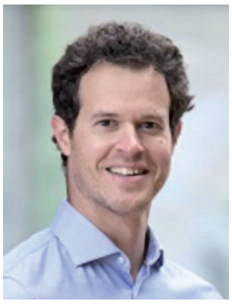
Consensus and Scientific Classification

Beckett Sterner*, Atriya Sen** and Joeri Witteveen***

*Arizona State University, School of Life Sciences,
427 E Tyler Mall, Tempe, AZ 85281, USA, <Bsterne1@asu.edu>
**University of New Orleans, Department of Computer Science,
2000 Lakeshore Drive, New Orleans, LA 70148, USA, <asen@uno.edu>
***University of Copenhagen, Department of Science Education,
Universitetsparken 5, 2100 København Ø, Denmark, <jw@ind.ku.dk>

Beckett Sterner is an Assistant Professor in History and Philosophy of Science in the School of Life Sciences at Arizona State University. He received his PhD from the University of Chicago's Committee on the Conceptual and Historical Studies of Science. His research investigates the driving question of what knowledge is needed for people to work together while differing in fundamental ways. Some of his recent projects include explaining ambiguity in scientific language, assumptions underwriting objectivity in statistics, and regulative ideals for data commons.

Atriya Sen is an Assistant Professor in Computer Science. He earned his PhD at Rensselaer Polytechnic Institute. His research interests include artificial Intelligence, particularly AI & explanation, computational logic, cognitive AI systems, philosophy of AI, automated scientific discovery, computational philosophy, biodiversity informatics, AI for sustainability and conservation biology, and bioinformatics generally.



Joeri Witteveen is an assistant professor in history and philosophy of science at the University of Copenhagen in Denmark. He received an MSc degree from the London School of Economics and obtained a PhD from the Department of History and Philosophy of Science at the University of Cambridge. His research interests include the history and philosophy of biological taxonomy and nomenclature and the epistemic and ethical tensions between science and policy making at the interface of taxonomy and species conservation.

Sterner, Beckett, Atriya Sen and Joeri Witteveen. 2022. "Consensus and Scientific Classification." *KnowledgeOrganization* 49(4): 236-256. 137 references. DOI:10.5771/0943-7444-2022-4-236.

Abstract: Consensus about a classification is defined as agreement on a set of classes (concepts or categories) and their relations (such as generic relations and whole-part relations) for us in forming beliefs. While most research on scientific consensus has focused on consensus about a belief as a mark of truth, we highlight the importance of consensus in justifying shared classificatory language. What sort of consensus, if any, is the best basis for communicating and reasoning with scientific classifications? We describe an often-overlooked coordinative role for consensus that leverage agreement on how to disagree such that actors involved can still achieve one or more shared aims even when they do not agree on shared beliefs or categories. Looking forward, we suggest that investigating structures and methods for coordinative consensus provides an important new direction for research on the epistemic foundations of knowledge organization.

Received: 29 March 2022; Accepted 15 August 2022

Keywords: scientific consensus, classification structure, taxonomy, authorities

1.0 Introduction

This article reviews the growing body of literature on consensus and scientific classification in order to distinguish different roles and grounds for consensus in the production and use of scientific classifications. We follow Henry Bliss's definition, (cited in Hjørland 2017), in understanding a classification to be "a series or system of classes arranged in some order according to some principles or conception, purpose or interest, or some combination of such." Hull (1998) also provides a useful analysis of the component parts of a classification: "The fundamental elements of any classification are its theoretical commitments, basic units and the criteria for ordering these basic units into a classification." We define consensus about a classification is defined as agreement on a set of classes (concepts or categories) and their relations (such as generic relations and whole-part relations) for us in forming beliefs, and we understand scientific classifications here to be those produced and used as part of scientific research.

Consensus as an outcome of a group deliberative process is often seen as an important marker of knowledge and prerequisite for collective action (Miller 2013; Stegenga 2016; Brooy, Pratt and Kelaher 2020). Most discussion has focused on cases where the object of consensus is a belief or set of beliefs, such a collection of facts or principles (Miller 2019). In their attempts to explicate what makes a consensus belief a plausible indicator of knowledge, social epistemologists have mentioned such criteria as the absence of coercion, reliance on several lines of evidence, sufficiently large group size, inclusion of diverse and independent agents (Miller 2013; Tucker 2003; Goldman 2001). Meanwhile, global organizations and intergovernmental panels that operate at the intersection of science and policy try to meet these criteria to ensure that the expert consensus statements they produce are taken to reflect the latest knowledge on climate change (Beck et al. 2014), public health (Stegenga 2016; Bouwel and Oudheusden 2017), and biodiversity (Díaz-Reviriego, Turnhout and Beck 2019).

Agreeing on the truth of a shared proposition, though, presupposes common knowledge of the terms and meanings needed to express it. In other words, it is often (and quietly) assumed that consensus about classification can be passed over in the philosophical discussion of the epistemic import of consensus and dissensus in science. In many pluralistic contexts, whether interdisciplinary research projects in academia or policy decisions in a democratic society, shared terms and meanings cannot be taken for granted (Sullivan 2017; Weiskopf 2020; Sterner, Witteveen and Franz 2020). Moreover, the growing social reach of computer technologies and emphasis on re-use of scientific data are raising new challenges for communication and reasoning across local communities and organizational boundaries (Gerson 2008; Leonelli 2008; Ribes and Bowker 2009). In practice, scientific classifications generally provide limited rather than exhaustive coverage of terminology for a specific domain, but they also typically use or define terms from other domains, for example to characterize a deeper hierarchy of phenomena or concepts or to describe how experimental data from one domain is acquired and processed in another. These connections between classifications used in different domains are institutionalizing an emerging network of information exchange that has the potential to bring the concerns of previously very "distant" fields into close dialogue and interaction.

With these motivations in mind, the core question we aim to address is therefore, "What sort of consensus, if any, is the best basis for communicating and reasoning with scientific classifications in a domain?" In attempting to give an answer, one issue we need to address is whether a single best classification exists for the purposes of representing knowledge or phenomena in a particular domain. If this isn't available, i.e. when pluralism about classifications for a domain is warranted, then a second issue is how one can nonetheless ensure reliable communication and reasoning within and among domains.

In order to synthesize and clarify literature addressing these issues, we identify different types and aims of consensus that have typically operated in the background assumptions of particular classification projects. Section 1 presents two general types of consensus about classifications: consensus classification, which we define as agreement on a set of categories to use in forming beliefs, and coordinative consensus, which represents a second-order way of agreeing on how to disagree so that people can still share information. The types of consensus we describe are consistent with different views about the proper organization and content of scientific classifications, e.g. whether they should be genealogical or conform to a strict logical hierarchy (Gnoli 2018). Similarly, the issues we highlight are conceptually prior to common practical methods for classification such as ontology design, statistical clustering, or logical division (Blomqvist, Gangemi and Presutti 2009; Hjørland and Gnoli 2020; Fricke 2016).

Sections 3 and 4 discuss consensus classifications in more detail, including their philosophical interpretation and how consensus may be justified. Sections 5 and 6 then review approaches to coordinative consensus, including the role of Artificial Intelligence (AI) in articulating our collective ability to agree about how to disagree. We illustrate these general issues with examples of classifications drawn from a wide range of fields, including astrophysics, neuroscience, psychology, traditional ecological knowledge, and systematic biology. Additional relevant examples are addressed in Dick (2019) and Lee (2019).

Throughout the article, we highlight several important gaps in current research. Debates over the desirability and guiding principles for computer ontologies among scientists have been largely detached from the philosophical literature on consensus, which suggests the potential for fruitful sharing of problems and insights. In addition, questions about the existence, or even desirability, of consensus classifications raise parallel issues to the philosophical literature addressing the merits of consensus as a basis for belief. Furthermore, the practical reality of living with dissensus about classifications expands the philosophical landscape ways by forcing us to consider the implications of how we formalize communication and reasoning. In particular, we want to highlight Nicholas Rescher's (2000) book on pluralism "against the demand for consensus" as an accessible and far-reaching guide to this broader landscape that merits renewed interest and engagement by scientists and philosophers. Arguably the most important gap we identify, though, is the dearth of studies investigating how the types of consensus we describe below are realized for classifications through specific institutional or governance processes (e.g. comparable to peer-review or National Institutes of Health consensus committees). Much more work is needed to investigate the relationship between the types and methods of consensus researchers pursue for scientific classifications.

2.0 Epistemic roles for consensus in scientific classification

Most philosophical discussion of consensus to date has focused on beliefs, actions and values as the primary objects of agreement. In the context of scientific inquiry, the role of consensus about beliefs (about states of affairs) in making and assenting to epistemic claims has been discussed at length by social epistemologists. In this section, we give a brief overview of recent philosophical discussions about the role of consensus and dissensus in social epistemology.

2.1 Relationship to knowledge

Consensus in the sense of bare agreement on some matter is not a reliable marker of knowledge. Consensus must be produced in the right way for it to be epistemically significant. This has led several theorists to distinguish between different kinds of consensus, such as *accidental* vs. *essential* consensus (Fuller 2002), or *de facto* consensus ("faktisch erzielte Konsensus") vs. *rational* consensus ("vernünftige Konsensus") (Habermas 1971). Only the second forms of consensus in these pairs represent epistemically justified kinds of consensus, rooted in certain shared standards of evidence, relevance, and/or rationality. Accidental or de facto consensus, on the other hand, manifest a fortuitous convergence of views or opinions, with no further epistemic import.

When should we have epistemic trust in a consensus? Following Miller (2019), we distinguish *cognitive*, *social*, and *hybrid* accounts of knowledge-based consensus. Cognitive accounts identify consensus beliefs as knowledge if they are consistently successful in prediction and explanation, and show resistance to falsification (e.g. Oreskes 2007). A problem for cognitive accounts is that consensus is not principally required for a view to score well on cognitive criteria. Hence, such criteria are not about consensus *per se*. Fringe views could qualify as more epistemically significant than the consensus on a cognitive account.

