
Animal Behaviour	62,628	Animal Science and Zoology Insect Science Small Animals
Developmental Biology	27,215	Developmental Biology
Ecology and Evolution	129,045	Ecological Modeling Ecology Ecology, Evolution, Behavior and Systematics
Generalist journals	67,967	Agricultural and Biological Sciences (miscellaneous)
Microbiology	267,334	Applied Microbiology and Biotechnology Cell Biology Microbiology Microbiology (medical)
Molecular Biology	587,402	Biochemistry Biochemistry, Genetics and Molecular Biology (miscellaneous) Genetics Genetics (clinical) Molecular Biology Molecular Medicine Structural Biology
Neuroscience	153,377	Behavioral Neuroscience Cellular and Molecular Neuroscience Cognitive Neuroscience Developmental Neuroscience Neuroscience (miscellaneous) Sensory Systems
Plant Science	93,934	Plant Science
Psychology	154,651	Developmental and Educational Psychology Experimental and Cognitive Psychology Neuropsychology and Physiological Psychology Psychiatry and Mental Health Psychology (miscellaneous) Social Psychology

2.2 *Step 2. Semantic search*

Using a vector database, I queried the sentence-level embeddings of the subcorpus with the queries based on theoretical definitions of COMMUNICATION in a biological context (see discussion in section 3). For each discipline under investigation (see Table 1) and each query, the search selected top 10,000 best matching sentences from articles (without duplicating results in case a sentence was a top result for more than one query or more than one area of interest). To limit the noise, sentences shorter than 16 characters were omitted from the results (the number was selected arbitrarily, to limit the appearance of short headers and similar fragments containing 2-3 words). The algorithm uses cosine similarity as the similarity score, a standard metric for analysis of similarity between vectors (the score is the cosine of the angle between the vectors in the representation space, with a maximum of 1 for proportional vectors). For each resulting sentence, I also extracted the corresponding paragraph (using annotations from S2ORC or an arbitrary number of 10 neighbouring sentences, where paragraph annotations were not available).

The final number of results depended on the amount of overlap between the queries. For the current study, this was 84,110 sentences and corresponding paragraphs (sample results are presented in subsection B.1, and full results are available in the online supplement). The choice of 10,000 most similar articles for each query and discipline pair is arbitrary. This criterion was chosen to ensure the comparability across queries and disciplines (for example, the average similarity score for the results of each query becomes interpretable in this context) and to provide a sufficient amount of data for further analysis. The alternative would be to select a cut-off point for the similarity score. However, considering the lack of direct interpretability of cosine similarity score values, I judged it to be a more arbitrary approach than the former.

2.3 *Step 3. Topic modelling*

For a large-scale investigation into the references to communication, I conducted topic modelling of the paragraphs selected with semantic search. The study used the BERTopic algorithm (Grootendorst 2022; see the comparison of BERTopic and LDA in Lamirel, Lareau, and Malaterre 2025). Based on vector embeddings for sentences and paragraphs calculated with the MiniLMv2 model in Step 1. Subcorpus selection and preparation, BERTopic uses dimensionality reduction and clustering algorithms (respectively, the uniform manifold projection, UMAP, see McInnes, Healy, and Melville 2020; and hierarchical density-based clustering, HDBSCAN, see Campello, Moulavi, and Sander 2013). This unsupervised procedure identifies semantic clusters in the dataset which are treated as "topics".

For each such topic, BERTopic constructs a representation using class-based term frequency—–inverse document frequency [c-TF—IDF; see Spärck Jones (1972)], an algorithm which selects most characteristic words for each topic. Further, it can construct user-specified representations which take syntactic features into account (I constructed two such representations, "Nouns + modifiers" and "Verbs + modifiers" which selects nouns or verbs with their respective modifiers from the lists generated by c-TF—IDF), or which rely on vector embeddings (using the KeyBERT algorithm, see Grootendorst 2020, with Maximum Marginal Relevance technique to reduce redundancy of the keywords).

The number of topics is selected by the algorithm. It tends to produce fine-grained topics with limited interpretability, but the topics can be reduced into more coarse-grained clusters by merging topics with most similar representations (judged by their vector embeddings). We select the final number of topics using standard coherence metrics, C_v and U_{mass} (see the comparison and discussion in Röder, Both, and Hinneburg 2015), to find a coherence/granularity trade-off. The selection is confirmed through manual, qualitative inspection of the produced topics and documents grouped into those topics. In the current study I selected a model with 50 topics (including outliers).

Finally, BERTopic introduces an additional "outlier" topic which collects documents that did not fit in any of the clusters, and as such is not interpretable. To limit the number of such texts, the algorithm attempts to assign documents initially classified as outliers to their best-fitting topics. However, a number of outlier documents usually remains. These are normally omitted from the analysis.

2.4 *Step 4. Word Sketch*

Fine-grained semantic structure of relevant terms is explored through the Word Sketch algorithm (Kilgarriff et al. 2014). Word Sketch allows for an analysis of the collocations of target words, i.e., words which notably co-occur with the targets, while retaining the syntactic structure. In this way it is possible to explore collocates which perform particular syntactic functions (e.g., are modifiers of the target words, or verbs for which the target is a subject or an object). The Word Sketch algorithm produced lists of top 20 collocates for "communication" from different syntactic categories.

2.5 *Step 5. Conceptual analysis and close reading*

The quantitative data produced in the course of computational analysis, as well as the documents selected in semantic search, are submitted to close reading and conceptual analysis, using standard methods of philosophy of science. This step draws from naturalistically-oriented approach to conceptual analysis (see the essays in Braddon-Mitchell and Nola 2009) and conceptual engineering (Cappelen 2018; cf. Rorot and Miłkowski 2024). There are important limitations to the evidence provided by computational analysis (see Pence 2025) which these qualitative, interpretative procedures can mitigate. Importantly, the analyses conducted in the paper strive to be descriptively accurate of the actual conceptual practice of the use of the notions of interest. Nevertheless, the proposal has normative implications as it reveals the conceptual "norms researchers espouse [and] how they reason about them" (Miłkowski and Nowakowski 2025, p. 9).

3 DEFINING "COMMUNICATION"

The existing philosophical treatments of biological communication focus on the use of the concept in reference to animal behaviour studies, an area of research known as animal communication, and to the evolution of human language. In this context, Thom Scott-Phillips (2008) distinguished between two primary frameworks of viewing commmunication: informational and influence-based.

The former, also known in terms of the "signalling model" (see Stegmann 2016), is primarily developed to bridge the gap between the animal and human communication systems (e.g., Bennett 1979; Green 2007). This model views animal communication as a "more basic" form of human linguistic interactions. Although human communication may require additional conceptual apparatus (at least according to (Bennett 1979), who points towards the framework advanced by H. Paul Grice, e.g., (Grice 1957, 1969, 1995)), it shares with other animal communication systems the foundational role of information transmission. According to the informational view, communication consists of those behaviours which have a function of transmitting information.

The alternative, influence-based view focuses on the adaptive benefits that (putatively) communicative interactions and behaviours confer upon the organisms involved—senders and receivers. This view is exemplified by one of the most widely used definitions of communication, offered initially by John Maynard Smith (2000), and improved upon in his later work with David Harper (Maynard Smith and Harper 2003). These authors define communication through the notion of a "signal"—the vehicle of a communicative interaction, but without the informational implications. For Maynard Smith and Harper (2003, p. 3), signal is "any act or structure which alters the behaviour of other organisms, which evolved because of that effect, and which is effective because the receiver's response has also evolved." This adaptationist view has been preferred by a number of theorists due to the purely syntactic nature of Shannon information (Scott-Phillips 2008) or the apparent difficulties of linking the informational view with evolutionary principles (Owren, Rendall, and Ryan 2010; Rendall, Owren, and Ryan 2009).

The influence-based approach has also been defended more recently within the organizational account of biology (see Moreno and Mossio 2015). Frick, Bich, and Moreno (2019) extend the original, adaptationist focus of the definition proposed by Maynard Smith and Harper, and Scott-Phillips following them, with a broader notion of functional influence, which can be characterized depending on the theoretician's preferred notion of function—avoiding some of the criticisms the neo-Darwinian adaptationism has garnered (most importantly, Gould and Lewontin 1979).

