Talking Cells: the unity of ‘communication’ from bacteria to brains

Wiktor Rorot

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

July 10, 2026

Abstract

The concept of “communication” has in recent years been applied across the life and cognitive sciences to a wide-ranging set of phenomena: from coordination in bacterial biofilms and metazoan development, through the operation of the nervous system, to the social behaviours of animals. No philosophical appraisal of this use of the concept yet exists.

This paper analyses the term, on the basis of a corpus study of over 1.1 million scientific articles. For concepts referring to such diverse phenomena, the expected verdict is pluralist, with a distinct meaning of the term for each domain of application, and limited, metaphorical or analogical resemblance between them. The evidence points the other way. Across domains, “communication” is used to pick out one and the same causal organization, which indicates a shared core meaning rather than a family of loosely related senses.

The term effectively serves as a mechanism schema, providing an abstract description of a mechanism type. The organization consists of four components—a sender and a receiver, a transferred mark, an enabling structure, and a function—which are instantiated differently across fields. The principal epistemic benefit is coarse-graining of biological phenomena. Representing a process as communication fixes a causally relevant grain of description, one at which points of experimental intervention become visible and implementation detail remains available. Additionally, the abstract character of “communication” allows the concept to serve in the integration of biological research across scales and domains, enhancing scientific understanding of life and mind.

1 INTRODUCTION

The notion of COMMUNICATION is increasingly applied to a diverse range of phenomena in biological and cognitive research. It is used for intracellular signalling in gene regulatory networks, 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, or even microbiome-gut-brain communication). At the other end of the scale it describes the social behaviours of animals, from alarm calls, to the complex games of authority and power in human politics.

This widespread use of COMMUNICATION is bound with the place that information theory has come to play in contemporary biological and psychological research, since its inception in the 1950s and 1960s (Cartwright et al. 2024; Gatenby and Frieden 2007; Gleick 2011; Kull 2000). The connection shows in the title of Shannon's founding work, *A Mathematical Theory of Communication* (Shannon 1948; Shannon and Weaver 1964). Yet where INFORMATION has become one of the central topics of philosophy of biology (e.g., Anta and Sánchez-Dorado 2025; Bartlett et al. 2025; Jablonka 2002; Godfrey-Smith 2007; Maynard Smith 2000; Oyama 2000), its close relatives COMMUNICATION or SIGNALLING have gone almost unexamined.

In the current paper, I offer a philosophical appraisal of the concept of BIOLOGICAL COMMUNICATION across its wide-ranging uses. The primary questions in this context are: what is communication? Do bacteria and human beings engage in one process, or is the shared term a metaphor, perhaps an unfortunate one? What work does the concept do in the conceptual practice of life and cognitive sciences?

For concepts spread this widely, the ready answer is pluralist (e.g., Chang 2012; Ludwig and Ruphy 2026): a term used across such different domains is associated with a distinct concept in each, and any resemblance among them holds only by analogy or metaphor (e.g., Gentner and Jeziorski 1993; Hesse 1966) or by the transfer of a model template (Knuuttila 2021). The framework of *patchwork concepts* (Wilson 2006; Haueis 2024) gives this its sharpest form: a concept is a patchwork of domain-specific *patches* with no shared core meaning, held together only by a common strategy for extending the term to new cases. The range of contexts in which COMMUNICATION is put to use, surveyed above and throughout the paper, makes such a treatment the natural default.

Unexpectedly, a detailed investigation of the actual use of the term indicates against this default. Across the scales at which it is applied, COMMUNICATION is found to pick out a single causal organization. This shared causal organization constitutes a core meaning of the concept, and as such the notion is not a patchwork. Researchers appear to treat the widely different instances of communication as identifying the same type of process. If that is so, then a unified view of biological communication is required.

The causal structure COMMUNICATION picks out consists of four components: two parties, a sender and a receiver; a mark transferred between them; a structure that enables the transfer; and a function the transfer serves in some larger process. The first three recall components of the standard Shannonian model of information transmission (Shannon 1948). The fourth, the functional role, is recovered from the results and forms no part of that model. Hence, COMMUNICATION constitutes a mechanism schema (Machamer, Darden, and Craver 2000), an abstract description of a mechanism type that each field fills in with its own parts and activities. COMMUNICATION fixes a causally relevant grain of description, one at which sites for experimental intervention become visible while implementation detail stays available. The concept is parsimonious, and it guides the investigation of living and cognitive systems rather than merely labelling their interactions. Indeed, as I will argue below, some initial evidence points to the empirical benefits communicative framing brings—for instance, in research on morphogenesis, where viewing cell interactions in terms of communication enables novel experimental manipulation [e.g., Cervera et al. (2020); see section 5].

To substantiate this point, I introduce the results of 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 the concept and the various research practices it is involved in. This large scale picture of the many uses of communication is supplemented with a careful philosophical analysis of a quantitatively selected sample from papers. This allows me to develop a general account of BIOLOGICAL COMMUNICATION and uncover the epistemic functions the concept serves—its role in explanation, description, modelling, and experimentation.

The structure of the paper is the following: I begin by reviewing how digital study of scientific papers can inform our views on the conceptual practice of science (section 2). Next, in section 3, I briefly summarize existing literature on ANIMAL COMMUNICATION, which has received the most philosophical attention among the different references to biological communication. I review the informational and influence-based definitions of animal communication and motivate querying the corpus inclusively across that divide. In section 4, I develop an argument by negation. Initially, I assume the default pluralist position, drawing on the notion of patchwork concepts (Haueis 2024), to analyse the different uses of "communication". Contrary to the assumption, the analysis suggests a unified, if parsimonious, picture: the different uses of COMMUNICATION are connected by a shared property, the causal organization they pick out. 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, a mechanism schema, and an experimentally fruitful abstraction.