Social accounts attempt to address this by identifying knowledge-based consensus based on social factors alone (Tucker 2003; Miller 2013). These accounts present requirements on the social and procedural organization of epistemic communities, such that if their inquiries yield consensus, they are plausibly knowledge-based. For example, Tucker (2003) argues that knowledge is the best explanation for consensus if the latter is uncoerced, uniquely heterogeneous, and based on a large population. Critics of social accounts nevertheless point out that social criteria fail to exclude actual and imagined cases in which a consensus is clearly accidental rather than knowledge-based.

Finally, hybrid accounts aim to combine the best of both cognitive and social criteria. Miller (2013), for example, argues that consensus yields knowledge when it meets the cognitive criteria of calibration of evidential standards and consilience of evidence as well as the social criterion of a diversity of perspectives.

Rescher (2000) and Beatty & Moore (2010) criticize the view that consensus plus some extra cognitive or social ingredients is a reliable indicator of knowledge. Consensus has no intrinsic significance for deciding what counts as knowledge, and should therefore not be made into the aim of scientific inquiry. Rescher argues that rationality does not necessitate consensus. Variation in people's standpoints, cognitive values, experiences and methodologies is compatible with rationality of inquiry. From this perspective, perfect epistemic consensus only obtains under ideal conditions that cannot be realized in practice. In short, "consensuality is neither a requisite for, nor a consequence of, rationality in the conduct of inquiry" (Rescher 2000, 12). Problems of underdetermination in inductive contexts and overdetermination in reductive inquiry make rational disagreement based on different cognitive and contextual values almost inescapable (Rescher 2000, 37ff). Though a Kuhnian picture of normal science may be taken to suggest that research is predominantly consensus-based puzzle solving (as opposed to the Popperian embrace of dissent), even this Kuhnian consensus is perennially unstable.

Beatty and Moore go further, by arguing that the existence of a (near) consensus can be a good reason to doubt the veracity of a position. They contend that the existence of substantial minority view is an indicator that the dominant position is being tested against a worthy alternative. In their view, the aim of inquiry should not be consensus, but "deliberative acceptance", which is understood as acceptance *by the group* of a majority position based on the quality of their deliberation, in spite of the persistence of minority views held by individual participants (Beatty and Moore 2010, 209). "Disagreement at the individual level is not so worrisome if it is combined with acceptance at the group level and with deliberative acceptance in particular." (Beatty and Moore 2010, 210).

2.2 Relationship to communication

The idea that situations featuring co-existing agreement and disagreement could have special merit is also linked to an observation Rescher makes about the communication of conceptions that we will explore and elaborate on in more depth. Rescher notes that communication does not require common conceptions or agreement about the meaning of terms, but "like dancing, it calls for *co-ordination* rather than *agreement*" (Rescher 2000, 138). We agree, and will explore the requisites for successful coordination in the context of communicating information using classifications as well as about the classifications themselves.

Understood as infrastructure for communication and reasoning, classifications provide a shared set of rules about how people should apply a set of regulated terms and what inferences one may draw from applications of those terms in particular cases (Bowker and Star 1999; Bowker et al. 2009). Multiple "design" choices are involved in the production and maintenance of a classification: which terms should be included in its scope, how to express (represent) the corresponding rules of usage, and what reasons are legitimate to offer in settling answers for both of these questions? Theoretical and technological advances in knowledge representation and reasoning are also opening up new ways to make the content of a classification explicit and translate effectively among incompatible classification systems. As we'll see below, alternative approaches to representing classifications carry different pragmatic affordances for communication and reasoning and hence potential for collective action.

To implement or revise a classification is thus to make a social intervention, and the value of a classification, e.g. for communication and discovery, depends on the social processes supporting it. Henry Bliss's (1929) work stands out in this regard as an early argument for the value of a general consensus classification of knowledge for the functional organization of society (Hjørland 2016). Highlighting the close connection he saw between constructing shared classifications and progress, Bliss quoted the American Engineering Standards Committee's 1928 yearbook motto: "Standardization is dynamic, not static. It means, not to stand still, but to move forward together." Bliss's core aim, though, was to replace the Dewey Decimal system with what he saw as a more theoretically principled way of organizing library materials, and in this regard he largely failed to win converts to his proposed classification (Hjørland 2016). His experience helps motivate the importance of understanding the what, when, how, and why of consensus as an adequate basis for a shared classification.

An important point missing from Bliss's early treatment is the recognition that consensus can operate on multiple levels. In what follows, we will discuss two forms of consensus about classification: consensus on a single best classification, and consensus about how to coordinate among incompatible classifications. Both forms of consensus provide regulative ideals for standardizing the application of labels, but they differ in their "level" of application. Consensus about the best classification applies at the first-order level of standardizing how the terms are applied across the entirety of the relevant community. A consensus classification therefore constitutes substantive agreement on the meanings of terms.

Coordinative consensus, in contrast, allows disagreement in the application of classificatory terms but standardizes how this disagreement is characterized and managed. A coordinative consensus therefore constitutes procedural agreement on how to declare and manage conflicting associations of terms and meanings. Coordinative consensus thus enables persistent use of multiple classifications within a community. Sufficient agreement must exist on the nature of conflicts or changes among these alternative classifications in order to avoid fragmentation. With this shared picture of people's first-order disagreements, it is then possible to determine what second-order information about the alternative classifications is required to maintain accurate communication and reasoning among the community.

Understanding the implications of these two forms of consensus is only becoming more important with the growing reach of computational reasoning across organizational boundaries and communities (Gerson 2008). The continued development of semantic web, linked data, and other associated technologies are expanding the impacts of classifications on scientific practice (Leonelli 2013a; 2016). In the past several decades, the ideal of a consensus classification has been especially influential in standardizing terminologies for describing biological data using computer ontologies (Sterner, Witteveen and Franz 2020). At the same time, computer technology is also lowering barriers to creating novel or customized classifications, facilitating local benefits for communities but also proliferating the number of historical and competing versions that people have to manage. Understanding of how different forms and implementations of consensus are appropriate to different contexts is likely to drive both practical and theoretical advances in our knowledge.

An early study of ontology development in geoscience, for example, found that initial support for the categories in the ontology developed through a largely informal and iterative process of domain expert recruitment, training, and input (Ribes and Bowker 2009). Ontology developers led "concept workshops" that sought to elicit socially distributed knowledge from geoscientists in a way that was representative of the larger geoscience community. The disagreements and debates that emerged in those workshops among the domain scientists signaled to the group that a larger process of engagement was important to gain support and adoption for the ontology. The initial group decided to make the ontology in development public on the web and invited comments, with "the most enthusiastic (and prestigious) geoscience colleagues... invited to the concept-space workshops themselves, to act as participants" (Ribes and Bowker 2009, 213). No formal determination of consensus was made, or indeed whose opinion would be counted in such an effort, but participants consciously understood themselves as standing in for the larger community as representatives and that the process would be stronger if it was not executed by a top-down, exclusive group. This is more elaborate than the "bunch of guys sitting around a table" method one participant described for the DSM in psychiatry, but it highlights the social and political dimensions involved in legitimating scientific classifications (Kendler and Parnas 2012, 141).

Additionally, the network of geoscientists and ontology developers involved developed multiple strategies for handling persistent disagreements. We quote from Ribes and Bowkers list at length to illustrate the richness of strategies involved:

1. Decrease granularity: deal with the issue at a higher conceptual level, where the domain has established a stronger consensus...
2. Working agreement: encourage the domain experts to form a temporary working consensus in order to continue the ontology development process at hand. [For example, one participant] suggests a solution relative to the "science questions" geologists would like to address, rather than to a more abstract goal of serving the entire community of geoscientists.
3. Rain-check: leave the problem aside for a later time; experts may be able to resolve the issue with a re-examination of evidence, review of the literature, further consultation with experts or, in the long term, production of new evidence... To some extent this makes it difficult for the modelers to create drafts or prototype ontologies, however, the process of development is iterative and often takes months.
4. Represent the uncertainty: this is a technical solution within semantic modeling which affords encoding and representing disagreement, uncertainties, ambiguities or ambivalences. Within [the project] this was a promised capability of ontologies; however, since it was still an experimental option, we never saw it employed in practice. We expect that expressing domain knowledge as a form of uncertainty is itself a practical skill – and one perpetually controversial amongst scientists.

An important distinction to keep in mind is the difference between consensus on a classification and its status as an authoritative reference standard. In biodiversity science, for example, different organizations task taxonomic experts to develop comprehensive lists of species for use in downstream goals, such as assessing extinction risk, studying responses to climate change, or studying evolutionary history. These lists are amalgamations of results from many different research papers, which often propose and defend a handful of new or revised species at a time. Some lists aggressively adopt the latest published species under the assumption that the most recent science is best, albeit with certain exceptions, while others are more conservative and wait for a preponderance of evidence or community support to emerge. It is moreover common for there to be multiple authoritative lists for a biological group in use and maintained at a time which disagree on the species they recognize. There have been four major checklists for birds for decades (Le Page et al. 2014), for example, and mammals are currently covered by the Mammal Diversity Database, Bats of the World, and Mammal Species of the World checklists. The International Union for the Conservation of Nature (IUCN) also maintains its own taxonomic list of animal, plant, and fungi species for the purposes of assessing extinction risk in its Red List of Threatened Species. The differences between these lists demonstrate the absence of total taxonomic consensus, but they are all recognized as valid reference standards for research use. To the extent that consensus emerges among these lists, it often arises by users "voting with their feet," i.e. indicating their support by using one list over the others (Sterner, Gilbert and Franz 2020). Having illustrated some of the organizational and social complexities involved in scientific classifications, we now turn to consider the grounds on which consensus can be treated as a desirable marker of quality in the first place.