Both informational and influence-based definitions were traditionally developed to capture several paradigmatic instances, including human (*Homo sapiens*) communication, vervet monkeys' (*Chlorocebus pygerythrus*) alarm calls, or honey bees' (*Apis mellifera*) waggle dance, and processes at other scales of biological organization have only been mentioned in passing. Despite the theoretical justifications for distinguishing between the focus on "information" and "influence", the actual practice in animal communication cannot be so neatly divided. One source which clearly indicates this is the list of different definitions of communication collected in the *Historisches Wörterbuch der Biologie* [HWB; Toepfer (2011)] and the associated BioConcepts online database (Toepfer 2024). While these resources are not exhaustive and limited to major positions in the literature, they indicate that definitions adopted by researchers often cut across the theoretical distinctions introduced by Scott-Phillips (2008) (see the selection in Appendix A).

Hence, to track the use of "communication" inclusively, without premature theoretical decisions, I have used both the definitions based on Scott-Phillips' review, and the examples from HWB and BioConcepts, to query the corpus during the semantic search (Step 2. Semantic search). A sample of the queries used is presented in Appendix A. Importantly, while these definitions, as already discussed, have been offered in reference to animal communication, they successfully allow us to pick out the other areas of biological communication of interest in the current study, as highlighted in the following section.

4 RESULTS

4.1 *Topics as patches*

The diversity of contexts in which communication is used across life and cognitive sciences is well visible in the results (see the sample in subsection B.1). To capture this diversity we may consider the different topics identified in the data (see Table 2) through the lens of the "patchwork concepts" approach (Wilson 2006; Haueis 2024). As an attempt to capture the polysemy of various scientific concepts, the patchwork framework proposes to view the different uses of a concept as different patches, defined through "the length scale [i.e., spatiotemporal granularity] of investigation, what technique is involved, the specific domain of application, and which property scientists target" (Haueis 2024, p. 742). As such the patchwork concepts do not have any "core" nor "total meaning", beyond the meanings associated with individual "patches." The notion of a "patch" has been adapted metaphorically by Wilson (2006) to highlight the idea that the individual uses or senses are connected solely by local relations. These local relations may result from the overlaps of the domain or scale at which the two patches are used. Effectively, patches identify different classes of phenomena that are captured with the same term. The terms acquire novel meanings when a "general reasoning strategy" is extended to new cases.

One of the examples (Haueis 2024, p. 749-750) discusses, drawing on his own previous work (Haueis 2021), is that of a *cortical column*. In this case, the general reasoning strategy instructs researchers to look for vertical structures in the brain and then determine whether neurons within those structures have a similar function. However, the techniques for specifying the structure and function differ significantly across spatial scales of neural organization. Columns have been originally analysed at the scale of "hypercolumns" using electrophysiological recordings, with the similarity of function of neurons within the column being approximated via uniformity of the recorded responses (see Hubel and Wiesel 1977). Later the concept has been extended to the neural meso- and microscale. At the microscale, the columns are identified *ex vivo* using Golgi staining, and the functional organization is inferred from the dominance of vertical over horizontal connections between neurons (see Mountcastle 1997). Each of these scale-dependent applications of the CORTICAL COLUMN forms an individual patch.

Indeed, most of the topics identified can be interpreted as picking out a patch-like category of communicative phenomena: identifying the scale, technique, domain, and property. The three most frequent topics refer to different biological systems that extensively rely on electrical or chemical signalling: the nervous (topic 0, NERVOUS SYSTEM) and immune systems (topic 3, IMMUNE SYSTEM), as well as gene networks which control gene expression (topic 1, QUORUM SENSING). Other topics pick out different forms of animal-animal communication (e.g., topics 5—SOCIAL FACTORS IN LANGUAGE, 4—SOCIAL LEARNING, 20—HUMAN INTERSPECIES INTERACTIONS). Second, the topics partially highlight the functional role of various communication systems, picking out the involvement of communication systems in cancerogenesis (topic 3), social organization (topics 1, 8—SOCIAL BEHAVIOURS, and 4, in different ways) or cognition (topic 0). Together, these two dimensions seem to delineate distinct domains of COMMUNICATION.

The scales for different patches of communication emerge from the qualitative assessment of topic similarities based on the similarity matrix (presented in Figure 2, clusters A-E mentioned below are marked in the figure). The topics can be grouped into several clusters: for instance, cluster A includes topics SOCIAL FACTORS IN LANGUAGE (5), COMMUNICATION IN

EDUCATION (14), FAMILY COMMUNICATION (37)—all formulated on the scale of social communication. Similarly, cluster D, which includes topics such as SOCIAL LEARNING (4), HUMAN INTERSPECIES INTERACTIONS (20), or ANIMAL MOVEMENT (9), focuses on the scale of social communication, but extends beyond linguistic interactions. Cluster B, on the other hand, groups topics concerned with the scale of subcellular components involved in signalling—this includes topics QUORUM SENSING (1), PLANT STRESS (21) or GENETIC EVOLUTION (27). A slightly larger scale of cellular processes is grouped by cluster E, which includes topics NOTCH DEVELOPMENT (6), STEM CELLS (18) or MITOCHONDRIA (42). Finally, cluster C captures an intermediate scale of signalling within or between organisms, e.g., in topic IMMUNE SYSTEM (3) or topic PARASITE (34). This clustering is reflected by the topological structure of the topic space, as represented by UMAP 2D-reduced document embeddings (Figure 3), where we see that topics align along a similar structure—from the smallest to largest spatiotemporal scale.[4]

In fact, the topic model, at least in some cases, provides directly the other two components necessary to define patches—techniques, and the properties they target—included either in the representation of the topic produced by BERTopic, or in representative documents. These sometimes occur separately, as in the case of properties without mention of techniques: vocalizations (topic 4), specific chemical substances or biochemical compounds, including calcium (topic 12), pheromones (topic 11), proteins such as Wnt or Notch [^notch] (topic 6). However, they also come packaged together: as in the case of questionnaires (technique) of relationship satisfaction (property; topic 37), models (technique, albeit quite a broad one) of the dynamics of disease spread (property; topic 35), and genome-wide associations (technique) of phenotypic traits (property; topic 27). This "patch" interpretation of topics is presented in subsection B.2.

4.2 *Causal structure of communication*

Despite the differences between the patches, they are clearly picking out a shared causal process, characterized primarily by three components:

1. The involvement of two distinguishable entities, the sender/communicator and the receiver/audience;
2. The transfer of "something", whether it is a message encoded into signal or a piece of evidence for informative intention;
3. The existence of a structure that allows this transfer: for example, in the form of a shared code or the presence of appropriate intentions on both communicating sides.

Further, communication is usually considered as embedded within a broader system, where it serves a particular function—an obvious statement, but one that highlights that instances of communication will be identified as involved in other processes. This can be considered a fourth component of the putative general reasoning strategy, one that has been overlooked by the theoretical accounts discussed before:

4. The functional role of a communicative process.

4. Because of the nature of mathematical operations involved in dimensionality reduction with UMAP, and considering its stochastic nature and the impact of hyperparameter choice, the structure the visualization highlights needs to be interpreted cautiously—hence, here I rely primarily on the parallels between structures highlighted by UMAP and topic similarities.

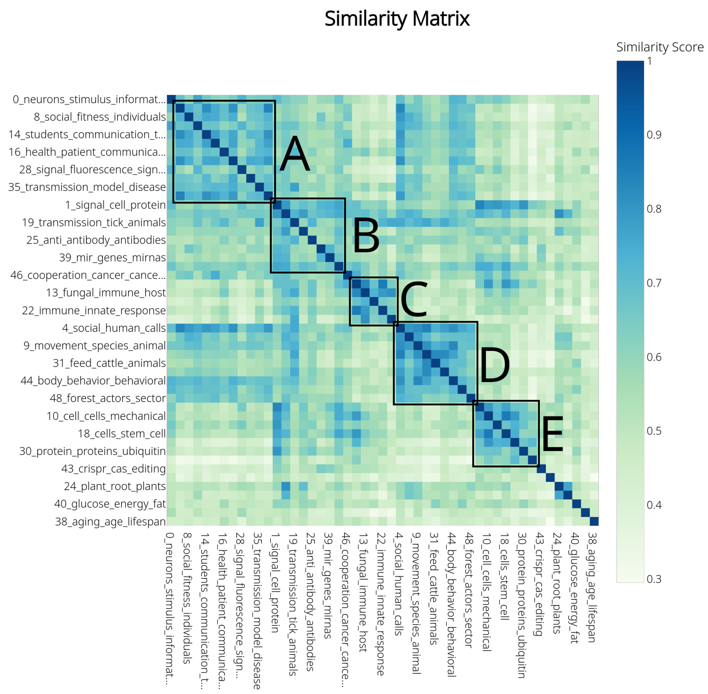
Table 2: 5 most frequent topics with the number of documents classified as belonging to that topic (Count). I have chosen the topic names through qualitative review of the different representations produced by BERTopic and a close reading of the representative documents selected by the algorithm. The table includes the name and the representation produced by KeyBERT, limited to 5 top terms. Full list of topics with complete representations (20 terms for each representation) is available in the online supplementary materials.