2 CONCEPTUAL PRACTICE OF SCIENCE THROUGH THE DIGITAL LENS

Tools of digital philosophy of science (Lean, Rivelli, and Pence 2023; Miłkowski and Nowakowski 2025; Pence and Ramsey 2018; Pence 2022b, 2025) allow for accounting for the full scope of the concept of COMMUNICATION without falling into the known pitfalls of standard philosophical methods, like case studies (see Pitt 2001). Computational methods provide a more comprehensive sample of the use of concepts (e.g., Malaterre and Léonard 2024; Miłkowski 2023; Overton 2013; Pence 2022a). Drawing on tools from 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 (of scientific texts—publications, peer reviews, blog posts, etc.) with qualitative close reading using standard philosophical methodologies (for a detailed discussion of the methodology see Rorot 2025).

Dominant approaches in philosophy of science in practice are suspicious of the image of science codified in its published outputs (see the discussion in Lean, Rivelli, and Pence 2023). However, the theoretical and conceptual domains of scientific practice function somewhat differently. This can be captured by viewing this conceptual dimension as a "system of practice" (Chang 2014). While scientific theories are standardly regarded as the outcomes of scientific research, they also play a *practical* role in the scientific day-to-day by shaping the research design (e.g., Brown 1996) or restricting the possibility space for investigation (e.g., Love 2013; Wilson 2006).

One way to view this role is through the cognitive lens on science (e.g., Giere 2010; Nersessian 2010; Thagard 1990, 2012), in particular in the form of recently developed "cognitive metascience" (Miłkowski 2022, 2023). The cognitive approach broadly views theorizing as a cognitive practice (Callebaut 2013; Chang 2017; Cheon and Machery 2016), and Miłkowski's version casts theories as perspectival "cognitive artifacts". Such artifacts are entities used to "maintain, display, or operate upon information in order to serve a representational function and that affect human cognitive performance" (Norman 1991, p. 11), that extend beyond (e.g., generalize or abstract) the available data. Through this lens, scientific practice can be analysed within the framework of cognitive science. It can be viewed as an instance of distributed cognition (Afeltowicz and Wachowski 2015; Giere and Moffatt 2003; Hutchins 2006; Osbeck and Nersessian 2014; Zhang and Norman 1994), where the"cognitive labor is distributed among lab members, or among multiple collaborating teams" (Miłkowski 2022, p. 221). The external form of the cognitive artifacts plays a central role in those contexts, as it is what enables this division of cognitive labour: the sharing of theories depends on (and is shaped by) their expression—primarily linguistic and often textual (for visual representations of theories see e.g., Elkins 2001; Meynell 2013; Perini 2005; Sartori 2025).

Considering the limited use of "theories" in life and mind research (e.g., Rosenberg and McShea 2012, ch. 2; see also Hardcastle 1996; Keller 2000; Ross 2025; for an alternative view see Sober 2018), I opt here to focus on the study of individual concepts and the larger frameworks they form. The standard psychological notion of concepts views them as "the basic units of knowledge [… which] consist of the information people associate with classes of objects and entities in the world" (Fini, Falcinelli, and Borghi 2025; this definition is motivated by the broader perspective of e.g., Barsalou 1999; Borghi et al. 2017; Borghi et al. 2019; Borghi et al. 2023; Murphy 2004). Frameworks, then, are "network[s] […] of interlinked concepts that together provide a comprehensive understanding of a phenomenon," "support one another, articulate their respective phenomena", and introduce "ontological, epistemological, and methodological assumptions" (Jabareen 2009).

Crucially, conceptual frameworks impact the identification, individuation and classification of phenomena by scientists (cf. Kant [1781] 1998). 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). Hence, considering again the cognitive metascience view, it makes sense to assume that some such effects might be present in scientific practice as well. This *constitutive* role of conceptual frameworks makes them a crucial part of practice, and highlights a methodological consequence the paper turns on: identifying the same conceptual structure across fields provides evidence that those fields individuate phenomena under a common template, conceiving them as instances of the same kind of process. In the current context, a distributional study of the use of "communication" can therefore uncover how scientific practice views processes described with this term.

The study reported here 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]

To explore this large scale corpus at a semantic level, I draw on a number of methods from distributional semantics: topic modelling, semantic search, and collocation analysis. These

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 available through the Allen Institute for Artificial Intelligence, its publisher and maintainer, see https://www.semanticscholar.org/product/api (note that as of July 2026, a free API key is required).

methods are motivated by a longstanding linguistic tradition which claims that the meaning of a word can be approximated by looking at its neighbourhood across texts (see e.g., Firth 1957; notably, this view has Wittgensteinian origins, Wittgenstein 1968). One of the benefits is that through computational methods, a numerical representation—vector embedding—of the semantics of a given document (at arbitrary granularities—from individual words to whole papers) can be created from such word co-occurrence data. This 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. The overall design of the study is presented in Figure 1, and the individual steps are fully described in Appendix A.

The selection of queries comes with a caveat: as the semantic search identifies sentences and paragraphs that are semantically similar to the queries, any selection of queries will skew the results. When interpreting the results of the study it is important to consider that some uses of the central concept—radically different from what queries specify—may have gone unnoticed. This is relatively smaller worry in the case of a term like "communication", where the space of possible technical definitions is restricted by the ordinary meaning of the word, but might matter more when specialist terms (like "information" or "representation") are explored using analogous digital methods.

Two features of the design limit this risk. First, the queries combine informational definitions with influence-based ones, cutting across theoretical distinctions (see section 3); a structure recovered across both is not an artifact of either formulation. Second, and more to the point, the main result of the paper—four components specifying the causal organization of COMMUNICATION—is not only read off the queries. The components are recovered from the grammatical collocates of "communication" in the retrieved passages (which relies on the fields' own prose, see Step 4. Word Sketch), and they recur in domains remote from the animal-signalling and language cases for which the queries were devised (see section 3), from quorum sensing to Notch signalling. The queries fix what is retrieved, but not the causal organization recovered from usage.

Through the constitutive role of conceptual frameworks, the results of the analysis will motivate two central arguments: first, they establish the shared structure of COMMUNICATION across fields (in section 4), and second, they uncover the epistemic work the concept does in scientific practice (in section 5).