3.0 Grounds for consensus classification

There are multiple strategies for grounding the epistemic merit of consensus as a basis for selecting the best classification. As Rescher points out, "It is never *just* consensus we want but the *right sort* of consensus—that is, a consensus produced in the right way—one that obtains for good and cogent reasons" (Rescher 2000, 16). That people agree on some point is not generally sufficient to justify its use, although we note below that adopting a shared convention can sometimes be justified on practical grounds such as efficiency. We describe two important distinctions scientists and philosophers have used to characterize "good and cogent reasons" for agreeing on a classification: operational versus theoretical, and natural versus artificial (Hjørland 2020, Section 4). For a further discussion of warrants as a basis for the selection of classificatory terms in library science, see (Martínez-Ávila and Budd 2017). We also discuss how the entrenchment of a classification in common use can go either way in this respect as a good practical reason for agreement or an obstacle to improvement.

3.1 Operational versus theoretical grounds

The first distinction is between operational versus theoretical foundations for defining consensus terms, which is also related to historical discussion of real versus nominal kinds (Sullivan 2017; Ludwig 2018; Sterner, Witteveen and Franz 2020; Mattu and Sullivan 2020). As we use it here, the distinction is between justifying the categories in a consensus classification based on agreement about the correct interpretation of information about the world versus the reliability of certain procedures or sources of information about the world.

For example, should the categories in computer ontologies classifying anatomical parts of organisms in different species be based on evolutionary homologies or structural similarity? The classification of the parts and properties of organisms into anatomical categories serves to standardize their description for use in biomedicine, evolutionary biology, and ecology. Using evolutionary homology as a theoretical basis for defining anatomical features is potentially problematic in this regard, in that our knowledge of homology relationships depends on our certainty about what happened deep in evolutionary history, frequently tens or hundreds of millions of years ago. As a result, an anatomy ontology whose terms always designated homologies would have to deal with regular instability from advances in our best knowledge about evolutionary relationships among traits and controversies about whether some traits are indeed homologies.

Scientists developing the Hymenoptera Anatomy Ontology (HAO), for instance, argue that "The dynamic nature of homology hypotheses conflicts with the HAO's goal of unambiguous circumscription of anatomical concepts, and, as such, overt reference to homology hypotheses are avoided in constructing HAO definitions" (Seltmann et al. 2012, 79). Instead, they argue that structural equivalence, i.e. the identical relative placement of a part in the bodies of different species, provides a more reliable operational basis for classification. Whether a structural basis to anatomy provides an adequate foundation for discovering, communicating, and reasoning about evolutionary homologies is contested, however (Deans, Yoder and Balhoff 2012; Franz and Goldstein 2013).

Theory-neutrality has also been appealing for psychologists attempting to converge on an intersubjectively reliable basis for classifying psychological disorders. The atheoretical aims of the Diagnostic and Statistical Manual of Mental Disorders III (DSM-III) committee stand out in particular (Cooper 2005; 2017b). "Mental health professionals operate within numerous different theoretical frameworks, and there were fears that a D.S.M. based on any one particular theory would alienate many practitioners. The D.S.M.-III committee sought to avoid this problem by basing the D.S.M. on no theory at all." (Cooper 2005, 78). As a result, the introductory text proclaimed, "The approach taken in DSM-III is atheoretical with regard to etiology ... except for those disorders for which this is well established ... The major justification for the generally atheoretical approach ... is that the inclusion of etiological theories would be an obstacle to use of the manual by clinicians of varying theoretical orientations" (as quoted in Tsou 2016, 413). This ideal was widely criticized both in theory and in execution for particular disease definitions in DSM-III, and was abandoned in later versions, at least as an explicit and universal principle (Tabb 2015; Tsou 2016; Cooper 2017a). As Rachel Cooper suggests, a more modest effort at theory-neutrality would look for descriptive terms that are operationally reliable and neutral with respect to many or all of the alternative theories under consideration in the field at some time (Cooper 2005, Ch 3).

3.2 Natural versus artificial grounds

The second major distinction we will address is between natural and artificial classifications (Müller-Wille 2013; Wilkins and Ebach 2013). We suggest this can be usefully subdivided further to separate: 1) whether the categories in a classification are meant to describe how nature is or the way we conceptualize it (Hacking 1991; Mattu and Sullivan 2020), and 2) whether the classification is meant to be general purpose or special purpose (e.g. Dupré 1995; Ludwig 2018). Perhaps the most common understanding of a natural classification is that it describes how nature is and (therefore) provides a useful basis for all purposes. Similarly, an artificial classification is commonly understood as a description of how we conceive of nature that is suited for some special interest we bring to the phenomena as humans but may not accurately represent nature. Whether describing nature and providing a general-purpose classification always travel together is contested, however, and we discuss this in more detail in the following section.

In this vein, it is also important to note that requiring global agreement on whether classificatory theories should represent reality or concepts presents a false choice for science. The position that computer ontologies should describe reality, not our concepts, has been historically influential in the development of the Open Biological and Biomedical Ontologies Foundry (OBO Foundry), which is one of the largest repositories and standard-setting groups for ontologies in science (Smith et al. 2007). This view was defended forcefully by philosopher Barry Smith and bioinformatician Werner Ceusters in a series of papers in the early 2000s, which pointed out how existing biomedical terminologies flip-flopped between defining terms as describing concepts or natural phenomena (Smith and Ceusters 2003; Ceusters, Smith and Goldberg 2005; Smith 2006; Smith and Ceusters 2006). Such inconsistencies lead to invalid semantic reasoning when a higher-level category, for example, is defined as a concept while a subcategory within it refers to a natural process.

(As an aside, we note that in computer science and knowledge representation, the word "ontology", following Husserl, has come to be dissociated from the question of existence and used in a purely formal sense (Smith 1987). However, inversely, the study of ontologies may be considered a formal approach to the study of metaphysics or science. An example of the former is Abstract Object Theory, which may be seen as an extension of mathematical platonism (E. Zalta 1983). An example of the latter, formal physics, is discussed in Example I of Section 6.3 below.)

Ceusters and Smith then argued that ontologies adopted as community standards should only include categories referring to real natural phenomena. This would imply excluding categories for hypothetical phenomena not yet demonstrated, for instance, as well as fictions, idealizations, or groupings shown not to exist. They argued this approach was best positioned to lead to a natural classification system that maximized interoperability and reuse of terms across domains. An extended debate from 2010–2013, however, surfaced doubts about what work the idea of "real" phenomena was doing for ontology design and disagreements about how to handle uncertainty and common modeling practices such as idealization (see Sterner, Witteveen and Franz 2020 for historical overview and further references). Recent research on referent tracking in clinical databases and ecology provides a response to some of these concerns by more explicitly handling changes in belief about the existence and properties of particular diseases or patient conditions (Hogan and Ceusters 2016; Diller et al. 2020). Separately, many of Smith and Ceusters' proposed design principles for ontologies reflect an Aristotelian approach to metaphysics that is hard to reconcile with contemporary particle physics or evolutionary biology (Arp, Smith and Spear 2015), and no global fundamental ontology has emerged as a consensus to unify existing scientific domain systems (Biagetti 2020). In Section 5, we return to discuss positive approaches in current information science approaches that are able to handle conflicting semantics among co-existing ontologies without leading to invalid conclusions.

It is also worth noting that institutional entrenchment of a (merely) conventional classification can also ground consensus on the basis of shared efficiencies. Perhaps the best-known example of this outcome is the Dewey Decimal Classification (DDC) for books. (Outside knowledge organization, one could also point to QWERTY keyboards and continued use in some countries of the English system of units as examples.) At the time of the DDC's development in the 1870s, most libraries in the U.S. used a fixed shelving location for books based on their date of purchase for the collection. Placing books by topic in a standard alphanumeric sequence therefore represented a major gain in finding related works within a library and navigating the holdings of unfamiliar institutions. The sequence of topics described by the DDC, however, does not reflect a systematic theory of the organization of scholarly knowledge, which has been a consistent source of criticism since its origin. As the critic Henry Bliss experienced with his proposed alternative, once a system becomes widely adopted it takes tremendous organizational power and effort to change, and this barrier can simply overwhelm the virtues of any competitors (Garfield 1974).

4.0 Pluralism and classification

Substantial disagreement remains about whether and why we should expect a single best consensus classification to exist for any domain of science (Hjørland 2017, Section 4). In the natural sciences, if anywhere, one might expect we should be able to set aside our human differences of opinions and values to converge on a single way of representing the world. On the other hand, one might think nature exceeds our ready grasp in this fashion, or may be less definitive in its answers to our questions than hoped. This latter view suggests that, "if the multitude assent and applaud, men ought immediately to examine themselves as to what blunder or fault they may have committed" (as quoted in Rescher 2000, 29; original saying from Phocion, adapted by Francis Bacon, Novum Organon I.77). Having set out an array of cogent reasons for consensus in the prior section, we now turn to consider in principle and in practice arguments about what conclusions we can draw from them. For further discussion of whether such arguments for pluralism lead us to a radical position of permitting any way of classifying things as legitimate, see Slater (2017) and Conix (2019).

4.1 In-principle arguments

How might aiming for a natural classification grounded in scientific theory fall short of achieving consensus? A basic commitment at issue here is whether converging on a comprehensive set of truths a phenomenon entails converging on a single representation encapsulating those truths, e.g. a scientific theory. A major contribution of recent philosophy of science in this respect has been to articulate and clarify arguments for pluralism in metaphysics. This literature is often overlooked in discussions about designing computer ontologies, and so represents a valuable and unexplored area of interaction. This gap is especially evident in work on ontological realism approach, which does not acknowledge or engage with established literature on pluralist metaphysics, for example in relation to biological species (c.f. Arp, Smith and Spear 2015).

One influential argument for pluralist metaphysics is premised on the claim that nature is too complex to be captured within any single model or theory that we can derive. (An example of a different starting premise would be that some natural phenomena are fundamentally indeterminate, e.g. lacking precise boundaries (Keil, Keuck and Hauswald 2017). Sandra Mitchell, for example, has shown how different explanatory factors social insect behavior are captured by multiple models that can't be unified because their individual epistemic contributions depend on idealizing the phenomenon in incompatible ways (Mitchell 2003; 2009). Margaret Morrison has made a similar point about the family of models physicists use to understand the behavior of atomic nuclei in different experimental settings (Morrison 2011). For related critiques of essentialism. the idea that natural phenomena can be classified by unchanging and universal necessary and sufficient conditions, see Dupré (1995) and Binney (2015). Also, recent work on perspectival realism (Crețu and Massimi 2020) will be of special interest for those wanting to maintain a realist stance while accommodating metaphysical pluralism.