Topic	Count	Name	KeyBERT + MMR
0	7161	NERVOUS SYSTEM	stimuli, cortical, neurons, cortex, stimulus
3	6013	IMMUNE SYSTEM	immunotherapy, immune cells, immune response, innate immune, immunity
1	5211	QUORUM SENSING	signaling, gene expression, proteins, bacteria, pathways
2	4634	PLANT PATHOGENS	pathogens, signaling pathways, arabidopsis, biosynthesis, triggered immunity
6	4241	NOTCH DEVELOPMENT	notch signaling, signaling pathway, notch1, wnt signaling, signaling

All components are highlighted by the characteristic collocations of “communication” in the results of semantic search, as identified by the Word Sketch algorithm (Step 4. Word Sketch):

- Component (1) is indicated both by modifiers (“intercellular”, “interpersonal”, “cell” communication) or by nouns and verbs that “communication” modifies (communication “partner” or “network”).
- Component (2) is referred to mainly through modifiers, such as “verbal”, “non-verbal”, “gestural”, “chemical”, “multimodal” or “vocal” communication.
- Component (3) is most directly included in the nouns and verbs modified by “communication”: communication “strategy”, “style”, “method” or “capacity”, which indicate the constraints imposed, communication “system”, “channel”, “technology” or “tool” which highlight the materiality of the structure involved. It is also highlighted by the verbs that take “communication” as an object: they indicate that communication needs to be “mediated”, “facilitated”, “enabled” or “enhanced”; as well as verbs that take it as a subject: communication “requires”, “relies”, “depends” or “needs” a number of preconditions to be met. This spotlights the constraints on the structure of communicative processes.
- Component (4), the function, is indicated by the verbs that have “communication” as a subject: communication can “help”, “play”, “provide”, “allow”, “serve” or “influence” other processes. As such, adjective predicates of “communication” provide an evaluative framework within which communication can be regarded as “important”, “essential” or “crucial” but also “effective” or “successful.” These are rather generic and do not indicate any specific functions, but the broader involvement of communication in a variety of processes.

Figure 2: Topic similarity matrix with qualitatively identified groupings of topics marked. Clusters A, D capture the scale of social interactions, B and E—cellular and subcellular signalling, and C groups topics covering signalling within or between organisms. For fuller description see the text.



The analysis undertaken here arrives at a similar position to the one formulated by Arnon Levy (2011) in his discussion of the use of the notion of "information" in biology. In that context, Levy (2011) advocates for a form of fictionalism with regard to biological information—understood primarily as semantic. In his view, information acts as a "liminal metaphor", where in practice its metaphoric character is backgrounded to highlight the epistemic benefits it brings. But it is metaphoric nonetheless: Levy notes the epistemic hedging (e.g., the use of scare quotes) of ascriptions of semantic properties in the literature and supports this argument by stating that this is "the most natural way to take" information talk (Levy 2011, p. 647) given the widespread use of information across many domains of biology. While perhaps not the strongest argument, Levy's viewpoint is quite understandable, especially when we consider that semantic ascriptions are made indirectly in the literature, even if no explicit hedging is involved.[5] Crucially, as Levy's analysis highlights (2011, pp. 649-650), the metaphor of biological (semantic) information can be beneficial because the information talk picks out a particular causal pattern characterized by:

1) Directionality: the separation of sender and receiver (spatially or temporally), often with a boundary, and the associated direction of influence;
2) Connecting variation: information is invoked to highlight systematic covariation of two components, especially when they are connected through a complex mechanism with many intermediate links;

5. I explore this topic in detail as part of my doctoral dissertation, see Rorot (2025).

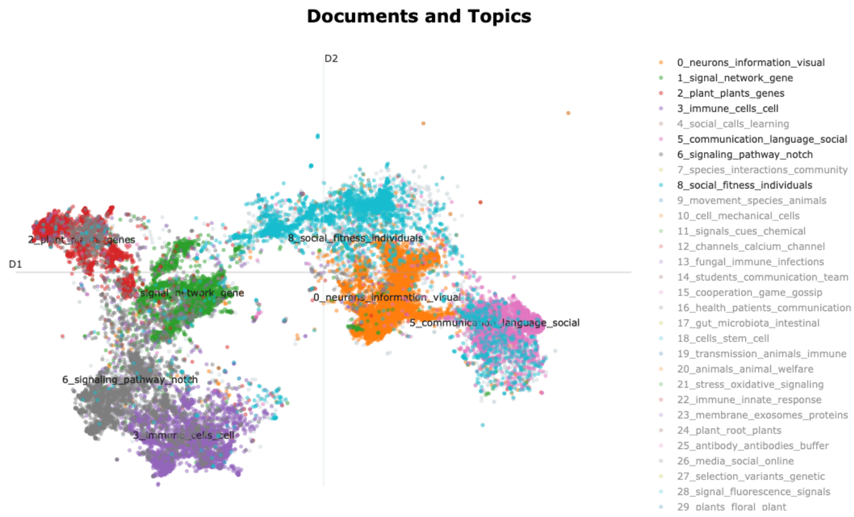


Figure 3: Two-dimensional representation of document vector embeddings, calculated by BERTopic using UMAP algorithm. The documents are clustered by topics based on the topic model. Each point represents a 2D-reduced vector embedding of a document (paragraph). Colors correspond to the topics. A subset of most popular topics is selected, for enhanced readability of the figure—the full, interactive version is available in the online supplement at: https://osf.io/z342v/overview?view_only=aca8dfbaa299468fbd3ae35472d74ae2.

3) Activity of the sender and receiver: though this is metaphorical according to Levy, informational language allows for specifying a particular component of the process as the active sender or receiver, and another—as a passive signal.

This pattern corresponds to the proposed components of communication in an interesting way. It presumes the distinction of a sender and receiver (component 1), and introduces some additional restrictions into what marks and structures (components 2 and 3) can support communication. Crucially, Levy derives these conclusions by tracing how *information* is used in biology, in particular in the context of debates surrounding genetic information (e.g., Godfrey-Smith 2000; Griffiths 2001; Jablonka 2002; Maynard Smith 2000; Sarkar 1996, 2000). The match between the causal structure Levy proposes as the target of informational ascriptions (metaphors, in Levy's view), and the communication structure I have identified, further motivates the identification of a particular causal organization of biological processes as the shared target of the many patches of communication. But it is more than a general reasoning strategy, as the description applies directly in the different contexts, without the need for specifying the components in a technique-involving way—a point that suggests that "communication" is a uniform concept across the many uses, despite the patchwork-like structure.

5 EPISTEMIC GOALS OF "COMMUNICATION"

The remaining question is what epistemic benefits using the concept COMMUNICATION provides, compared with a direct description of the outlined causal organization.

The philosophical literature on scientific concepts, roughly split between the more traditional "semantic" (e.g., Hanson 1958; Feyerabend 1962; Kuhn [1962] 2012) and the "cognitive" approach (e.g., Nersessian 2010; Thagard 1990; Thagard 2012; more recently also Miłkowski 2022; Miłkowski 2023; the division into "semantic" and "cognitive" views follows the review in Cheon and Machery 2016), has focused on the epistemic functions or goals of concepts—as separate from theories or models—only to a limited degree. One exception comes from Brigandt (2010), who has suggested that epistemic goals are a part of the semantic contents of a concept. The goals define a concept along with reference and its inferential role ("a subset of the beliefs scientists have about the term's referent", p. 22), and they enable normative evaluation of conceptual change. Refining this proposal, Cheon (2025) rejects the broadly semantic framework in which Brigandt's proposal was embedded, instead opting for a notion of concepts as "information-complexes". However, he retains the notion of epistemic functions, understanding it in terms of functional meta-information "encoding information of 'for what' the first-order information [that makes up the concept] is used" (p. 378). Rather than considering a separate taxonomy of conceptual functions, both authors tie them to the aims of science (e.g., @ Laudan 1986; Potochnik 2017), or more modestly to the local aims of a particular research programme that relies on a given concept.