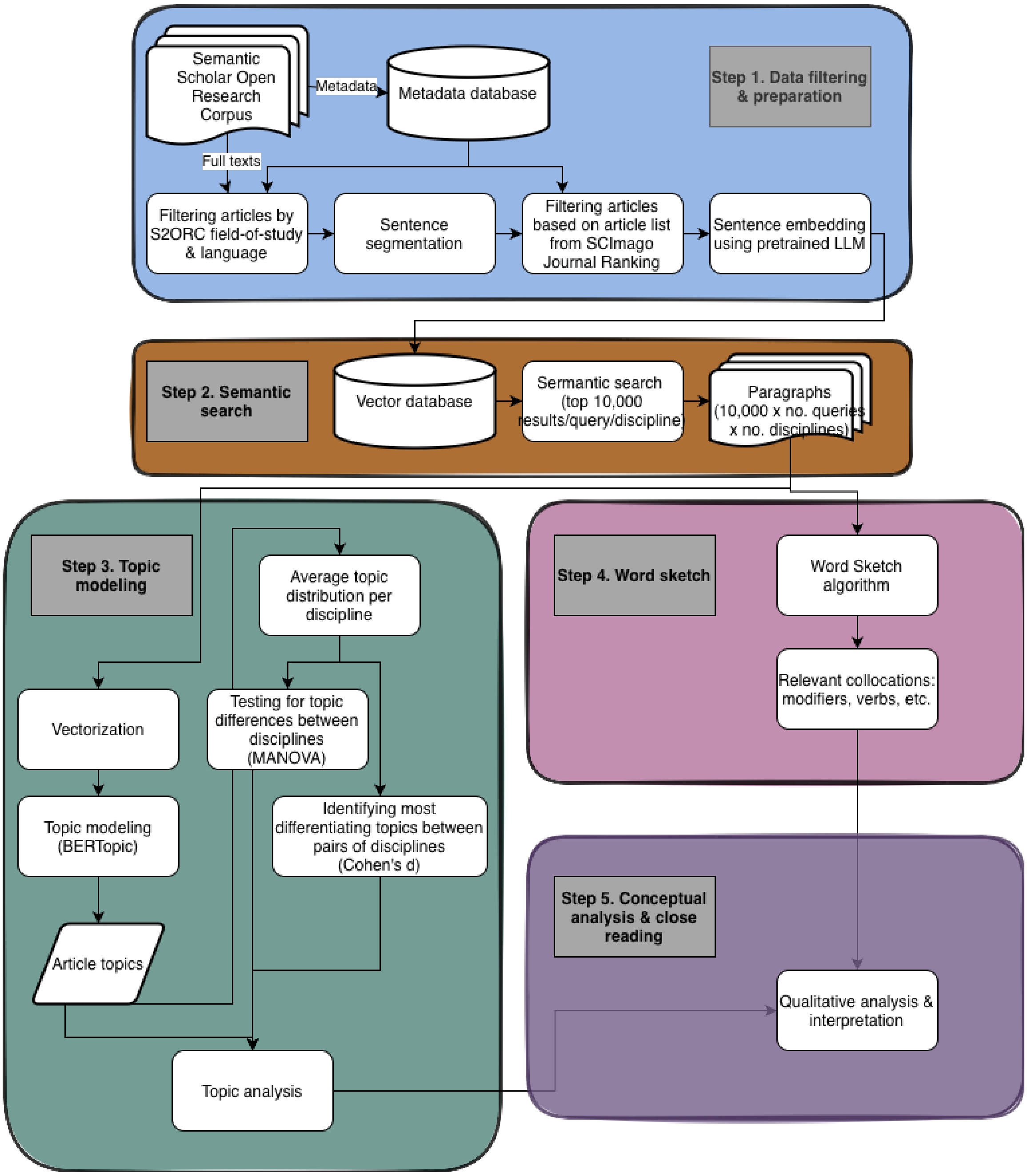
3 DEFINING "COMMUNICATION"

To track the uses of communication in the corpus, the selection of queries for semantic search (see details in Appendix A) relies on the existing philosophical treatments of biological communication. These focus almost exclusively on the use of the concept in reference to animal communication and to the evolution of human language.[2] In this context, Thom Scott-Phillips (2008) distinguished between two primary frameworks of viewing communication: informational and influence-based.

The former, also known as the "signalling model" (see Stegmann 2016), is primarily developed to bridge the gap between the animal and human communication systems (e.g., Bennett 1979;

2. This is also the case for existing treatments of SIGNALLING (e.g., Skyrms 2010).

Figure 1: Flowchart summarizing the research design. Individual steps are fully described in Appendix A. Most importantly, in step 2, vector embeddings of constructed queries (relying on definitions of BIOLOGICAL COMMUNICATION from the literature, see section 3 and Appendix B) are used to pick out highly similar sentences from the corpus, based on distance in the embedding space. These sentences are interpreted as referring to COMMUNICATION. In step 3, vector embeddings of these results are clustered to classify contexts in which COMMUNICATION is mentioned. In step 4, search results from step 2 are analysed at word level to identify collocates of theoretically relevant terms ("communication" itself). In step 5, the results of steps 2-4 are analysed qualitatively and used as evidence for conceptual analysis.



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), a 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, and honey bees' (*Apis mellifera*) waggle dance. 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 B).

Hence, to track the use of "communication" inclusively, without premature theoretical decisions (and to limit the bias of the results, as noted in section 2), I have used both 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 B (full query list is included in the online supplement). While these definitions, as already discussed, have been offered in reference to animal communication, they successfully 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 C.1). To capture this diversity, consider the different topics identified in the data (see Table 1) through the lens of the "patchwork concepts" approach (Wilson 2006; Haueis 2024)—assuming the "default" pluralist position on such widely used concepts. 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). The patchwork concepts do not have any "core" or "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, pp. 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 (the technique), with the similarity of function of neurons within the column being approximated via uniformity of the recorded responses [the property targeted by the technique; 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 (the technique), and the functional organization is inferred from the dominance of vertical over horizontal connections between neurons [the property targeted by the technique; see Mountcastle (1997)]. Each of these scale-dependent applications of the CORTICAL COLUMN forms an individual patch.

Indeed, most of the topics identified in the results can be interpreted as picking out a patch-like category of communicative phenomena: identifying the scale, technique, domain, and property. First, 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

Table 1: 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 (see subsection A.3), limited to 5 top terms. Full list of topics with complete representations (20 terms for each representation) is available in the online supplementary materials.