What about a natural classification grounded on reliable operational methods? In response, a second line of argument for classificatory pluralism starts from the diversity of scientific questions or ways of investigating nature and shows that no single way of classifying nature is universally optimal. This argument is often presented in a way that also invokes the idea that nature's complexity overflows the expressive capacity of our representations. This is classically illustrated by the debate over how to define species, where prioritizing reproductive isolation, evolutionary history, and ecology can lead to irreconcilable species classifications (Ereshefsky 1992). It has also been argued to apply even to the periodic table as a classification of the elements, where the point of contention is not the list of elements per se, but whether there is a single objective basis for grouping them according to similarities in their various properties (Scerri 2006, 288).

There is also a more direct path to classificatory pluralism from the diversity of scientific goals, which does not invoke any special claims about complexity. As philosopher David Danks argues, "ontological pluralism can arise simply from our inability to directly observe the world's 'true' ontological structure, combined with our multiple mathematical criteria for when we 'understand the world.'" (Danks 2015, 3603). For instance, we might grant that science pursues a single overarching goal, to "understand the world". Nonetheless, we are unable to directly observe how well we are doing on that goal, as we lack the ability to simply look and see how our beliefs compare to the ontology of the world. Instead, we have multiple indirect ways of evaluating this aim, e.g. in terms of how well we do at prediction and explanation. In principle, these indirect evaluations could cohere and lead us to a single true ontology, but in fact they sometimes diverge for deep and persistent (e.g. mathematical) reasons.

4.2 In-practice arguments

The arguments for classificatory pluralism we have surveyed so far fall short of clinching the generalized conclusion that communities will have better outcomes if they maintain multiple concurrent classification systems. Pluralism is not a panacea in this respect, and as a general approach it poses important trade-offs or dangers in its own right. In-principle arguments for the legitimacy of a pluralistic stance may be overridden by major costs that accrue in putting this into practice, including fragmentation of knowledge under different classifications, lower economies of scale, and new barriers to translation or synthesis (Bowker and Star 1999; Bowker 2000; Leonelli 2016; Sullivan 2017). Pragmatics about the cognitive, social, and infrastructural capacities of a community enter into the picture here as an additional layer affecting outcomes (Cooper 2005; Sterner and Franz 2017; Sterner, Gilbert and Franz 2020). The broad-minded virtue of being a pluralist can thus become a vice without a calibrated sense of where to draw the limits: "the greatest danger of pluralism is that it provides no means or even motivation for reducing conceptual luxuriance" (Hull 1987, 178).

Reflecting the importance of situated pragmatics, historical and social science studies of classifications support a contextual approach to assessing the merits of adopting and maintaining a consensus classification. In some circumstances, pluralism can generate obstacles to adoption or distract from high-value use cases. This has arguably been the case for anatomical terminology in many areas of systematic biology, where many scientists have maintained idiosyncratic language and definitions (Deans, Yoder and Balhoff 2012). Too much disconnected variation has substantially hampered efforts to digitize knowledge about species' traits and make the most of new computational reasoning tools and measurement technologies. It is also possible for projects to invest too much up front in supporting pluralistic approaches when this does not lead to short-term growth in users or funding. For instance, both The Arabidopsis Information Resource (TAIR) and the NIH cancer informatics project, caBIG, initially defined a broad set of ambitions to handle many types and formats of biological data (Leonelli 2013b). After receiving about $350 million in funding, caBIG was eventually shut down after having invested too much money in generating an inclusive suite of software products that frequently demanded high-levels of technical sophistication and local support for cancer research centers to adopt. TAIR eventually restricted its focus to a narrower set of data and a single species that constituted the dominant service value it provided to the community, but which kept it from pursuing emerging new research areas in plant biology.

Many people also place high value on following a shared, authoritative standard for classifying phenomena. It is undeniably simpler in certain respects if everyone agrees to follow the same standard, although only if that standard does not overload some users with high customization or workaround costs in their local settings. In the history of astronomy, for example, the Draper (a.k.a. Harvard) Classification for stars held considerable attraction as a consensus standard for the international community in part because the number of stars classified under it far exceeded any other option (DeVorkin 1981; Dick 2013). Its formal adoption by the International Union for Cooperation in Solar Research in 1910, however, did not signal consensus on the underlying temporal dynamics ("evolution") of stars. The classification held operational rather than theoretical value in this respect, reflecting the scientific community's increasing interest in research projects that only worked with international support.

The Draper Classification as adopted also did not represent the best system in all respects, as shown by subsequent modifications that merged in desirable features from competitors (DeVorkin 1981, 49). In sum (DeVorkin 1981, 48):

> Almost all astronomers nevertheless agreed that one classification was needed, whatever might be the course of stellar evolution, and that the classification to choose was Pickering's. That his Draper classification was the one adopted, when it was becoming quite clear that it did not provide an unambiguous classification of giant and dwarf stars, was the result of several other factors: the vast bulk of the stars classified in the system; the complexity of the system-intermediate between the simplest systems (Secchi and Vogel) and the most involved systems (Maury, Lockyer); the apparent neutrality of the system regarding one or another scheme of stellar evolution; the physical significance of the system as a temperature sequence; the circumstances affording Pickering a major role in the development of the consensus itself.

Scientists have also argued for the pragmatic value of a single, authoritative classification in the domains of psychology and biological systematics (Garnett and Christidis 2017; Poland and Tekin 2017). Nonetheless, the pursuit of a consensus classification has the downside of eliminating the linguistic resources necessary to express valuable dissent. The virtues of standardization are context-sensitive in this regard, and when implemented under inapt conditions can lead to the exclusion, conflation, or erasure of local variation needed for successful collective action.

Increasing recognition of the value of Traditional Ecological Knowledge (TEK) to inform conservation and sustainability efforts provides a useful example (Tengö et al. 2017; Ludwig 2018; Weiskopf 2020; Yeboah et al. 2020). Even as Western researchers have been involved in creating the conditions that made climate change and global biodiversity possible, they have also often downgraded the validity of non-academic knowledge acquired by indigenous peoples, who have proven to be some of the most effective stewards of nature worldwide (Garnett et al. 2018; Ludwig and El-Hani 2020). The pursuit of a "unified" system of scientific classification for biodiversity and ecology can prove to be epistemically harmful if the very premises on which such a system would be acceptable make it impossible to include other valuable ways of knowing. Similarly, consensus as an ideal may also have negative moral and political consequences when standardizing the use of certain categories entails the abandonment or rejection of independently valuable alternatives. Biological systematics has a long history of colonialist erasure (and exploitation) of Indigenous language, place names, and knowledge in this respect, although some recent efforts are exploring how to engage in indigenous-oriented research (Walter et al. 2020; McAllister et al. 2020). It is critical to recognize that the goal of making local circumstances globally intelligible, e.g. for conservation and biodiversity monitoring, can in fact enable the exploitation and decline that making such portable knowledge is meant to prevent (Rye and Kurniawan 2017; Fraser 2019; Chase, Chase and Chase 2020; Rubis and Theriault 2020).

Another source of difficulties for consensus classification comes from how the implementation of standardized categories can strengthen and entrench obstacles to possible lines of inquiry. Sociologist Geoffrey Bowker made this point forcefully for biodiversity as a sort of Matthew effect (the rich tend to get richer and the poor poorer) for what we name, catalogue, and measure in the living world. "If certain kinds of entities and certain kinds of context are being excluded from entering into the databases we are creating, and those entities and contexts share the feature that they are singular in space and time, then we are producing a set of models of the world which, despite its frequent historicity, is constraining us generally to converge on descriptions of the world in terms of repeatable entities: not because the world is so, but because this is the nature of our data structures" (Bowker 2000, 655). More recently, Franz and Sterner have argued that aggregating biodiversity data using an imputed, global taxonomy has multiple downsides for community, including lower trust and engagement in data portals, increased barriers to recognition for dissenting taxonomic viewpoints, and reduced data quality (Franz and Sterner 2018; Sterner, Gilbert and Franz 2020).

A different example of consensus-seeking classification leading to unwanted and entrenched side effects can be found in the DSM and its effects on psychiatric research. Some authors have suggested that the standardized criteria for psychological disorders are not in fact primarily used by practicing clinicians, who typically consult the precise diagnostic criteria only for unfamiliar cases (Phillips et al. 2012). Instead, students use the DSM as part of their formal training, and researchers use it most essentially to address the operational need for uniform study populations. Given the non-mechanistic and fuzzy character of many diagnostic definitions, it is then "an ironic side effect of the diagnostic criteria that they may impede research by confining research efforts to criteria-determined questions" (Phillips et al. 2012, 2). The limitations of the DSM for research use have in fact led to a complementary, though also contested, system called the Research Domain Criteria (RDoC), introduced by the National Institute of Mental Health (Poland and Tekin 2017).

Finally, ontologies and databases in cognitive neuroscience, such as BrainMap and Neurosynth, have arguably helped to stratify terminologies without aligning concepts (Sullivan 2017). As a result, meta-analyses based on literature linked to labels in ontologies can yield spurious results, when those labels abstract away from conceptual disagreements and differences in experimental paradigms and protocols used in the generation of the relevant datasets. Ontologies in cognitive neuroscience thus not only "run afoul of the realist aim" to which they aspire, but they put it further out of reach by burying references to conceptual differences that would need to be resolved for realism to be obtained (Sullivan 2017, 134).