As an example, Brigandt (2010) analyses the case of the concept of "gene" as it changed from classical to contemporary molecular genetics (here I follow his discussion in Brigandt 2010, pp. 26-31). Initially, in Mendelian genetics, its epistemic goal was the prediction of patterns of inheritance and the explanation of phenotypic differences between individuals. For the molecular concept established in 1960s and 1970s, after the discovery of DNA, the primary purpose is to explain in causal-mechanistic terms how individual genes bring about their molecular products. These two aims are not independent, as the molecular concept accounts for the development of phenotypic traits, filling in a gap left by the classical concept. For Brigandt, this relates to the different inferential roles of the two concepts, with the molecular one integrating novel empirical findings about the material structure of genes. Yet the change of the inferential role is rational as the epistemic goals of the two concepts are tightly linked.

COMMUNICATION does not seem to fit the bill equally well as "gene" does. While the associated causal structure can figure in causal-mechanistic explanations, whenever "communication" is referenced, the details of that structure are usually implied rather than explicitly considered, as the analysis in previous section shows. In relation to the closely related concept of "information", Oliver Lean (2023) suggests that its primary epistemic function is precisely complementary to that of causal reasoning. His argument, building on interventionist account of causation (Woodward 2004, 2010) and an interpretation of biological practice, is that "information" captures the environmental "interventions" on agents, just as "causation" captures agent's interventions on the world. In his view,

> those [causal and informational—WR] types of reasoning, while in a sense distinct, are intimately related because perception and action are ultimately inseparable: The two jointly constitute an iterative feedback loop, and hence each can only be understood in connection with the other [...]. Nevertheless it is possible and often useful to decompose that overall process and to view it either from the causal or informational perspective; that is, to view a system either as a target of action or as a conduit for the information that guides action. (Lean 2023, p. 110)

In a sense, COMMUNICATION in the view laid out here identifies such action-perception feedback loops of the sort that Lean envisages as a primary biological explanation structure. This aligns well with an emerging view of "scale-free biology" (Fields and Levin 2020), an attempt to extend existing biological syntheses—modern synthesis (Huxley 2010), evo-devo (Gould 1977; Oyama 2000), as well as the extended evolutionary synthesis (Pigliucci and Müller 2010)—into the domain of cognitive and neuroscience through the notions of "information" and "communication". However, evidence from the corpus suggests that such an integrative function is not the main epistemic goal of the concept.

Adopting the (meta)cognitive perspective on science enables us to consider the epistemic goals of concepts not only through the taxonomies of aims of science, but also drawing on the roles that concepts play in cognitive processes. In psychology and cognitive science it is now commonly accepted that there are "top-down" effects in various domains of cognition in which our concepts, previous knowledge, or expectations shape what we perceive and experience (e.g., Clark 2013; Clark 2016; Hohwy 2013; for arguments against this point consider Firestone and Scholl 2016). Such a broadly Kantian perspective (Kant [1781] 1998; see Swanson 2016) suggests that individual concepts and larger conceptual frameworks, understood as networks of interlinked concepts (see also Jabareen 2009), can impact the identification, individuation, and classification of phenomena by scientists.

That is where the main epistemic benefit of COMMUNICATION appears to be located. Accounting for biological phenomena, COMMUNICATION centres the interactive character of the processes (cf. Brancazio 2023; Dingemanse et al. 2023; McGann 2024), at the same time simplifying them by abstracting from their causal-mechanistic details (though these details remain epistemically accessible, as shown in detail in section 4). The unique granularity of description afforded by the concept allows for identifying biological phenomena that otherwise, at a finer or coarser level, would escape the attention of researchers. Although not necessarily emergent in the technical sense, the interplay of complexity (the many elements involved in a communicative exchange) and their strictly "pragmatic" or functional character (component 4 in section 4) posits communicative interactions as unique explanatory or predictive causal nodes. These interactions necessarily have a particular material makeup, but can also be characterized in structural or organizational terms, which motivates the original, analogical application of the term "communication" in the first place.

The cognitive view on science (in particular, the metacognitive version advanced by Miłkowski 2022, 2023) views different forms of scientific representations—theories, models, concepts, and frameworks—as cognitive artifacts. Instead of rejecting the representational function of scientific concepts (as a strongly artifactualist approach, such as the one advanced by (Knuuttila 2005, 2011, 2021), would suggest), the approach views cognitive artifacts following Norman (1991), as used to "maintain, display, or operate upon information in order to serve a representational function and […] affect human cognitive performance" (p. 11). Their content "cannot be fully accounted for in terms of data [and] provide[s] a perspective on the phenomena under investigation" (Miłkowski 2023, p. 187).

The concept of BIOLOGICAL COMMUNICATION as a cognitive tool allows a rendering of biological processes which captures their complexity, at the same time enabling epistemic access for researchers. Consider the growing interest in the role of the body in cognition—both the peripheral nervous system, and non-neural tissues—and the considerations in the following example from the results:

> In an effort to explore how internal state coordinates behaviors, we focus on the body-to-brain direction of communication. We constrained our search to endocrine and sensory systems and a few basic needs over a range of timescales. […]

> This signaling occurs over timescales that might differ from those traditionally used to measure activity at the neural or behavioral levels and will likely reveal novel and exciting new mechanisms of communication between the body and the brain. (235403109, 153)

Here, researchers explicitly indicate that the benefit of the communicative framing is that it enables to identify timescales that more traditional perspective would have missed, leading to uncovering hitherto unknown mechanisms engaged in cognition beyond the brain. In a significantly different context, one example reads: ' > Upon sensing mechanical stimuli, mechanosensitive molecules can initiate or modulate a wide variety of specific intracellular signaling pathways […]. This signaling can lead to diverse series of functional and/or structural responses in the involved cells and tissues[,] the most extensively investigated of these are highlighted in each section of this Review (e.g. shape changes, migratory events and differentiation). (218985273, 126)

Again we see that reference to COMMUNICATION (via the term "signalling") introduces a unique coarse-graining of the studied processes (here on the spatial scale), where identifying a signalling event as such uncovers the relevant components involved in the process (signalling pathways). Furthermore, this example clearly highlights the causal powers ascribed to communicative events, highlighting key points for experimental intervention. Consider a related case of Notch signalling, also involved in development, here in the case of *Drosophila*, the fly:

> Notch signalling in this tissue is not limited to lateral cell contacts: a network of dynamic, actin-based protrusions at the basal side of the epithelium aids signal propagation over longer distances […]. This type of protrusion-mediated signalling […] helps ensure the gradual emergence and refinement of a pattern of well-spaced S[ensory] O[rgan] P[recursor]s. (1774933, 11)

In this passage, the notion of "signalling" enables researchers to select the interesting components of the developmental phenomena, identifying relevant marks and structures (components 2 and 3) and picking out the relevant spatiotemporal scale for the description of the process, the scale at which the relevant causal interactions are taking place.

Importantly, material research tools such as modern microscopes (e.g., X-ray diffraction microscopes or scanning tunneling microscopes, see the analyses in Hacking [1983] 2007; Barad 2007), enable epistemic access to phenomena by interfering with them. Similarly, concepts as cognitive artifacts of science are not neutral and interfere with the available data by highlighting some dependencies and hiding others. Indeed, BIOLOGICAL COMMUNICATION emphasizes the structural or organizational aspect of the processes, while hiding the material (biochemical) details. A description of Notch signalling brackets the complexities involved in just Notch protein-ligand interaction which is *the* signalling event. These elements would be highlighted in a more standard mechanistic description of the process, but such a molecular account would not allow access to the cell-level outcomes: the differentiation and segregation of the cells, etc.