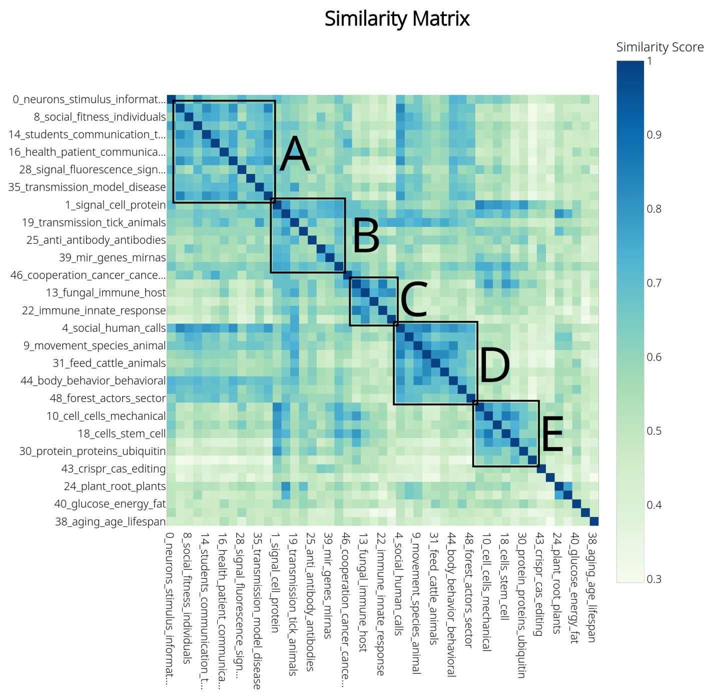
Topic	Count	Name	KeyBERT + MMR
0	7161	NERVOUS SYSTEM	stimuli, cortical, neurons, cortex, stimulus
3	6013	IMMUNE SYSTEM	immunotherapy, immune cells, immune response, innate immune, immunity
1	5211	QUORUM SENSING	signalling, gene expression, proteins, bacteria, pathways
2	4634	PLANT PATHOGENS	pathogens,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

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 the topics align along a similar structure—from the smallest to largest spatiotemporal scale.[3]

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: vocal-

3. 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.

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.



izations (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 for most frequent topics is presented in subsection C.3.

4.2 *Causal structure of communication*

However, despite the differences between the patches, the uses pick out a shared causal process. This organization has four components. Three are familiar from existing accounts of communication:

1. *Two (or more) parties*—at least one sender (communicator) and at least one receiver (audience), distinguishable from one another;
2. *A transferred mark*—something passed between them, whether a message encoded into a signal or a piece of evidence for an informative intention;
3. *An enabling structure*—a condition that makes the transfer possible, such as a shared code or appropriate intentions on both sides;

The fourth is recovered here and is not usually treated as a distinct structural component (see Appendix B). Communication is embedded within a broader system in which it serves a

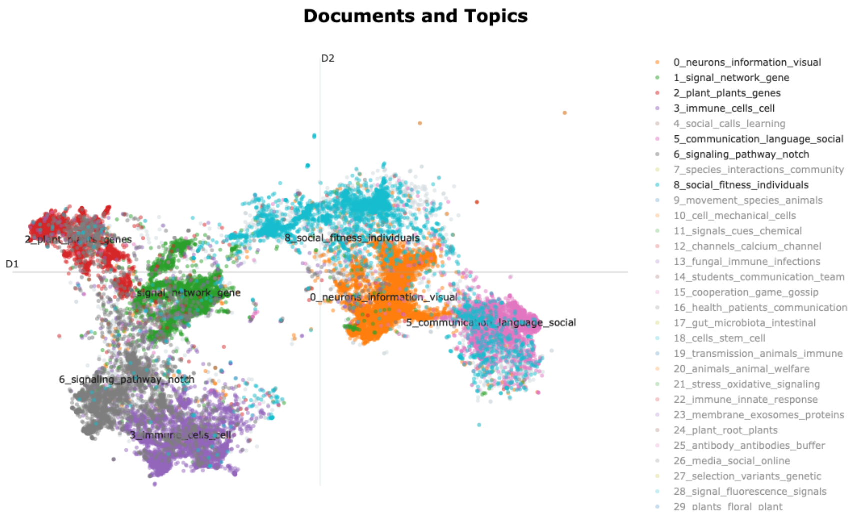


Figure 3: Two-dimensional representation of document vector embeddings, calculated by BERTopic using UMAP algorithm. The documents are clustered by topics based on the topic model. Each point represents a 2D-reduced vector embedding of a document (paragraph). Colors correspond to the topics. A subset 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.

particular function, so that representing a process as communication marks it as implicated in further processes:

4. *A function*: the role the communicative process performs within the system that contains it.

Each of the four components is marked by characteristic collocations of "communication" in the results of semantic search, drawn from different putative "patches" (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. Adjective predicates of "communication" provide an evaluative framework within which communication can be regarded as "important", "essential" or "crucial" but also "effective" or "successful." While not indicating any specific functions, these collocations indicate the broader involvement of communication in a variety of processes.

By the account adopted earlier [Haueis (2024); see subsection 4.1], a patchwork concept has no core meaning: each patch is individuated by four features together—the scale of investigation, the technique applied, the domain of application, and the property targeted—and the patches are unified only pragmatically through a general reasoning strategy, which informs researchers how to extend the concept to a novel domain, instantiating it through domain-specific techniques and properties. The uses of COMMUNICATION vary along three of those axes. They range across spatiotemporal scales (Figure 3), they are detected by techniques as different as electrophysiology and questionnaire study (highlighted in subsection C.3), and they belong to domains from molecular biology to sociolinguistics. Yet, they share a common property: across the varied applications they target the same causal organization with the four components just outlined.