5.0 Coordinative consensus

As has been recognized in the political science and philosophy literature (Rescher 2006; Niemeyer and Dryzek 2007; Golding 2013; Muldoon 2016), consensus is possible on different orders of abstraction. As we described in Section 3, first-order consensus about a classification consists of everyone agreeing to use the same categories. Second-order consensus, by contrast, consists of everyone agreeing about how to disagree about the choice of categories while maintaining certain collective abilities, such as being able to exchange information despite conflicting classifications. Rescher again states the point nicely: "What is crucial for your ability to communicate with me–to convey to me information about your beliefs, plans, or values–is not that we have a commonality of beliefs or ideas, and so stand in a consensus of some sort, but something quite different. It pivots on the recipient's capacity to *interpret*–to make good inferential sense of the meanings that the declarer is able to send. In the final analysis the matter is not one of an *agreement* between parties but of a *co-ordination* between them on the basis of a recipient's unilateral capacity effectively (successfully) to interpret the substantive" (Rescher 2000, 148).

Coordinated action nonetheless does require some amount of agreement on the matter of how to get things done (Star and Griesemer 1989, Gerson 2008). Getting things done together despite substantial differences in beliefs and concepts therefore represents a non-trivial constraint on the scope of pluralism that can be sustained in practical situations of limited resources and motivations. Several authors in philosophy of science have begun to explore the implications of this point (Sullivan 2017; Weiskopf 2020; Ludwig and El-Hani 2020; Sterner, Witteveen and Franz 2020), and it is an established topic of research in the social sciences (see, for example, the journal *Computer-Supported Cooperative Work*). Coordination without universally shared classifications is also closely related to the idea of interoperability in data science (Zeng 2019, Section 5.3).

One important distinguishing feature of coordination as a second-order endeavor is that it can remain shallow and partial rather than entailing convergence and unification in the long run (Weiskopf 2020). In other words, a group of people may maintain the ability to talk and act together despite persistent disagreements on language as a "steady state" under certain circumstances. Persistent coordinative pluralism in this form may be desirable in at least two contexts: first, where joint communication and reasoning are desirable but stronger forms of integration or unification are not currently possible; and second, where developing forms of integration or unification are not desirable in themselves. Philosopher David Weiskopf argues this latter case may be common for interactions between Indigenous Peoples and their ways of knowing with Western scientific communities and ways of knowing. "The model of knowledge integration is one on which both bodies of knowledge come together into a single overarching whole that nevertheless, in practice, often ends up being dominated by the most politically and economically powerful party" (Weiskopf 2020, 9). In contrast, the coordinative approach he advocates "replaces this model with one in which both bodies briefly come into contact and separate, each having been changed but neither having been subsumed. This approach promises a more accurate representation of traditional knowledge and of the ways that asymmetrically positioned communities negotiate these exchanges" (Weiskopf 2020, 9).

Reaping the benefits of coordination generally takes substantial investment in dialogue and articulating new ideas or practices; simply putting existing terminologies, data, or other resources in one place is insufficient (Sullivan 2017; Weiskopf 2020; Sterner, Witteveen and Franz 2020).

Indeed, computer scientists and industry researchers have devoted massive efforts toward automating data integration in large-scale repositories of information called “data lakes,” but the resulting computational methods are typically designed to serve the aims of one organization rather than establish consensus of either form in the community at large. In contrast, the *Cognitive Atlas* project in cognitive science allows experts “to identify and document when they agree or disagree about (a) how a term designating a construct is defined and (b) the asserted relationships between different constructs” (Sullivan 2017, 135). As Sullivan observes, these features provide means for identifying conceptual and methodological differences, but not for resolving them by integration. In a related spirit, the Digital Archaeological Repository (tDAR) allows users to specify and align external ontologies with the repository’s internal ontology resources in order to carry out reproducible and automated data integration (McManamon et al. 2017; Altschul et al. 2017; 2018).

Sterner, Witteveen and Franz (2020) identify general conditions for when the conditions of coordination are sufficiently regular and recurring to merit formalization in their own right. In contrast to approaches based on the use of consensus definitions for classificatory terms, they introduce a “Coordinative Consensus Principle (CCP): the design of a formal classificatory system for expressing a body of data should be grounded in a consensus standard for coordinating the application of names of classified entities, even if the meanings and extensions of those names haven’t been settled” (Sterner, Witteveen and Franz 2020, 8). They illustrate how this principle operates in biological taxonomy, where a many-to-many relationship generally holds between taxonomic names and their meanings. Despite this linguistic complexity, systematists are able to agree about correct or incorrect applications of a name according to a usage specified in a particular taxonomic treatment. Such agreement on the correct application of a name in the face of disagreement about its proper meaning is possible because systematists generally follow nomenclatural rules that fix the theoretical referent of a name using a designated type (e.g. a type specimen), and they additionally provide operational circumscriptions for the referent, i.e. hypotheses that describe which organisms should be classified under the name. This approach makes it possible for experts to share and receive information from incompatible sources because they can interpret how the source’s meaning differs from their own preferred view.

A variety of computational approaches are emerging for characterizing the second-order information required to maintain communication and reasoning among a community despite persistent variation in terms and meanings. In the context of biomedical ontologies, for example, researchers have investigated how different semantic relationships can help document changes across versions (Mungall 2019), and even automatically propagate changes in definition using logic reasoning (Groß, Pruski and Rahm 2016). In biodiversity data science, a more promising approach has been to sidestep reliance on explicit theoretical definitions for terms, which are largely absent in machine-readable form, and instead characterize semantic relationships using logical vocabulary from Region Connection Calculus (Franz and Peet 2009; Franz et al. 2015; 2016; Sen et al. 2020). Similar issues arise in Geographic Information System (GIS) research, where machine learning methods are often essential to operationalize theoretical concepts but then introduce new complications for documenting, versioning, and aligning differences in classifications across users and time (Gupta and Gahegan 2020).

In circumstances where precise alignments are needed, coordinative consensus presupposes rather than replaces the work of transdisciplinary inquiry that generates new integration among alternative disciplinary or social worldviews. Extensive research in history and philosophy of science has shown that scientific discovery often proceeds through analogical or metaphorical reasoning (Gibbs Jr. 2008; Nersessian 2010). Novel or newly associated categories across social worlds acquire stable formal characteristics as they become important to carrying out and managing routine shared activities. There is now an established trading zone for standardized lists of biological species linking basic research in systematics to applied uses in conservation, for example, but major disagreements remain about whether substantive or coordinative consensus provides the best grounds for meeting the aims of both fields (Franz et al. 2020; Garnett et al. 2020). Coordinative consensus then becomes of practical value when actors maintain fairly stable different ways of talking about the same things and they find their activities depend on reliably understanding what the others are saying or doing.

6.0 Knowledge representation and organization in AI

For centuries, scholars have speculated that what we are calling coordinative consensus may be more readily achieved, and perhaps even be predicated upon, the help of mathematical reasoning, and that it may be further catalyzed by a computational mechanism. Instead of relying on the “recipient’s unilateral capacity effectively (successfully) to interpret the substantive”, one may therefore pin one’s hopes instead on the ability of both transmitter and recipient to compose in, and interpret, a common mathematical language. As early as the 17th century, Gottfried Wilhelm Leibniz imagined a language able to express accurately and unambiguously any thought conceivable to the human mind (Smith 1992), such as an individual’s beliefs, plans, or values. Once so expressed, however, one person’s thoughts may still be beyond the capabilities of any or all other humans to interpret. The intended recipients, for instance may be unable to infer how exactly the communicated sentence in this language relates to their own beliefs, plans, and values. Leibniz had a plan for this: the invention of a set of mathematical rules for manipulating sentences, and also, the invention of a mathematical procedure for applying the rules to these sentences, to determine their truth, or falsehood, or inferential relationship. In light of all this mechanism, a disagreement between any number of humans on any matter may be resolved, so Leibniz hoped, by the rote application of his procedure to the statements: a matter of calculation, not rhetoric.

Here we encounter the first of the two challenges we will address in this section: how does one achieve consensus about the language in which coordinative consensus is to be achieved? Within this first challenge, there are two issues to consider: representation and computation. The quest for a language able to represent "any thought conceivable by the human mind" has proven elusive, but our language must allow the representation of the essential subject-matter of the disagreement at hand. Secondly, the language should ideally be such that the inferential consequentiality of a statement in the language from a set of other statements in the language should be feasible to establish computationally. (We leave aside the issue of whether such a computation may be performed efficiently, and restrict ourselves merely to the question of whether it may be performed at all, or not.)

The second challenge appears once the first has been addressed. Given a language, are there shared representational structures, expressible in that language, upon which both the communicator and the interpreter may agree? If so, what are they? It is common in certain areas of the AI literature to use the word "ontology" to denote these structures. This emphasizes that the sender and receiver may agree upon some shared understanding of reality or concepts describing a specific domain, in order to more effectively represent and interpret thoughts in the shared language. There are several questions we may ask here. Firstly, why are shared representational structures necessary? We must respond to the notion that automated approaches to knowledge organization eliminate the need for social and computational processes of deliberation and agreement about these structures. Further, what shared representational structures may even parties in disagreement agree upon? Do they exist? Is there one or many kinds of such structure?

The language Leibniz envisioned, the *characteristica universalis* (universal characteristic), and the rules that rational argumentation in it was to obey, the *calculus ratiocinator*, foreshadowed the 19th century development of mathematical logic. The rote mathematical procedure Leibniz envisioned for applying this language (and its rules of argumentation) to reason in turn foreshadowed the early 20th century notion of computation. Leibniz's subsequent idea, which may have been exactly the point of the previous ones, of using a machine, such as his own "Stepped Reckoner", to execute this procedure, then foreshadowed the mid-20th century birth of the field of AI. (Leibniz did not require or expect that this machine, a mere computational device, have any real understanding of human thought processes, let alone have any thought processes, in the human sense, of its own. He therefore avoids from the onset any concerns over whether machines can exhibit "real" intelligence in some sense; in modern terms, he envisions a "weak" and not a "strong" AI (Strickland 2014). (See also Searle's "Chinese Room" Argument in Searle (2006).)