Nevertheless, just as phenomena uncovered by microscopes are real, so is BIOLOGICAL COMMUNICATION. Even if the initial application of communicative perspective is motivated by analogy, as discussed in the previous section, in the contexts identified in the corpus, this analogy can be understood as a general reasoning strategy, and the shared causal structure of the processes it allows to pick out enables particular experimental interventions in a number of cases. For instance, disrupting the bioelectrical signals in *Planaria* impacts their regeneration and induces a significant modification in their body plan, by leading to a generation of multiple heads (see Oviedo et al. 2010; Durant et al. 2017; Cervera et al. 2020). Overall, the communicative conceptualization tracks a robust feature of the studied phenomena (in the sense of Wimsatt 2007).

6 CONCLUSIONS

The purpose of the paper is to evaluate the conceptualization of biological and cognitive processes in terms of COMMUNICATION which has been increasingly common in the scientific literature. To this end, I explored a corpus of approx. 1.1 million scientific articles across life and cognitive science. The corpus study adopts a mixed-methods approach, connecting computational quantitative analysis with qualitative close reading, building on existing methodologies of digital philosophy of science.

The scale of the corpus enables unique insights into broad patterns of use of communicative terms. I explore the various contexts where communication is referenced and on this basis, and suggest that despite the variety of spatiotemporal scales at which communicative processes are posited, it is best to view BIOLOGICAL COMMUNICATION as identifying one kind of process. The process is characterized by a shared causal organization with 4 main components:

(1) the transfer of a *certain mark*,
(2) the involvement of *distinguishable* entities, $and
(3) the presence of a *structure* enabling such transfer.
(4) Communication so understood is embedded within other causal processes, where it performs a particular function.

The epistemic benefits of ECOLOGICAL COMMUNICATION appear to be twofold: first, COMMUNICATION identifies a particular explanation structure, perhaps unique to the explanation of processes within biological systems. In this context, the concept enables an integration of knowledge across domains of biological and cognitive research, similarly to what the proponents of "scale-free biology" (Fields and Levin 2020) synthesis submit. However, this goal, while theoretically consistent with the data found in the corpus, was not explicitly formulated in the analysed articles. Instead, the researchers using the notion appear to be more interested in the ability of COMMUNICATION to capture a particular, causally relevant spatiotemporal scale in the complexity of biological and cognitive phenomena. At that scale, communicative interactions are causal nodes offering epistemic and predictive benefits for researchers. Representing biological phenomena in terms of communication allows identifying opportunities for experimental intervention, and capture inter-level interactions, where lower-level processes (the molecular details of signalling events, for instance) result in higher-level outcomes (cell-level events, such as differentiation or cellular migration in development). The success of such interventions strengthens the view of BIOLOGICAL COMMUNICATION as a robust phenomenon across the scales of biological organization. At the same time, COMMUNICATION offers important theoretical benefits beyond the existing "semiotic" conceptualizations, in particular in terms of INFORMATION or LANGUAGE, as it underscores the processual, interactive character of biological and cognitive processes (e.g., Brancazio 2023; Dingemanse et al. 2023; Nicholson and Dupré 2018).

A APPENDIX 1. QUERIES

A sample of definition-based queries used for semantic search. A full list of queries is available in the online supplementary materials.

Some sentences were slightly modified to reformulate them as directly definitional and to remove ellipses and unnecessary fragments that would impact the vector encoding of sentences.

The IDs encode the database from which the query was drawn (either (Toepfer 2011) or (Toepfer 2024). Similarity score values are average cosine similarities between encoded vectors of the given query and the 10,000 results from the corpus produced by that query, and have been approximated to four decimal places.

1. ID: 0_INFO
   Query: Communication can be defined in terms of information transfer from a sender to a receiver by means of a signal.
   Similarity Score: $\mu = 0.3733$, $\sigma = 0.1099$
2. ID: 1_ADAPT
   Query: Communication can be defined in terms of adaptive influence which a sender exerts on a receiver by means of a signal.
   Similarity Score: $\mu = 0.4157$, $\sigma = 0.0789$
3. ID: 2_BC_DEF1
   Query: Communication is the transmission or exchange of information or knowledge between individual organisms.
   Similarity Score: $\mu = 0.4175$, $\sigma = 0.1170$
4. ID: 3_BC_DEF1_EX1
   Query: Communication between animals involves the giving off by one individual of some chemical or physical signal, that, on being received by another, influences its behaviour.
   Similarity Score: $\mu = 0.4431$, $\sigma = 0.0615$
5. ID: 4_BC_DEF1_EX2
   Query: Communication among animals involves the transmission of information or some other commodity from one participant to another.
   Similarity Score: $\mu = 0.4653$, $\sigma = 0.0808$
6. ID: 13_BC_DEF2
   Query: Communication is the physical connection between the parts of a body.
   Similarity Score: $\mu = 0.4037$, $\sigma = 0.1159$
7. ID: 14_BC_DEF2_EX1
   Query: Highly integrated systems in which a communication system exists between the different molecules; This means that the molecules are informed at every moment of what is going on around them and that they work not in one way but in a highly coordinated manner.
   Similarity Score: $\mu = 0.4250$, $\sigma = 0.0363$
8. ID: 15_HWB_EX1
   Query: Communication necessitates the existence of a code shared between two or more individuals whose use is mutually beneficial to its possessors, i.e. increases fitness.
   Similarity Score: $\mu = 0.3917$, $\sigma = 0.0309$
9. ID: 16_HWB_EX2
   Query: Communication is the transmission of information from one animal to another.
   Similarity Score: $\mu = 0.4175$, $\sigma = 0.0883$

10. ID: 17_HWB_EX3
    QUERY: Communication is the process in which actors use specially designed [i.e. designed by natural selection] signals or displays to modify the behaviour of reactors.
    SIMILARITY SCORE: $\mu = 0.4138$, $\sigma = 0.1061$

B APPENDIX 2. SAMPLE RESULTS

B.1 *Manually selected representative results of semantic search*

A sample results of semantic search. The examples were selected from the results with highest cosine similarity (across all queries) and the query corresponding to that score is indicated below. The examples below were selected qualitatively to show the diversity of the results to the reader, all the analyses reported in the paper were conducted using the full data (available in the online supplementary materials). The reference is provided in parenthesis and includes the CorpusID from S2ORC and the sentence number—I use this convention throughout the article when referring to examples from the corpus analysis, to distinguish them from other citations.

1. QUERY ID: 5; SIMILARITY SCORE: 0.8085; RESEARCH AREA: Animal Behavior
   RESULT: As communication usually occurs in a network of several animals in signalling and receiving range of each other [24], the emotional state of a caller may influence the behaviour of several individuals, addressees in direct interactions and bystanders alike. (5566073, 24)
2. QUERY ID: 20; SIMILARITY SCORE: 0.6442; RESEARCH AREA: Developmental Biology
   RESULT: In sum, RA signaling is a key regulator of anteroposterior patterning orchestrated by Hox codes in at least general ectoderm, central nervous system and endoderm of chordates. (10613111, 174)
3. QUERY ID: 6; SIMILARITY SCORE: 0.6312; RESEARCH AREA: Developmental Biology
   RESULT: Intercellular communication is a key biological process that enables cells to coordinate their responses spatially and temporally to physiological changes. (229344625, 128)
4. QUERY ID: 4; SIMILARITY SCORE: 0.7613; RESEARCH AREA: Ecology and Evolution
   RESULT: Animal communication includes a vast array of signalling systems, ranging from the warning colouration of noxious insects to the complexity of human language. (18825221, 8)
5. QUERY ID: 18; SIMILARITY SCORE: 0.6849; RESEARCH AREA: Microbiology
   RESULT: Specifically, transcallosal information exchange is thought to ensure a good balance between excitation and inhibition of contralateral brain activation during bimanual performance to suppress inadequate mirror movements [3,12,19]. (251734952, 131)
6. QUERY ID: 14; SIMILARITY SCORE: 0.5740; RESEARCH AREA: Molecular Biology
   RESULT: At the molecular level, this has led to the emergence of multiple signaling systems perceived by both interacting parties. (264953331, 9)
7. QUERY ID: 26; SIMILARITY SCORE: 0.7253; RESEARCH AREA: Neuroscience
   RESULT: A channel signal is a linear combination of electrode signals that spatially filters the original multielectrodes recording. (54725834, 36)
8. QUERY ID: 25; SIMILARITY SCORE: 0.6461; RESEARCH AREA: Plant Science

RESULT: This signaling occurs over timescales that might differ from those traditionally used to measure activity at the neural or behavioral levels and will likely reveal novel and exciting new mechanisms of communication between the body and the brain. (235403109, 153)

B.2 *Patches of "communication" corresponding to the 5 most popular topics identified in the dataset*

In some cases, techniques and properties were supplemented based on qualitative analysis of articles assigned to this topic and broader domain knowledge, as these elements were not always readily available from the topic representation and from representative documents, as picked out by BERTopic.