In contrast to patchwork concepts, as in the case of a cortical column [analysed in Haueis (2021); see also subsection 4.1], the different techniques are not *constitutive* of a patch, serving to fix the meaning at a particular scale. Instead, in the case of COMMUNICATION, they are *detective*, locating the transferred mark within a domain without entering into what makes the process count as communication. In effect, the four components allow for multiple realizations. The transferred mark can be chemical, vocal, or gestural. The structure is a co-evolved molecular interaction between two proteins in one domain, and a configuration of intentions in another. The components are fixed at type level, and individuated by the function they perform, but allow for different subclasses of token realizers. But this does not lead to a polysemy of COMMUNICATION as the same types of components can be traced across all uses noted in the corpus: each patch predicates all four components and differs only in how each is filled.

What is recovered from the data, then, is a core meaning: a property predicated of every instance that fixes the identity of the phenomenon independently of the scale, technique, and domain. As such, COMMUNICATION displays genuine semantic unity, not just pragmatic connection characteristic of patchwork concepts. This result can now be read through the constitutive role of concepts set out in section 2. A conceptual structure shared across fields is evidence that those fields individuate their phenomena under a common template, not merely that a word is shared between them. The semantic unity of COMMUNICATION is therefore at the same time a claim about scientific practice: the disciplines that deploy the concept pick out one phenomenon, as it is constituted in their practice, applying a single concept at an abstract grain rather than a family of concepts held together by a shared procedure for coining them.

The causal structure identified here resembles the one Arnon Levy (2011) has noted in his discussion of BIOLOGICAL INFORMATION. As Levy's analysis highlights (2011, pp. 649-650), references to biological (semantic) information pick 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;

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.

However, Levy adopts the ready pluralist answer, advocating for a form of fictionalism. 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. This runs contrary to what corpus results analysed here show for "communication". Across the different uses, researchers do not hedge references to communication.[4] Considering the generally dominant realist views of scientists (see Beebe and Dellsén 2020), it motivates interpreting the concept in a realist fashion. Despite these differing ontological commitments, the parallel between causal structure of information Levy notes and the organization of communication identified in the current study further strengthens the claim of unity of COMMUNICATION through the common property of causal organization of the processes described with the term.

5 EPISTEMIC GOALS OF "COMMUNICATION"

Given the unity of BIOLOGICAL COMMUNICATION emerging from the data, 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 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 (e.g., Hanson 1958; Feyerabend 1962; Kuhn [1962] 2012), 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.

COMMUNICATION does not seem to fit the bill well. 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

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

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. Building on interventionist account of causation (Woodward 2004, 2010) he argues that "information" captures the environmental "interventions" on agents, just as "causation" captures agent's interventions on the world.

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". While this integrative function is not the explicitly formulated in the corpus, this is where a large potential epistemic payoff of this concept lays.

Through the cognitive perspective on science the epistemic goals of concepts can be considered beyond the taxonomies of aims of science. As noted in section 2, concepts have an important top-down role in which our previous knowledge or expectations shape what we perceive and experience. Individual concepts and larger conceptual frameworks can impact the identification, individuation, and classification of phenomena by scientists.

5.1 *"Communication" in discovery and experimentation*

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 implementation 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) positions 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 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 as used to store and modify information and to affect cognitive performance of researchers. 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 them 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.

In this regard concepts resemble 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), which 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.

5.2 *"Communication" as a mechanism schema*

In view of the unique granularity of COMMUNICATION, it can be viewed as a "mechanism schema". The notion of a schema has been a component of the neo-mechanist approach from its inception (Machamer, Darden, and Craver 2000; see also Craver and Darden 2001; Darden 2002). Schemas are characterized as abstract (incomplete) descriptions of types of mechanisms, that can be filled out with descriptions of known parts and activities. They can be offered at various levels of abstraction and generality. They can guide experimental work, specifying what entities or activities to look for (due to the constraints that types included in the schema introduce), effectively guiding mechanism discovery, and specifying sites for experimental intervention even if full description is not known.

In these respects, COMMUNICATION fits perfectly the definition of a mechanism schema. As highlighted throughout the paper, description in terms of COMMUNICATION specifies a particular, abstract organization of the process, highlighting types of components and activities involved. These types need to be additionally specified and pinpointed to particular physical instantiations in biological systems, leading to the apparent patchwork-like structure of the concept. But the constraints COMMUNICATION introduces effectively guides experimental practices. 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)—an intervention that relies on viewing morphogenesis as a communicative process.

As Machamer et al. note (2000, p. 16), schemas frequently figure as activities in higher-level mechanisms. This captures the "functional" character of communicative interactions visible in the analysis (section 4). Representing the process in the schema of communication allows for focusing on the higher-level mechanism, while sufficiently substantiating the activities that bring it about. Such a "functional" reference to a mechanism schema can only be successful if there is sufficient agreement on the existence of a mechanism type described by the schema.

Importantly, the account of COMMUNICATION offered here is highly parsimonious. This level of abstraction provides a highly generalizable notion that can be effectively applied across a wide range of scales of biological organization (although in general higher abstraction needs not lead to a wider scope, e.g., Yoshida and Love 2025). The parsimony of this proposal grounds its integrative promise, which allows for appreciating the similarities of structure across scales. This integration enhances the opportunities for understanding the processes captured as COMMUNICATION, regardless of the complexity of the biological instantiation (Potochnik 2017; Rorot and Miłkowski 2024; Wilkenfeld 2019). Further, it motivates knowledge and tool transfer across the fields, for instance, as is increasingly the case in the basal cognition research (e.g., Levin et al. 2021; Lyon et al. 2021), where neuroscientific knowledge is applied to questions from developmental or plant biology (and vice versa).

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 I suggest that despite the variety of spatiotemporal scales at which communicative processes are posited, whenever referring to BIOLOGICAL COMMUNICATION, researchers aim to identify instances of one kind of process, characterized by the shared causal organization set out in subsection 4.2: (1) two parties, (2) a transferred mark, (3) an enabling structure, and (4) a function.

Such a notion of communication is a mechanism schema—a (partially) abstracted description of a type of mechanism, that can be instantiated by a variety of processes (and hence, allows for multiple further specifications). In this context, the epistemic benefits of BIOLOGICAL COMMUNICATION appear to be twofold: first, 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 capturing 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 accounts strengthens the view of BIOLOGICAL COMMUNICATION as a robust (in the sense of Wimsatt 2007) phenomenon across the scales of biological organization. At the same time, COMMUNICATION offers important theoretical benefits beyond the existing "semiotic" conceptualizations of biological phenomena, in particular in terms of INFORMATION, as it underscores the processual, interactive character of biological and cognitive processes (e.g., Brancazio 2023; Dingemanse et al. 2023; Nicholson and Dupré 2018).