The construction of this language, procedure, and machine, was easier said than done, and Leibniz eventually admitted failure (Leibniz, Electress of Hanover and Charlotte 2011). Three and a half centuries later, a truly *characteristica universalis* has yet to be successfully designed. However, there now exists a selection of mathematical languages and corresponding inference calculi, i.e., mathematical *logics*, capable of representing and reasoning over a wide variety of human thoughts (Kutz et al. 2010). For many of these, there also now exists a computational method for determining deductive entailment of its sentences, in modern terms a decision procedure (Boolos, Burgess and Jeffrey 2002). And perhaps most impressively of all, there exists a machine, the *Turing Machine*, able to execute any such computation (Petzold 2008). Indeed, we use the phrase "computational method" to refer to a method computable by a Turing Machine. This assumption, known as the Church-Turing thesis, is widely thought to be very astute, and a practicable device in violation of it is yet to be discovered or engineered, and not for lack of trying (Copeland 2017).

The rest of this section is segmented into four sub-sections, the first two corresponding to the two challenges outlined above. The first subsection addresses representation and computation, i.e., consensus about language. The second sub-section takes up consensus about ontology. The third sub-section comprises of three examples of consensus about both language and ontology. Finally, in the fourth subsection, we discuss the question of whether consensus about the organization of knowledge is even necessary.

6.1 Consensus about mathematical language and logic

We have identified two criteria for choosing the logical language in which consensus is to be achieved:

1. The language must allow the representation of the essential subject-matter of the disagreement at hand.
2. The language should be such that the inferential consequentiality of a statement in the language from a set of other statements in the language should be feasible to establish computationally.

It is well known that, unfortunately, these two criteria are often at odds with each other: highly expressive languages are often such that the second criterion is not met (Levesque 1986). Such languages are commonly termed "undecidable." Another category of languages is "semi-decidable," which is to say that, given enough computational resources, it is possible to determine inferential consequentiality if it holds, but the lack of inferential consequentiality may not be determinable (Boolos, Burgess and Jeffrey 2002). Computable languages, moreover, often do not satisfy the first criterion (Levesque 1986). In addition to this matter, there is the consideration of computational complexity, which dictates how efficiently this computational procedure may be performed, which generally exhibits exactly the same trade-off between expressibility and practicality (Arora and Barak 2009). There is great interest, therefore, in identifying "minimally expressive" languages, i.e., languages barely expressive enough to be sufficient, while retaining therefore the most favorable computability and complexity properties possible under the circumstances (Baader, Horrocks and Sattler 2004).

For example, the language of propositional logic is known to be decidable, whereas the language of classical first-order logic is known to be only semi-decidable (Boolos, Burgess and Jeffrey 2002). Various logics known collectively as description logics have been designed to be minimally expressive languages for various practical representational problems, and compared to the former two logics, are intermediate in their expressiveness (Baader, Horrocks and Sattler 2004). The kinds of statements representable in all of these logics have been studied in depth by computer scientists, and will not be discussed here.

We now look at an example of representation and ontological pluralism in a domain close to Leibniz's heart, and indeed one of the primary motivations for his *characteristica*: the *scientia universalis*, or universal language for the laws of physics. The proper language for representing scientific theories has been debated since the early days of analytic philosophy. The logical positivists, such as the members of the so-called Vienna Circle, propounded their axiomatization in (mostly) classical logic (Uebel 2020). The problem of axiomatizing Einstein's formulation of his theory of relativity, for instance, was taken up first by philosopher Hans Reichenbach (Reichenbach 2006). A more recent approach taken by "the Hungarian group" (Hajnal Andréka, István Németi, Gergely Székely and Judit Madarász, all affiliated with the Alfréd Rényi Institute of Mathematics) seeks to axiomatize Einstein's special and general theories of relativity in first-order logic with identity, and deduce well-known results as theorems of the axioms.

Michéle Friend has pointed out (Friend 2015) a pluralism inherent in the Hungarian group's approach: the axioms are fallible, and not considered laws of physics. Consensus on the axiom system is achieved by considering its deductive closure, i.e., how well it predicts observations and known physical laws. An axiom may be considered falsified if its addition to an ostensibly sound existing axiom system results in the prediction of an untruth. The axiom system is not unique (hence the pluralism), and many different formulations are possible for the same physical theory. Indeed, comparing the deductive closures of multiple axiom systems is an important part of choosing one.

The question of theory representation language is also relevant when AI techniques are employed to induce these theories from relevant data: an inductive paradigm of computational scientific discovery. For example, using inductive logic programming we may choose to induce theories in either first or second-order logic, and this choice has significant impact on both the quality and readability/understandability of the induced theory.

6.2 Consensus about domain-independent ontology

In this section we focus specifically on domain-independent, also known as upper-level, ontologies in the AI literature. In contrast, earlier sections primarily addressed domain-specific ontologies, which are also relevant to the problem of consensus in an analogous manner. Recall that common aims of an upper-level ontologies are to support semantic interoperability across domains and represent consensus categories across viewpoints.

In the AI literature, several ontologies such as SUMO and Upper Cyc have been proposed, typically expressed or expressible in the OWL description logic (Matuszek et al. 2006; Niles and Pease 2001). Also common are ontological frameworks that address specific challenges of representability. We turn to two such examples. What is important in them is that consensus about the representational structures is independent of consensus about the facts they represent; yet, it is this shared structural conceptualization that enables these facts to be effectively represented, i.e., representational consensus in some form was *necessary* to reason in the example domains.

The first example considers the event calculus, which may be used to effectively model time and change. The notorious *frame problem* in AI (Dennett 1984) arises from the monotonicity of classical logic, and deals with the problem of describing the non-effects of actions over time without having to enumerate them. This is therefore a case of representability where the above logics are not sufficient, and we must turn to a "non-classical" logic. (These are also sometimes called philosophical logics, as opposed to mathematical logics, since the mathematical properties of the non-classical logics are not nearly as well developed.)

Cognitive systems in AI have the ability to represent statements about the mental states of human or artificial agents. Mental states such as belief, intention, knowledge and so on, are representable in epistemic logics, most of which are also modal logics (Rendsvig and Symons 2019). Cognitive AI systems thus aim to account for human thought processes, which, as we have stated previously, are representable in epistemic logics, most of which are also modal logics. An additional concern is ethics; ethical concepts are representable in deontic modal logics. Many AI systems need to reason about time, change, and action, and there is a long history of AI engineering to develop ontologies for this purpose (Mueller 2014). The event calculus takes the approach of axiomatizing this domain and reifying fluents (a condition which can change over time), and has been employed by many AI systems, including the original Jeopardy! playing IBM Watson. Tentacular AI envisions a multi-agent architecture where agents reason computationally in DCEC*, an epistemic deontic modal logic with the event calculus, using ShadowProver, a natural-deduction automated theorem prover (Bringsjord et al. 2018). This approach has been applied to problems arising in smart city design and value ethics.

The second example considers the case of qualitative reasoning, which may be used to robustly model uncertain systems. In classical logic, a proposition is either true (valid) or false (invalid): it cannot be uncertain. In many-valued logics there are more truth values. For example, in Kleene logics a proposition may (additionally) be unknown. This is different from being able to represent the degree of certainty to which the truth or falsity, or for that matter, the degree of the state of being unknown, is known i.e. representing uncertainty or vagueness in the extension and nature of referents. Probabilistic logics incorporate a measure of uncertainty, but they are not the only means to do so.

Qualitative reasoning is an ontological framework for representing quantities that are only qualitatively known, thereby providing an alternative to estimating numerical probabilities for representing uncertainty in AI systems (Forbus 1997). Consensus about such an ontology might consist of consensus about symbolic or subjective probabilities, such as degrees of belief. This approach has been used to achieve consensus on the resolution of well-known paradoxes of rationality such as the lottery paradox. Another example is qualitative spatial reasoning, i.e., reasoning about spatial relationships that are only qualitatively known. Reasoning with qualitative spatial calculi has been employed in applications as diverse as reasoning about geo-spatial entities and about biological taxonomies (Franz et al. 2016; Cheng and Ludäscher 2019).

6.3 Can AI replace the need for consensus?

Having reviewed some of the increasingly sophisticated ways that AI aspires to Leibniz's dream of a universal language for communication and reasoning, one might reasonably wonder whether AI could in fact eliminate the need for consensus in scientific classification. In brief: is the problem of figuring out how to share classifications just for humans, or computers, too? We close with some reflections on this question and use it to highlight the importance of human-computer interaction in the future of classification as part of the scientific process. In his article "Is classification necessary after Google?", Birger Hjørland (Hjørland 2012) considers the question of whether the activity of classification is even worth the effort, given the existence of search engines such as Google, which "students use far more than they use library catalogs in order to find what they need." As Hjørland points out, this is a question of whether computer algorithms are able to do a "100% satisfactory job" at the task of information retrieval, without the need for classification.

The general problem of information retrieval is to identify a subset of information which is relevant for an information need within a large amount of knowledge. What counts as relevant is of course highly situational, reflecting the urgency, rigor, and evaluative criteria of what the user is trying to do as well as their background knowledge and expertise. Similarly, the best method for extracting relevant results will also depend on the features of the pooled information and processing resources available; one can look, for example, to the challenges that volume, variety, and velocity pose for big data analysis (Kitchin 2014). As Hjørland points out, even in human classification, criteria are often an empirical and subjective matter.

From this perspective, information retrieval is an AI-complete problem, which among other skills, requires commonsense knowledge and cognitive computation to attempt to solve. It has been extensively argued that knowledge organization is necessary for general artificial intelligence, although it remains a controversial issue (Marcus and Davis 2019). This is not to mention that the problems of computer vision and natural language understanding, both arguably AI-complete tasks, may be viewed as subsets of the IR problem.