1. TOPIC ID: 0; TOPIC NAME: NERVOUS SYSTEM
   SCALE: Multiple interacting cells; DOMAIN: Neurons, neural populations (cortex, brain); TECHNIQUE: Electrophysiological and imaging methods: single cell recordings, electrical stimulation, two-photon calcium imaging etc.; PROPERTY: Activation pattern
2. TOPIC ID: 3; TOPIC NAME: IMMUNE SYSTEM
   SCALE: Multiple interacting cells; DOMAIN: Immune system: immune and cancer cells, their components, incl. antibodies, systemic and local properties (genetic factors, biochemical compounds and their networked interactions), on the timescale of tumor development; TECHNIQUE: In vitro models (e.g., 3D microfluidic devices), flow cytometry, proximity-based intercellular labeling (e.g., LIPSTIC), RNA sequencing; PROPERTY: Tumor proliferation, patient survival
3. TOPIC ID: 1; TOPIC NAME: QUORUM SENSING
   SCALE: Multiple interacting cells; DOMAIN: Bacterial biofilms, colonies, cell signalling mechanisms (genes, proteins); TECHNIQUE: Microscopy, calcium fluorescent imaging, genetic probing, mathematical modelling (differential equations); PROPERTY: Biofilm growth, cell behaviours, signalling molecule densities
4. TOPIC ID: 2; TOPIC NAME: PLANT PATHOGENS
   SCALE: Individual cells; DOMAIN: Receptors, genetic and protein mechanisms within plant cells; TECHNIQUE: (unclear); PROPERTY: Cell behaviours (e.g., programmed cell death, biochemical changes: protein expressions, hormone production etc.)
5. TOPIC ID: 6; TOPIC NAME: NOTCH DEVELOPMENT
   SCALE: Individual cells; DOMAIN: Morphogenesis, cell development and specialization and their subcellular mechanisms; TECHNIQUE: Immunofluorescence, immunoblotting, genetic methods (reporter lines, knockouts), flow cytometry; PROPERTY: Gene expression, cell viability and specialization

REFERENCES

Barad, Karen. 2007. *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning.* Durham London: Duke University Press.

Bennett, Jonathan Francis. 1979. *Linguistic Behaviour.* London: Cambridge University Press.

Braddon-Mitchell, David, and Robert Nola, eds. 2009. *Conceptual Analysis and Philosophical Naturalism.* Cambridge, Mass: MIT Press.

Brancazio, Nick. 2023. "Interactive Agential Dynamics." *Synthese* 201, no. 6 (June): 221. https://doi.org/10.1007/s11229-023-04192-5.

Brigandt, Ingo. 2010. "The Epistemic Goal of a Concept: Accounting for the Rationality of Semantic Change and Variation." *Synthese* 177, no. 1 (November): 19–40. https://doi.org/10.1007/s11229-009-9623-8.

Campello, Ricardo J. G. B., Davoud Moulavi, and Joerg Sander. 2013. "Density-Based Clustering Based on Hierarchical Density Estimates." In *Advances in Knowledge Discovery and Data Mining,* edited by David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, et al., 7819:160–172. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-37456-2_14.

Cappelen, Herman. 2018. *Fixing Language: An Essay on Conceptual Engineering.* Oxford, United Kingdom ; New York, NY: Oxford University Press.

Cartwright, Julyan H. E., Jitka Čejková, Elena Fimmel, Simone Giannerini, Diego Luis Gonzalez, Greta Goracci, Clara Grácio, et al. 2024. "Information, Coding, and Biological Function: The Dynamics of Life." *Artificial Life* 30, no. 1 (February): 16–27. https://doi.org/10.1162/artl_a_00432.

Cervera, Javier, Salvador Meseguer, Michael Levin, and Salvador Mafe. 2020. "Bioelectrical Model of Head-Tail Patterning Based on Cell Ion Channels and Intercellular Gap Junctions." *Bioelectrochemistry* 132 (April): 107410. https://doi.org/10.1016/j.bioelechem.2019.107410.

Chadarevian, Soraya De, and Harmke Kamminga. 2003. *Molecularizing Biology and Medicine.* Taylor & Francis, September. https://doi.org/10.4324/9780203304235.

Cheon, Hyundeuk. 2025. "The Function of Scientific Concepts." *Foundations of Science* 30, no. 2 (June): 371–385. https://doi.org/10.1007/s10699-023-09929-6.

Cheon, Hyundeuk, and Edouard Machery. 2016. "Scientific Concepts." In *The Oxford Handbook of Philosophy of Science,* edited by Paul Humphreys. Oxford Handbooks Online. Oxford University Press, May. https://doi.org/10.1093/oxfordhb/9780199368815.013.9.

Clark, Andy. 2013. "Whatever next? Predictive Brains, Situated Agents, and the Future of Cognitive Science." *Behavioral and Brain Sciences* 36, no. 03 (June): 181–204. https://doi.org/10.1017/S0140525X12000477.

———. 2016. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind.* Oxford ; New York: Oxford University Press.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,* arXiv:1810.04805, May. Accessed January 27, 2023.

Dingemanse, Mark, Andreas Liesenfeld, Marlou Rasenberg, Saul Albert, Felix K. Ameka, Abeba Birhane, Dimitris Bolis, et al. 2023. "Beyond Single-Mindedness: A Figure-Ground Reversal for the Cognitive Sciences." *Cognitive Science* 47, no. 1 (January): e13230. https://doi.org/10.1111/cogs.13230.

Durant, Fallon, Junji Morokuma, Christopher Fields, Katherine Williams, Dany Spencer Adams, and Michael Levin. 2017. "Long-Term, Stochastic Editing of Regenerative Anatomy via Targeting Endogenous Bioelectric Gradients." *Biophysical Journal* 112, no. 10 (May): 2231–2243. https://doi.org/10.1016/j.bpj.2017.04.011.

Feyerabend, Paul K. 1962. "Explanation, Reduction, and Empiricism." In *Scientific Explanation, Space and Time,* edited by Herbert Feigl and Grover Maxwell, 28–97. Minnesota Studies in the Philosophy of Science 3. Minneapolis, MN: University of Minnesota Press.

Fields, Chris, and Michael Levin. 2020. "Scale-Free Biology: Integrating Evolutionary and Developmental Thinking." *BioEssays* 42, no. 8 (August): 1900228. https://doi.org/10.1002/bies.201900228.

Firestone, Chaz, and Brian J. Scholl. 2016. "Cognition Does Not Affect Perception: Evaluating the Evidence for "Top-down" Effects." *Behavioral and Brain Sciences* 39:e229. https://doi.org/10.1017/S0140525X15000965.

Firth, John Rupert. 1957. "A Synopsis of Linguistic Theory 1930-55." In *Studies in Linguistic Analysis,* 1–32. Publications of the Philological Society. Oxford: Blackwell.

Frick, Ramiro, Leonardo Bich, and Alvaro Moreno. 2019. "An Organisational Approach to Biological Communication." *Acta Biotheoretica* 67, no. 2 (June): 103–128. https://doi.org/10.1007/s10441-019-09342-2.

Gatenby, Robert A., and B. Roy Frieden. 2007. "Information Theory in Living Systems, Methods, Applications, and Challenges." *Bulletin of Mathematical Biology* 69, no. 2 (February): 635–657. https://doi.org/10.1007/s11538-006-9141-5.

Gleick, James. 2011. *The Information: A History, a Theory, a Flood.* New York: Pantheon Books.

Godfrey-Smith, Peter. 2000. "Information, Arbitrariness, and Selection: Comments on Maynard Smith." *Philosophy of Science* 67 (2): 202–207.

Gould, Stephen J., and Richard C. Lewontin. 1979. "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme." *Proceedings of the Royal Society of London. Series B. Biological Sciences* 205, no. 1161 (September): 581–598. https://doi.org/10.1098/rspb.1979.0086.