7 ACKNOWLEDGMENTS

I want to thank my supervisors, Joanna Rączaszek-Leonardi and Marcin Miłkowski, for guiding my work on the doctoral project this paper grew out of. I have presented earlier versions of this work at the ISHPSSB conferences in Toronto in July 2023 and in Porto in July 2025, at the "Philosophy of Science: Past, Present, and Future" meeting at the University of Minnesota in October 2024, and at the Philosophy of Science Around the World online conference in November 2025 and want to thank all the audiences for helpful feedback. Most directly, this manuscript has greatly benefitted from the comments of Wiktor Lachowski, Szymon Miłkoś, Enrico Petracca, Daniel Piecka, Aline Potiron, Marcin Rabiza, and Lorenzo Sartori, as well as from discussions with Olesya Bondarenko, David Harrison, Urte Laukaityte, Nedah Nemati, and Charles Pence. Parts of this research have been carried out while I was a Writing-Up Fellow at the Konrad Lorenz Institute from October 2024 to March 2025, and the atmosphere of the KLI has been instrumental in bringing this project to fruition.

8 FUNDING

This work was supported by the National Science Centre, Poland, under research project no. UMO-2022/45/N/HS1/02434 awarded to WR.

9 LLM DISCLAIMER

Claude Opus 4.8 model was used in editing and proofreading the text.

A APPENDIX 1. METHODS

A.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.[5] 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 2) 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).[6] Table 2 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 spatiotemporal scale of biological processes that the areas of biology are concerned with. 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).

5. 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.

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

Table 2: 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 1.4 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

A.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 2) 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 vector comparison (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 fallback 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 C.1, and full results are available in the online supplement). The choice of 10,000 most similar documents 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.

A.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). KeyBERT calculates the vector distance (semantic similarity) between candidate keywords and topic embeddings, selects best candidates (least distance), and filters them with the Maximal Marginal Relevance algorithm, to minimize redundancy between 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). The final number of topics is selected 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, a model with 50 topics is selected (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.

A.4 *Step 4. Word Sketch*

Fine-grained semantic structure of relevant terms (in particular, "communication") is explored through the Word Sketch algorithm (Kilgarriff et al. 2014). Word Sketch uses lemmatized representation of the semantic search results with syntactic tags (constructed using syntactic tagging from spaCy Python package) which allows for an analysis of collocations of target words, retaining the syntactic structure (e.g., separating modifiers or verbs for which the target is a subject or an object). The Word Sketch algorithm produced lists of top 20 collocates from different syntactic categories, using a logDice statistic to establish typicality of co-occurrence, see (Rychlý 2008).

A.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. 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).

B APPENDIX 2. 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: $\mu = 0.4138$, $\sigma = 0.1061$

C APPENDIX 3. SAMPLE RESULTS

C.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)

c.2 *10 most frequent for the topic model of the uses of "communication"*

Table 3: 10 most frequent topics with the number of documents classified as belonging to that topic (Count), excluding the "outliers" topic which was ranked 8th overall. 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 (see Appendix A). The table includes all representations used, limited to top 5 terms for each. Base representation is the default c-TF-IDF algorithm (class-based term frequency—inverse document frequency, see subsection A.3). "Nouns and modifiers" and "Verbs and modifiers" use Spacy part-of-speech tagger with the `en_core_web_sm` model to extract terms from the c-TF-IDF which match specified syntactic patterns—in this case, nouns or verbs, and their modifiers (if any). "KeyBERT + MMR" calculates the vector distance (semantic similarity) between candidate keywords and topic embeddings, selects best candidates (least distance), and filters them with the Maximal Marginal Relevance algorithm, to minimize redundancy between keywords. Full list of topics is available in the online supplement at: https://osf.io/z342v/overview?view_only=aca8dfbaa299468fbd3ae35472d74ae2.

Topic	Count	Name	Base representation	Nouns and modifiers	KeyBERT + MMR	Verbs and modifiers
0	7161	NERVOUS SYSTEM	neurons, stimulus, information, visual, brain	neurons, stimulus, visual, information, brain	stimuli, cortical, neurons, cortex, stimulus	learning, reward, processing, used, using
3	6013	IMMUNE SYSTEM	immune, cells, cell, tumor, immune cells	immune, cells, tumor, cell, cancer	immunotherapy, immune cells, immune response, innate immune, immunity	innate, associated, signaling, related, study
1	5211	QUORUM SENSING	signal, cell, protein, gene, network	signal, cell, protein, network, gene	signaling, gene expression, proteins, bacteria, pathways	signaling, binding, using, used, based
2	4634	PLANT PATHOGENS	plant, plants, genes, signaling, defense	plant, plants, signaling, genes, resistance	pathogens, signaling pathways, arabidopsis, biosynthesis, triggered immunity	plant, signaling, pathogen, induced, triggered
6	4241	NOTCH DEVELOPMENT	signaling, pathway, notch, cell, wnt	notch, signaling, pathway, cell, wnt	notch signaling, signaling pathway, notch1, wnt signaling, signaling	notch, signaling, pathways, signalling, activated

Continued on next page

Table 3: (continued)

Topic	Count	Name	Base representation	Nouns and modifiers	KeyBERT + MMR	Verbs and modifiers
5	3785	SOCIAL FACTORS IN LANGUAGE	communication, language, social, children, conversation	communication, language, social, children, conversation	social communication, social interaction, conversations, communicative, autism	talk, use, turn, used, understanding
8	3690	SOCIAL BEHAVIORS	social, fitness, individuals, individual, group	fitness, social, individuals, individual, group	evolutionary, traits, mating, aggression, behavioral	group, given, based, study, used
4	3148	SOCIAL LEARNING	social, human, calls, learning, communication	social, human, calls, dogs, communication	primates, behavioral, chimpanzees, behaviors, species	learning, use, study, used, contexts
7	3065	ECOLOGICAL COMMUNITIES	species, interactions, community, communities, microbial	species, interactions, community, microbial, communities	microbial communities, ecological, ecosystem, species, phylogenetic	taxa, used, based, benefit, understanding
20	2182	HUMAN INTERSPECIES INTERACTIONS	animal, animals, welfare, human, dogs	animal, animals, dogs, pets, pet	attachment pets, companion animals, animal welfare, emotional attachment, animal assisted	care, use, stress, assisted, related