Nonetheless, sufficient regularities exist in these contextual aspects of information retrieval that pragmatically satisfactory solutions are common. From a sociological perspective, this is no accident: in addition to investing great efforts into constructing our concepts and environments so as to make our aims achievable, we are also regularized by our cultures and the physical environments. People moreover often revise the scope and content of their aims to match what's possible. Computational agents are not so different from humans in being bound by these social and environmental incentives and constraints, although humans and computational agents frequently differ on which are most important (Sterner and Franz 2017). In this respect, the process of seeking consensus on classification can be viewed as making explicit and formalizing shared infrastructure that enables individual and collective action among both humans and computers (Wilkinson et al. 2016).

One remaining and important distinction to make is then between knowledge organized by humans versus machines, i.e. by machine learning. This is common in information retrieval systems: the Google search engine, for example, relies on the Google Knowledge Graph, which has been learnt from "crawling" the web. The creation of this knowledge graph certainly involves the classification of knowledge into categories and instances, though by an algorithm and not a human. In this respect, it may only be possible to answer questions such as Hjørland's "How do we decide the criteria for assigning document A to class X?" only at the meta-level of the training of a machine learning algorithm; that is to say, "whatever criteria has been determined by such-and-such algorithm".

To sum up, then, we agree with Hjørland that classification remains necessary for information retrieval, but its implications for consensus are less clear. If one accepts the necessity of knowledge organization for artificial general intelligence, the conclusion that it is necessary for information retrieval is inescapable. Consensus in an AI-driven context, however, probably won't take the form of agreeing on a shared classification, per se, but rather agreeing on using the output of an algorithm and the language in which that output is expressed. One might call this a procedural rather than substantive consensus classification, since the computationally produced ontological structure itself can be highly pluralistic, and may or may not be human-understandable. Such a procedural context poses novel challenges for consensus at both first and second orders as we've discussed here.

7.0 Conclusion

We have distinguished two types of consensus related to scientific classifications: consensus on the classificatory categories for a research domain versus consensus on how to represent the conflicts between alternative classifications for a domain. We labeled these types consensus classification and coordinative consensus, respectively. We then summarized a series of arguments against the assumption that consensus classification is achievable or desirable for the aims of science, drawing on both philosophical and empirical studies in the literature. This motivated a deeper investigation of coordinative consensus, which preserves a community's ability to communicate and advance shared inquiry despite potentially fundamental disagreements. We presented examples of scientific communities exploring pluralistic classification practices and then reviewed relevant approaches from knowledge representation and AI that enable communication without presupposing first-order agreement on how information should be presented and interpreted.

Our approach has been to examine how consensus figures in the background assumptions of the aims and development of scientific classifications. One common assumption about science, for example, holds that it will eventually uncover a single true or at least most empirically adequate representation of nature. Additionally, the evidence supporting this best representation would be sufficiently strong to create consensus among researchers, including consensus on scientific classifications as a corollary. However, the examples of scientific practice and in-principle arguments we surveyed raise serious doubts about this assumption as a universal guide to either current research today and its ultimate future outcome.

In response, we suggest adopting a deliberative stance toward the driving question we posed in the introduction: "What sort of consensus, if any, is the best basis for communicating and reasoning with scientific classifications in a domain?" In any real situation, there are almost certainly some terms on which researchers working in a domain can agree, but there will also be others that resist incorporation into a consensus classification. The reasons for this may be heterogeneous, as suggested by the arguments for pluralism we surveyed above, but making a choice about what to do with the categories that escape first-order consensus is unavoidable: are they to be excluded as intractable or made interoperable through the alignment and translation methods we discussed for coordinative consensus? Is the goal to converge on a consensus classification as fast as possible or to maximize the benefits of coordination given the likely persistence of substantial dissensus for the foreseeable future?

Most discussions of ontology development, for example, fail to distinguish rigorously between the long-term goal of converging on a comprehensive consensus classification for a domain versus the short-term goal of developing a minimal standard classification that provides narrow support for widely shared aims such as data discovery and aggregation. Identifying areas of limited or partial agreement can just as well serve the goal of harnessing sustained pluralism within a domain by enabling an improved coordinative consensus. Understanding which options are available to a community, and how present actions are related to future goals, is essential to meaningful deliberation and discussion. It also relies on systematic knowledge of the aims, methods, and resources available to stakeholders in scientific classifications that is hard for any single researcher in the domain to acquire. Our discussion has hopefully highlighted important points of contact between diverse, interdisciplinary bodies of literature that deserve further development, including how the pursuit of the two types of consensus works or doesn't in particular domains and how computational methods and infrastructures can best support the aims of classification developers and stakeholders.

References

Altschul, Jeffrey H., Keith W. Kintigh, Terry H. Klein, William H. Doelle, Kelley A. Hays-Gilpin, Sarah A. Herr, Timothy A. Kohler et al. 2017. "Opinion: Fostering Synthesis in Archaeology to Advance Science and Benefit Society." *Proceedings of the National Academy of Sciences* 114 (42): 10999–11002. https://doi.org/10.1073/pnas.1715950114

Altschul, Jeffrey H., Keith W. Kintigh, Terry H. Klein, William H. Doelle, Kelley A. Hays-Gilpin, Sarah A. Herr, Timothy A. Kohler et al. 2018. "Fostering Collaborative Synthetic Research in Archaeology." *Advances in Archaeological Practice* 6 (1): 19–29. https://doi.org/10.1017/aap.2017.31

Arora, Sanjeev and Boaz Barak. 2009. *Computational Complexity: A Modern Approach*. Cambridge, UK: Cambridge University Press.

Arp, Robert, Barry Smith and Andrew D. Spear. 2015. *Building Ontologies with Basic Formal Ontology*. Cambridge, MA: MIT Press.

Baader, Franz, Ian Horrocks and Ulrike Sattler. 2004. "Description Logics." In *Handbook on Ontologies*, edited by Steffen Staab and Rudi Studer. International Handbooks on Information Systems. Berlin: Springer, 3–28. https://doi.org/10.1007/978-3-540-24750-0_1

Beatty, John and Alfred Moore. 2010. "Should We Aim for Consensus?" *Episteme* 7: 198–214. https://doi.org/10.3366/E1742360010000948

Beck, Silke, Maud Borie, Jason Chilvers, Alejandro Esguerra, Katja Heubach, Mike Hulme, Rolf Lidskog et al. 2014. "Towards a Reflexive Turn in the Governance of Global Environmental Expertise. The Cases of the IPCC and the IPBES." *GAIA - Ecological Perspectives for Science and Society* 23: 80–7. https://doi.org/10.14512/gaia.23.2.4

Biagetti, Maria Teresa. 2020. "Ontologies (as Knowledge Organization Systems)." In *ISKO Encyclopedia of Knowledge Organization*, edited by Birger Hjorland and Claudio Gnoli, https://www.isko.org/cyclo/ontologies#5

Binney, Nicholas. 2015. "Nosology, Ontology and Promiscuous Realism." *Journal of Evaluation in Clinical Practice* 21: 391–7. https://doi.org/10.1111/jep.12274

Bliss, Henry Evelyn. 1929. *The Organization of Knowledge and the System of the Sciences*. New York: Henry Holt. http://hdl.handle.net/2027/mdp.39015011706697.

Blomqvist, Eva, Aldo Gangemi and Valentina Presutti. 2009. "Experiments on Pattern-Based Ontology Design." In *Proceedings of the Fifth International Conference on Knowledge Capture*, K-CAP '09. New York, NY, USA: Association for Computing Machinery, 41–48. https://doi.org/10.1145/1597735.1597743

Boolos, George S., John P. Burgess and Richard C. Jeffrey. 2002. *Computability and Logic*. Cambridge, UK: Cambridge University Press.

Bouwel, Jeroen Van and Michiel Van Oudheusden. 2017. "Participation Beyond Consensus? Technology Assessments, Consensus Conferences and Democratic Modulation." *Social Epistemology* 31: 497–513. https://doi.org/10.1080/02691728.2017.1352624

Bowker, Geoffrey C. 2000. "Biodiversity Datadiversity." *Social Studies of Science* 30: 643–83. https://doi.org/10.1177/030631200030005001

Bowker, Geoffrey C., Karen Baker, Florence Millerand, and David Ribes. 2009. "Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment." In *International Handbook of Internet Research*, edited by Jeremy Hunsinger, Lisbeth Klastrup, and Matthew Allen. Dordrecht: Springer, 97–117. https://doi.org/10.1007/978-1-4020-9789-8_5

Bowker, Geoffrey C, and Susan Leigh Star. 1999. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press. http://cognet.mit.edu/book/sorting-things-out.

Bringsjord, Selmer, Naveen Sundar Govindarajulu, Atriya Sen, Matthew Peveler, Biplav Srivastava and Kartik Talamadupula. 2018. "Tentacular Artificial Intelligence, and the Architecture Thereof, Introduced." *ArXiv:1810.07007 [Cs]*, October. http://arxiv.org/abs/1810.07007.

Brooy, Camille La, Bridget Pratt and Margaret Kelaher. 2020. "What Is the Role of Consensus Statements in a Risk Society?" *Journal of Risk Research* 23: 664-77. https://doi.org/10.1080/13669877.2019.1628094

Ceusters, W., B. Smith and L. Goldberg. 2005. "A Terminological and Ontological Analysis of the NCI Thesaurus." *Methods of Information in Medicine* 44: 498–507. https://doi.org/10.1055/s-0038-1634000

Chase, Adrian S. Z., Diane Chase and Arlen Chase. 2020. "Ethics, New Colonialism, and Lidar Data: A Decade of Lidar in Maya Archaeology." *Journal of Computer Applications in Archaeology* 3: 51–62. https://doi.org/10.5334/jcaa.43

Cheng, Yi-Yun and Bertram Ludäscher. 2019. "Exploring Geopolitical Realities through Taxonomies: The Case of Taiwan." In *Proceedings of North American Symposium on Knowledge Organization 2019*. North American Symposium on Knowledge Organization (NASKO). https://www.ideals.illinois.edu/handle/2142/104013

Conix, Stijn. 2019. "Radical Pluralism, Classificatory Norms and the Legitimacy of Species Classifications." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 73 (February): 27–34. https://doi.org/10.1016/j.shpsc.2018.11.002

Cooper, Rachel. 2005. *Classifying Madness: A Philosophical Examination of the Diagnostic and Statistical Manual of Mental Disorders*. Dordrecht: Springer. https://doi.org/10.1007/1-4020-3345-1

Cooper, Rachel. 2017a. "A Modest Proposal." *Philosophy, Psychiatry, & Psychology* 24: 207–9. https://doi.org/10.1353/ppp.2017.0026

Cooper, Rachel. 2017b. "Diagnostic and Statistical Manual of Mental Disorders (DSM)." In *ISKO Encyclopedia of Knowledge Organization*, edited by Birger Hjorland and Claudio Gnoli. https://www.isko.org/cyclo/dsm.