Gould, Stephen Jay. 1977. *Ontogeny and Phylogeny.* Cambridge, Mass: Belknap Press of Harvard University Press.

Green, Mitchell S. 2007. *Self-Expression.* Oxford: Oxford University Press.

Grice, H. Paul. 1957. "Meaning." *The Philosophical Review* 66, no. 3 (July): 377. https://doi.org/10.2307/2182440.

———. 1969. "Utterer's Meaning and Intention." *The Philosophical Review* 78, no. 2 (April): 147. https://doi.org/10.2307/2184179.

———. 1995. *Studies in the Way of Words.* Cambridge, Mass.: Harvard Univ. Pr.

Griffiths, Paul E. 2001. "Genetic Information: A Metaphor in Search of a Theory." *Philosophy of Science* 68 (3): 394–412.

Grootendorst, Maarten. 2020. *KeyBERT: Minimal Keyword Extraction with BERT.* Zenodo. V. v0.1.3. https://doi.org/10.5281/zenodo.4461265.

———. 2022. "BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure." *arXiv preprint arXiv:2203.05794.*

Hacking, Ian. (1983) 2007. *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science.* Cambridge: Cambridge Univ. Press. https://doi.org/10.1017/CBO9780511814563.

Hanson, Norwood Russell. 1958. *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science.* Cambridge: Cambridge Univ. Pr.

Haueis, Philipp. 2021. "The Death of the Cortical Column? Patchwork Structure and Conceptual Retirement in Neuroscientific Practice." *Studies in History and Philosophy of Science Part A* 85 (February): 101–113. https://doi.org/10.1016/j.shpsa.2020.09.010.

———. 2024. "A Generalized Patchwork Approach to Scientific Concepts." *The British Journal for the Philosophy of Science* 75, no. 3 (September): 741–768. https://doi.org/10.1086/716179.

Hohwy, Jakob. 2013. *The Predictive Mind.* Oxford, United Kingdom ; New York, NY, United States of America: Oxford University Press.

Hubel, David H., and Torsten N. Wiesel. 1977. "Ferrier Lecture - Functional Architecture of Macaque Monkey Visual Cortex." *Proceedings of the Royal Society of London. Series B. Biological Sciences* 198, no. 1130 (July): 1–59. https://doi.org/10.1098/rspb.1977.0085.

Huxley, Julian. 2010. *Evolution: The Modern Synthesis.* Cambridge, Mass: MIT Press.

Jabareen, Yosef. 2009. "Building a Conceptual Framework: Philosophy, Definitions, and Procedure." *International Journal of Qualitative Methods* 8, no. 4 (December): 49–62. https://doi.org/10.1177/160940690900800406.

Jablonka, Eva. 2002. "Information: Its Interpretation, Its Inheritance, and Its Sharing." *Philosophy of Science* 69, no. 4 (December): 578–605. https://doi.org/10.1086/344621.

Kant, Immanuel. (1781) 1998. *Critique of Pure Reason.* Edited by Paul Guyer and Allen W. Wood. Cambridge ; New York: Cambridge University Press.

Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, and Vít Suchomel. 2014. "The Sketch Engine: Ten Years On." *Lexicography* 1, no. 1 (July): 7–36. https://doi.org/10.1007/s40607-014-0009-9.

Kinney, Rodney, Chloe Anastasiades, Russell Authur, Iz Beltagy, Jonathan Bragg, Alexandra Buraczynski, Isabel Cachola, et al. 2025. *The Semantic Scholar Open Data Platform,* 2301.10140, April. Accessed June 20, 2025. https://doi.org/10.48550/arXiv.2301.10140.

Knuuttila, Tarja. 2005. "Models, Representation, and Mediation." *Philosophy of Science* 72 (5): 1260–1271. https://doi.org/10.1086/508124.

———. 2011. "Modelling and Representing: An Artefactual Approach to Model-Based Representation." *Model-Based Representation in Scientific Practice* 42, no. 2 (June): 262–271. https://doi.org/10.1016/j.shpsa.2010.11.034.

———. 2021. "Epistemic Artifacts and the Modal Dimension of Modeling." *European Journal for Philosophy of Science* 11, no. 3 (September): 65. https://doi.org/10.1007/s13194-021-00374-5.

Kuhn, Thomas S. (1962) 2012. *The Structure of Scientific Revolutions.* Edited by Ian Hacking. Chicago ; London: The University of Chicago Press.

Kull, Kalevi. 2000. "Trends in Theoretical Biology: The 20th Century." *Aquinas* 43 (2): 235–249.

Lamirel, Jean-Charles, Francis Lareau, and Christophe Malaterre. 2025. "Old but Not Obsolete: Bag-of-Words vs. Embeddings in Topic Modeling." In *20th International Conference on Scientometrics & Informetrics.* July. https://doi.org/10.51408/issi2025_127.

Laudan, Larry. 1986. *Science and Values: The Aims of Science and Their Role in Scientific Debate.* v.3. Berkeley: University of California Press.

Lean, Oliver M. 2023. "Ideal Observations. Information and Causation in Biological Practice." In *From Biological Practice to Scientific Metaphysics,* edited by William C. Bausman, Janella K. Baxter, and Oliver M. Lean, 83–115. Minnesota Studies in the Philosophy of Science 23. Minneapolis, MN: University of Minnesota Press.

Lean, Oliver M., Luca Rivelli, and Charles H. Pence. 2023. "Digital Literature Analysis for Empirical Philosophy of Science." *The British Journal for the Philosophy of Science* 74, no. 4 (December): 875–898. https://doi.org/10.1086/715049.

Levy, Arnon. 2011. "Information in Biology: A Fictionalist Account." *Noûs* 45 (4): 640–657.

Lo, Kyle, Lucy Lu Wang, Mark Neumann, Rodney Kinney, and Dan S. Weld. 2020. *S2ORC: The Semantic Scholar Open Research Corpus,* arXiv:1911.02782, July. Accessed August 30, 2024.

Malaterre, Christophe, and Martin Léonard. 2024. "Epistemic Markers in the Scientific Discourse." *Philosophy of Science* 91, no. 1 (January): 151–174. https://doi.org/10.1017/psa.2023.97.

Maynard Smith, John. 2000. "The Concept of Information in Biology." *Philosophy of Science* 67, no. 2 (June): 177–194. https://doi.org/10.1086/392768.

Maynard Smith, John, and David Harper. 2003. *Animal Signals.* New York: Oxford University Press.

McGann, Marek. 2024. “Facing Life: The Messy Bodies of Enactive Cognitive Science.” *Phenomenology and the Cognitive Sciences* (January). https://doi.org/10.1007/s11097-024-09958-x.

McInnes, Leland, John Healy, and James Melville. 2020. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction,* arXiv:1802.03426, September. Accessed June 20, 2025. https://doi.org/10.48550/arXiv.1802.03426.

Miłkowski, Marcin. 2022. “Cognitive Artifacts and Their Virtues in Scientific Practice.” *Studies in Logic, Grammar and Rhetoric* 67, no. 3 (December): 219–246. https://doi.org/10.2478/slgr-2022-0012.

———. 2023. “Cognitive Metascience: A New Approach to the Study of Theories.” *Przegląd Psychologiczny* 66, no. 1 (October): 185–207. https://doi.org/10.31648/przegldpsychologiczny.9682.

Miłkowski, Marcin, and Przemysław Nowakowski. 2025. “Tracking Norms: A Plea for a Normative Digital Philosophy of Science.” *Synthese* 206, no. 3 (August): 120. https://doi.org/10.1007/s11229-025-05200-6.

Morange, Michel. 2020. *The Black Box of Biology: A History of the Molecular Revolution.* Translated by Matthew Cobb. Cambridge: Harvard University Press.

Moreno, Alvaro, and Matteo Mossio. 2015. *Biological Autonomy.* Vol. 12. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-9837-2.

Mountcastle, V. 1997. “The Columnar Organization of the Neocortex.” *Brain* 120, no. 4 (April): 701–722. https://doi.org/10.1093/brain/120.4.701.

Murdock, Jaimie, Colin Allen, Katy Börner, Robert Light, Simon McAlister, Andrew Ravenscroft, Robert Rose, et al. 2017. “Multi-Level Computational Methods for Interdisciplinary Research in the HathiTrust Digital Library.” *PLOS ONE* 12, no. 9 (September): e0184188. https://doi.org/10.1371/journal.pone.0184188.