C.3 *Patches of "communication" corresponding to the 10 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 labelling (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
6. Topic ID: 5
   Topic Name: SOCIAL FACTORS IN LANGUAGE
   Scale: Multicellular organisms
   Domain: Verbal linguistic communication: capacities and use
   Technique: Neuropsychological methods (e.g., study of aphasia patients), questionnaire-based methods, psychological experiments
   Property: Vocabulary, language processing efficiency
7. Topic ID: 8
   Topic Name: SOCIAL BEHAVIOURS
   Scale: Multicellular organisms, multiple interacting organisms
   Domain: Group behaviours, social behaviours of individuals
   Technique: Ethological studies, mathematical models
   Property: Behavioural patterns, fitness
8. Topic ID: 4
   Topic Name: SOCIAL LEARNING

Scale: Multiple interacting organisms
Domain: Social transmission of behaviour
Technique: Ethological studies
Property: Behavioral repertoires of individuals and populations

9. Topic ID: 7
   Topic Name: ECOLOGICAL COMMUNITIES
   Scale: Multiple interacting groups
   Domain: Ecological species interactions
   Technique: Microscopy, mathematical models
   Property: Diversity, fitness
10. Topic ID: 20
   Topic Name: HUMAN INTERSPECIES INTERACTIONS
   Scale: Multiple interacting organisms
   Domain: Human attitudes and behaviours towards other groups of animals
   Technique: Questionnaire-based methods, case studies
   Property: Human opinions, human welfare

REFERENCES

Afeltowicz, Łukasz, and Witold Wachowski. 2015. "How Far We Can Go Without Looking Under the Skin: The Bounds of Cognitive Science." *Studies in Logic, Grammar and Rhetoric* 40, no. 1 (March): 91–109. https://doi.org/10.1515/slgr-2015-0005.

Anta, Javier, and Julia Sánchez-Dorado. 2025. "The Pursuitworthiness of Informational Biology." *Biology & Philosophy* 40, no. 5 (October): 21. https://doi.org/10.1007/s10539-025-09996-y.

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

Barsalou, Lawrence W. 1999. "Perceptual Symbol Systems." *The Behavioral and Brain Sciences* 22, no. 4 (August): 577–609, discussion 610–660.

Bartlett, Stuart, Andrew W. Eckford, Matthew Egbert, Manasvi Lingam, Artemy Kolchinsky, Adam Frank, and Gourab Ghoshal. 2025. "Physics of Life: Exploring Information as a Distinctive Feature of Living Systems." *PRX Life* 3, no. 3 (September): 037003. https://doi.org/10.1103/rsx4-8x5f.

Beebe, James R., and Finnur Dellsén. 2020. "Scientific Realism in the Wild: An Empirical Study of Seven Sciences and History and Philosophy of Science." *Philosophy of Science* 87, no. 2 (April): 336–364. https://doi.org/10.1086/707552.

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

Borghi, Anna M., Laura Barca, Ferdinand Binkofski, Cristiano Castelfranchi, Giovanni Pezzulo, and Luca Tummolini. 2019. "Words as Social Tools: Language, Sociality and Inner Grounding in Abstract Concepts." *Physics of Life Reviews* 29 (July): 120–153. https://doi.org/10.1016/j.plrev.2018.12.001.

Borghi, Anna M., Ferdinand Binkofski, Cristiano Castelfranchi, Felice Cimatti, Claudia Scorolli, and Luca Tummolini. 2017. "The Challenge of Abstract Concepts." *Psychological Bulletin* 143, no. 3 (March): 263–292. https://doi.org/10.1037/bul0000089.

Borghi, Anna M., Albertyna Osińska, Andreas Roepstorff, and Joanna Rączaszek-Leonardi. 2023. "Concepts in Interaction: Social Engagement and Inner Experiences." *Philosophical Transactions of the Royal Society B: Biological Sciences* 378, no. 1870 (February): 20210351. https://doi.org/10.1098/rstb.2021.0351.

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.

Brown, Harold I. 1996. "The Methodological Roles of Theory in Science." In *The Scientific Nature of Geomorphology: Proceedings of the 27th Binghamton Symposium in Geomorphology, Held 27 - 29 September 1996,* edited by Bruce L. Rhoads and Colin E. Thorn, 3–20. Chichester: Wiley.

Callebaut, Werner. 2013. "Naturalizing Theorizing: Beyond a Theory of Biological Theories." *Biological Theory* 7, no. 4 (June): 413–429. https://doi.org/10.1007/s13752-013-0122-2.

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.

Chang, Hasok. 2012. *Is Water H_2O? Evidence, Realism and Pluralism.* v. 293. Dordrecht [The Netherlands]: Springer Verlag.

———. 2014. "Epistemic Activities and Systems of Practice." In *Science after the Practice Turn in the Philosophy, History, and Social Studies of Science,* edited by Léna Soler, Sjoerd Zwart, Vincent Israel-Jost, and Michael E. Lynch, 67–79. Routledge Studies in the Philosophy of Science 14. New York: Routledge.

———. 2017. "Operational Coherence as the Source of Truth." *Proceedings of the Aristotelian Society* 117, no. 2 (July): 103–122. https://doi.org/10.1093/arisoc/aox004.

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.

Craver, Carl F., and Lindley Darden. 2001. "Discovering Mechanisms in Neurobiology: The Case of Spatial Memory." In *Theory and Method In The Neurosciences,* edited by Peter K. Machamer, Rick Grush, and Peter McLaughlin, 112–137. Pittsburgh: University of Pittsburgh Press, March.

Darden, Lindley. 2002. "Strategies for Discovering Mechanisms: Schema Instantiation, Modular Subassembly, Forward/Backward Chaining." *Philosophy of Science* 69, no. S3 (September): S354–S365. https://doi.org/10.1086/341858.

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.

Elkins, James. 2001. *The Domain of Images.* Ithaca, NY: Cornell Univ. Press.

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.

Fini, Chiara, Ilenia Falcinelli, and Anna M. Borghi. 2025. "Technological Concepts: Object and Tool of Knowledge Outsourcing at Different Ages." *Topoi* (May). https://doi.org/10.1007/s11245-025-10206-3.

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.

Gentner, Dedre, and Michael Jeziorski. 1993. "The Shift from Metaphor to Analogy in Western Science." In *Metaphor and Thought,* 2nd ed, edited by Andrew Ortony, 447–480. Cambridge (Mass.): Cambridge university press.

Giere, Ronald N. 2010. *Explaining Science: A Cognitive Approach.* Chicago, IL: University of Chicago Press.

Giere, Ronald N., and Barton Moffatt. 2003. "Distributed Cognition: Where the Cognitive and the Social Merge." *Social Studies of Science* 33, no. 2 (April): 301–310. https://doi.org/10.1177/03063127030332017.

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

Godfrey-Smith, Peter. 2007. "Information in Biology." In *The Cambridge Companion to the Philosophy of Biology,* edited by David L. Hull and Michael Ruse, 103–119. Cambridge Companions to Philosophers. New York: Cambridge University Press.

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.

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.

Hardcastle, Valerie Gray. 1996. *How to Build a Theory in Cognitive Science.* Albany: State University of New York Press.

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.

Hesse, Mary. 1966. *Models and Analogies in Science.* Notre Dame: University of Notre Dame Press.

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.

Hutchins, Edwin. 2006. *Cognition in the Wild.* Cambridge, Mass.: MIT Press.

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.

Keller, Evelyn Fox. 2000. "Models of and Models for: Theory and Practice in Contemporary Biology." *Philosophy of Science* 67:S72–S86.

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.

Levin, Michael, Fred Keijzer, Pamela Lyon, and Detlev Arendt. 2021. "Uncovering Cognitive Similarities and Differences, Conservation and Innovation." *Philosophical Transactions of the Royal Society B: Biological Sciences* 376, no. 1821 (March): 20200458. https://doi.org/10.1098/rstb.2020.0458.

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.

Love, Alan C. 2013. "Theory Is as Theory Does: Scientific Practice and Theory Structure in Biology." *Biological Theory* 7, no. 4 (June): 325–337. https://doi.org/10.1007/s13752-012-0046-2.

Ludwig, David, and Stéphanie Ruphy. 2026. "Scientific Pluralism." Edited by Edward N. Zalta and Uri Nodelman. *The Stanford encyclopedia of philosophy.*

Lyon, Pamela, Fred Keijzer, Detlev Arendt, and Michael Levin. 2021. "Reframing Cognition: Getting down to Biological Basics." *Philosophical Transactions of the Royal Society B: Biological Sciences* 376, no. 1820 (March): 20190750. https://doi.org/10.1098/rstb.2019.0750.

Machamer, Peter, Lindley Darden, and Carl F. Craver. 2000. "Thinking about Mechanisms." *Philosophy of Science* 67 (1): 1–25. https://doi.org/10.1086/392759.

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.

Meynell, Letitia. 2013. "Parsing Pictures: On Analyzing the Content of Images in Science." *The Knowledge Engineering Review* 28, no. 3 (September): 327–345. https://doi.org/10.1017/S0269888913000271.

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.

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.

Murphy, Gregory L. 2004. *The Big Book of Concepts.* Cambridge, Mass.: MIT Press.

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.

Osbeck, Lisa M., and Nancy J. Nersessian. 2014. "Situating Distributed Cognition." *Philosophical Psychology* 27, no. 1 (January): 82–97. https://doi.org/10.1080/09515089.2013.829384.

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.

———. 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.

Perini, Laura. 2005. "The Truth in Pictures." *Philosophy of Science* 72, no. 1 (January): 262–285. https://doi.org/10.1086/426852.

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.

Rosenberg, Alexander, and Daniel W. McShea. 2012. *Philosophy of Biology: A Contemporary Introduction.* New York London: Routledge.

Ross, Lauren N. 2025. *Explanation in Biology.* Cambridge University Press, January. https://doi.org/10.1017/9781009300940.

Rychlý, Pavel. 2008. "A Lexicographer-Friendly Association Score." In *Proceedings of Recent Advances in Slavonic Natural Language Processing,* edited by Petr Sojka and Aleš Horák, 6–9. Masaryk University, Brno.

Sartori, Lorenzo. 2025. "Why We Love Pictures (for the Wrong Reasons): A Lesson from the Picture of a Black Hole." *Philosophy of Science* 92, no. 5 (December): 1338–1349. https://doi.org/10.1017/psa.2025.10148.

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.

Shannon, Claude E. 1948. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27, no. 3 (July): 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Shannon, Claude E., and Warren Weaver. 1964. *The Mathematical Theory of Communication.* Urbana: The University of Illinois Press.

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.

Skyrms, Brian. 2010. *Signals: Evolution, Learning, & Information.* Oxford ; New York: Oxford University Press.

Sober, Elliott. 2018. *Philosophy of Biology.* Boulder: Routledge.

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.

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.

———. 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.

Wilkenfeld, Daniel A. 2019. "Understanding as Compression." *Philosophical Studies* 176, no. 10 (October): 2807–2831. https://doi.org/10.1007/s11098-018-1152-1.

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.

Wittgenstein, Ludwig. 1968. *Philosophical investigations.* Translated by G. E. M. Anscombe. Oxford: Basil Blackwell.

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.

Yoshida, Yoshinari, and Alan C. Love. 2025. "Mechanisms and Principles: Two Approaches to Scientific Generalization." *European Journal for Philosophy of Science* 15, no. 2 (June): 26. https://doi.org/10.1007/s13194-025-00650-8.

Zhang, Jiaje, and Donald A. Norman. 1994. "Representations in Distributed Cognitive Tasks." *Cognitive Science* 18, no. 1 (January): 87–122. https://doi.org/10.1207/s15516709cog1801_3.