Nersessian, Nancy J. 2010. *Creating Scientific Concepts.* Cambridge, Mass.: MIT Press.

Nicholson, Daniel J., and John Dupré, eds. 2018. *Everything Flows: Towards a Processual Philosophy of Biology.* Oxford, United Kingdom: Oxford University Press.

Norman, Donald A. 1991. “Cognitive Artifacts.” In *Designing Interaction: Psychology at the Human-Computer Interface.* Edited by John M. Carroll, 17–38. Cambridge Series on Human-Computer Interaction, No. 4. New York, NY, US: Cambridge University Press.

OpenAI. 2023. *GPT-4 Technical Report,* arXiv:2303.08774, March. Accessed May 25, 2023.

Overton, James A. 2013. ““Explain” in Scientific Discourse.” *Synthese* 190, no. 8 (May): 1383–1405. https://doi.org/10.1007/s11229-012-0109-8.

Oviedo, Néstor J., Junji Morokuma, Peter Walentek, Ido P. Kema, Man Bock Gu, Joo-Myung Ahn, Jung Shan Hwang, Takashi Gojobori, and Michael Levin. 2010. “Long-Range Neural and Gap Junction Protein-Mediated Cues Control Polarity during Planarian Regeneration.” *Developmental Biology* 339, no. 1 (March): 188–199. https://doi.org/10.1016/j.ydbio.2009.12.012.

Owren, Michael J., Drew Rendall, and Michael J. Ryan. 2010. “Redefining Animal Signaling: Influence versus Information in Communication.” *Biology & Philosophy* 25, no. 5 (November): 755–780. https://doi.org/10.1007/s10539-010-9224-4.

Oyama, Susan. 2000. *The Ontogeny of Information: Developmental Systems and Evolution.* Durham, NC: Duke Univ. Press.

Pence, Charles H. 2022a. “How Not to Fight About Theory: The Debate Between Biometry and Mendelism in ”Nature”, 1890–1915.” In *The Dynamics of Science: Computational Frontiers in History and Philosophy of Science,* edited by Grant Ramsey and Andreas De Block. Pittsburgh, Pa: University of Pittsburgh press.

———. 2022b. “Testing and Discovery: Responding to Challenges to Digital Philosophy of Science.” *Metaphilosophy* 53, nos. 2-3 (April): 238–253. https://doi.org/10.1111/meta.12549.

Pence, Charles H. 2025. "Textual Analysis and Conceptual Cartography." In *Methods in Philosophy of Science: A User's Guide,* edited by Adrian Currie and Sophie Veigl. MIT Press.

Pence, Charles H., and Grant Ramsey. 2018. "How to Do Digital Philosophy of Science." *Philosophy of Science* 85, no. 5 (December): 930–941. https://doi.org/10.1086/699697.

Pennington, Jeffrey, Richard Socher, and Christopher Manning. 2014. "Glove: Global Vectors for Word Representation." In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP),* 1532–1543. Doha, Qatar: Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1162.

Pigliucci, Massimo, and Gerd Müller. 2010. "Elements of an Extended Evolutionary Synthesis." In *Evolution, the Extended Synthesis,* edited by Massimo Pigliucci and Gerd Müller, 3–17. Cambridge, Mass: MIT Press.

Pitt, Joseph C. 2001. "The Dilemma of Case Studies: Toward a Heraclitian Philosophy of Science." *Perspectives on Science* 9, no. 4 (December): 373–382. https://doi.org/10.1162/106361401760375785.

Potochnik, Angela. 2017. *Idealization and the Aims of Science.* Chicago: The University of Chicago Press.

Rendall, Drew, Michael J. Owren, and Michael J. Ryan. 2009. "What Do Animal Signals Mean?" *Animal Behaviour* 78, no. 2 (August): 233–240. https://doi.org/10.1016/j.anbehav.2009.06.007.

Röder, Michael, Andreas Both, and Alexander Hinneburg. 2015. "Exploring the Space of Topic Coherence Measures." In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining,* 399–408. Shanghai China: ACM, February. https://doi.org/10.1145/2684822.2685324.

Rorot, Wiktor. 2025. "Scale-Free Communication? An Investigation of the Use of the Concept "Communication" in Biology and Cognitive Sciences." PhD diss., University of Warsaw, December.

Rorot, Wiktor, and Marcin Miłkowski. 2024. "Empirical Evidence in Conceptual Engineering, or the Defense of 'Predictive Understanding'." In *Conceptual Engineering,* edited by Piotr Stalmaszczyk, 255–280. Brill | mentis, September. https://doi.org/10.30965/9783969753026_013.

Sarkar, Sahotra. 1996. "Biological Information: A Skeptical Look at Some Central Dogmas of Molecular Biology." *Boston Studies in the Philosophy of Science* 183:187–232.

———. 2000. "Information in Genetics and Developmental Biology: Comments on Maynard Smith." *Philosophy of Science* 67 (2): 208–213. https://doi.org/10.1086/392771.

Scott-Phillips, Thom. 2008. "Defining Biological Communication." *Journal of Evolutionary Biology* 21, no. 2 (March): 387–395. https://doi.org/10.1111/j.1420-9101.2007.01497.x.

Singh, Amanpreet, Mike D'Arcy, Arman Cohan, Doug Downey, and Sergey Feldman. 2023. *SciRepEval: A Multi-Format Benchmark for Scientific Document Representations,* arXiv:2211.13308, November. Accessed December 8, 2025. https://doi.org/10.48550/arXiv.2211.13308.

Spärck Jones, Karen. 1972. "A Statistical Interpretation of Term Specificity and Its Application in Retrieval." *Journal of Documentation* 28, no. 1 (January): 11–21. https://doi.org/10.1108/eb026526.

Stegmann, Ulrich E. 2016. "The Philosophy of Communication and Information." In *The Routledge Handbook of Philosophy of Information,* edited by Luciano Floridi. Routledge Handbooks in Philosophy. London ; New York: Routledge, Taylor & Francis Group.

Swanson, Link R. 2016. "The Predictive Processing Paradigm Has Roots in Kant." *Frontiers in Systems Neuroscience* 10 (October). https://doi.org/10.3389/fnsys.2016.00079.

Thagard, Paul. 1990. "Concepts and Conceptual Change." *Synthese* 82, no. 2 (February): 255–274. https://doi.org/10.1007/BF00413664.

———. 2012. *The Cognitive Science of Science: Explanation, Discovery, and Conceptual Change.* Cambridge (Mass.): MIT press.

Toepfer, Georg. 2011. *Historisches Wörterbuch der Biologie: Geschichte und Theorie der biologischen Grundbegriffe. Band 2: Gefühl — Organismus.* Stuttgart: J.B. Metzler. https://doi.org/10.1007/978-3-476-00455-0.

Toepfer, Georg. 2024. *BioConcepts: The Origin and Definition of Biological Concepts - A Multilingual Database.* Accessed November 15, 2024.

Wang, Wenhui, Hangbo Bao, Shaohan Huang, Li Dong, and Furu Wei. 2021. *MiniLMv2: Multi-Head Self-Attention Relation Distillation for Compressing Pretrained Transformers,* arXiv:2012.15828, June. Accessed February 5, 2026. https://doi.org/10.48550/arXiv.2012.15828.

Wang, Wenhui, Furu Wei, Li Dong, Hangbo Bao, Nan Yang, and Ming Zhou. 2020. *MiniLM: Deep Self-Attention Distillation for Task-Agnostic Compression of Pre-Trained Transformers,* arXiv:2002.10957, April. Accessed February 5, 2026. https://doi.org/10.48550/arXiv.2002.10957.

Wilson, Mark. 2006. *Wandering Significance: An Essay on Conceptual Behavior.* Oxford: Clarendon Press.

Wimsatt, William C. 2007. *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality.* Cambridge, Mass: Harvard University Press.

Woodward, James. 2004. *Making Things Happen: A Theory of Causal Explanation.* Oxford University PressNew York, January. https://doi.org/10.1093/0195155270.001.0001.

———. 2010. "Causation in Biology: Stability, Specificity, and the Choice of Levels of Explanation." *Biology & Philosophy* 25, no. 3 (June): 287–318. https://doi.org/10.1007/s10539-010-9200-z.