Copeland, B. Jack. 2017. "The Church-Turing Thesis." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2017/entries/church-turing/

Crețu, Ana-Maria, and Michela Massimi (Eds.). 2020. *Knowledge from a Human Point of View*. Cham: Springer International. https://doi.org/10.1007/978-3-030-27041-4

Danks, David. 2015. "Goal-Dependence in (Scientific) Ontology." *Synthese* 192: 3601–16.

Deans, Andrew R., Matthew J. Yoder and James P. Balhoff. 2012. "Time to Change How We Describe Biodiversity." *Trends in Ecology & Evolution* 27(2): 78–84. https://doi.org/10.1016/j.tree.2011.11.007

Dennett, Daniel C. 1984. "Cognitive Wheels: The Frame Problem of AI." In *Minds, Machines and Evolution: Philosophical Studies*, edited by C. Hookway. Cambridge, UK: Cambridge University Press.

DeVorkin, David H. 1981. "Community and Spectral Classification in Astrophysics: The Acceptance of E. C. Pickering's System in 1910." *Isis*: 29–49.

Díaz-Reviriego, I., E. Turnhout and S. Beck. 2019. "Participation and Inclusiveness in the Intergovernmental Science–Policy Platform on Biodiversity and Ecosystem Services." *Nature Sustainability* 2: 457–64. https://doi.org/10.1038/s41893-019-0290-6

Dick, Steven J. 2013. *Discovery and Classification in Astronomy: Controversy and Consensus*. Cambridge, UK: Cambridge University Press.

Dick, Steven J. 2019. "Astronomy's Three Kingdom System." In *ISKO Encyclopedia of Knowledge Organization*, edited by Birger Hjorland and Claudio Gnoli. https://www.isko.org/cyclo/3ks

Diller, Matthew, Evan Johnson, Amanda Hicks and William R. Hogan. 2020. "A Realism-Based Approach to an Ontological Representation of Symbiotic Interactions." *BMC Medical Informatics and Decision Making* 20: 258. https://doi.org/10.1186/s12911-020-01273-0

Dupré, John. 1995. *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, MA: Harvard University Press.

Ereshefsky, Marc. 1992. "Eliminative Pluralism." *Philosophy of Science* 59: 671–90. https://doi.org/10.1086/289701

Forbus, Kenneth D. 1997. "Qualitative Reasoning." In *CRC Handbook of Computer Science and Engineering*, edited by A. B. Tucker. Boca Raton, FL: CRC Press, 715–33.

Franz, Nico M., Mingmin Chen, Shizhuo Yu, Parisa Kianmajd, Shawn Bowers and Bertram Ludäscher. 2015. "Reasoning over Taxonomic Change: Exploring Alignments for the Perelleschus Use Case." *PloS One* 10(2): e0118247

Franz, Nico M. and Adam M. Goldstein. 2013. "Phenotype Ontologies: Are Homology Relations Central Enough? A Reply to Deans et Al." *Trends in Ecology & Evolution* 28(3): 131–32. https://doi.org/10.1016/j.tree.2012.08.001

Franz, Nico M. and R. K. Peet. 2009. "Perspectives: Towards a Language for Mapping Relationships among Taxonomic Concepts." *Systematics and Biodiversity* 7(1): 5–20.

Franz, Nico M., Naomi M. Pier, Deeann M. Reeder, Mingmin Chen, Shizhuo Yu, Parisa Kianmajd, Shawn Bowers and Bertram Ludäscher. 2016. "Two Influential Primate Classifications Logically Aligned." *Systematic Biology* 65: 561–82.

Franz, Nico M., Beckett Sterner, Nathan Upham and Kevin Cortés Hernández. 2020. "Redesigning the Trading Zone between Systematics and Conservation: Insights from Malagasy Mouse Lemur Classifications, 1982 to Present." *Biodiversity Information Science and Standards* 4: e59234.

Franz, Nico M. and Beckett W Sterner. 2018. "To Increase Trust, Change the Social Design behind Aggregated Biodiversity Data." *Database* 2018. https://doi.org/10.1093/database/bax100

Fraser, Alistair. 2019. "Land Grab/Data Grab: Precision Agriculture and Its New Horizons." *The Journal of Peasant Studies* 46: 893–912. https://doi.org/10.1080/03066150.2017.1415887

Fricke, Martin. 2016. "Logical Division." In *ISKO Encyclopedia of Knowledge Organization*, edited by Birger Hjorland and Claudio Gnoli. https://www.isko.org/cyclo/logical_division

Friend, Michéle. 2015. "On the Epistemological Significance of the Hungarian Project." *Synthese* 192: 2035–51.

Fuller, Steve. 2002. *Social Epistemology*. Bloomington, IN: Indiana University Press.

Garfield, Eugene. 1974. "The 'Other' Immortal: a Memorable Day with Henry E. Bliss". *Wilson Library Bulletin* 49: 288-92.

Garnett, Stephen T., Neil D. Burgess, John E. Fa, Álvaro Fernández-Llamazares, Zsolt Molnár, Cathy J. Robinson, James E. M. Watson et al. 2018. "A Spatial Overview of the Global Importance of Indigenous Lands for Conservation." *Nature Sustainability* 1: 369–74. https://doi.org/10.1038/s41893-018-0100-6

Garnett, Stephen T. and Les Christidis. 2017. "Taxonomy Anarchy Hampers Conservation." *Nature* 546 (7656): 25–7. https://doi.org/10.1038/546025a

Garnett, Stephen T., Les Christidis, Stijn Conix, Mark J. Costello, Frank E. Zachos, Olaf S. Bánki, Yiming Bao et al. 2020. "Principles for Creating a Single Authoritative List of the World's Species." *PLOS Biology* 18(7): e3000736. https://doi.org/10.1371/journal.pbio.3000736

Gerson, Elihu M. 2008. "Reach, Bracket, and the Limits of Rationalized Coordination: Some Challenges for CSCW." In *Resources, Co-Evolution and Artifacts. Computer Supported Cooperative Work*. London: Springer, 193–220. https://doi.org/10.1007/978-1-84628-901-9_8

Gibbs Jr., Raymond W. 2008. *The Cambridge Handbook of Metaphor and Thought*. Cambridge University Press.

Gnoli, Claudio. 2018. "Genealogical Classification (IEKO)." In *ISKO Encyclopedia of Knowledge Organization*, edited by Birger Hjorland and Claudio Gnoli. https://www.isko.org/cyclo/genealogical.

Golding, Clinton. 2013. "We Made Progress: Collective Epistemic Progress in Dialogue without Consensus." *Journal of Philosophy of Education* 47: 423–40. https://doi.org/10.1111/1467-9752.12010

Groß, Anika, Cédric Pruski and Erhard Rahm. 2016. "Evolution of Biomedical Ontologies and Mappings: Overview of Recent Approaches." *Computational and Structural Biotechnology Journal* 14 (January): 333–40. https://doi.org/10.1016/j.csbj.2016.08.002

Gupta, Prashant and Mark Gahegan. 2020. "Categories Are in Flux, but Their Computational Representations Are Fixed: That's a Problem." *Transactions in GIS* 24: 291–314. https://doi.org/10.1111/tgis.12602

Habermas, Jürgen. 1971. *Toward a Rational Society: Student Protest, Science, and Politics*. Boston, MA: Beacon Press.

Hacking, Ian. 1991. "A Tradition of Natural Kinds." *Philosophical Studies* 61: 109–26. https://doi.org/10.1007/BF00385836

Hjørland, Birger. 2012. "Is Classification Necessary after Google?" *Journal of Documentation* 68: 299–317. https://doi.org/10.1108/00220411211225557

Hjørland, Birger. 2016. "Knowledge Organization." In *Encyclopedia of Knowledge Organization*, edited by Birger Hjorland and Claudio Gnoli. https://www.isko.org/cyclo/knowledge_organization

Hjørland, Birger. 2017. "Classification." *Knowledge Organization* 44: 97–128.

Hjørland, Birger, and Claudio Gnoli. 2020. "Statistical Classification." In *ISKO Encyclopedia of Knowledge Organization*, edited by Birger Hjorland and Claudio Gnoli. https://www.isko.org/cyclo/statistical

Hogan, William R. and Werner Ceusters. 2016. "Diagnosis, Misdiagnosis, Lucky Guess, Hearsay, and More: An Ontological Analysis." *Journal of Biomedical Semantics* 7: 54. https://doi.org/10.1186/s13326-016-0098-5

Hull, David. 1987. "Genealogical Actors in Ecological Roles." *Biology and Philosophy* 2: 168–84. https://doi.org/10.1007/BF00057961

Keil, Geert, Lara Keuck and Rico Hauswald. 2017. *Vagueness in Psychiatry*. Oxford: Oxford University Press.

Kendler, Kenneth S. and Josef Parnas (Eds.). 2012. *Philosophical Issues in Psychiatry II: Nosology*. Oxford: Oxford University Press.

Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. London: SAGE. https://doi.org/10.4135/9781473909472

Lee, Deborah. 2019. "Hornbostel-Sachs Classification." In *ISKO Encyclopedai of Knowledge Organization*, edited by Birger Hjorland and Claudio Gnoli. https://www.isko.org/cyclo/hornbostel

Leibniz, Gottfried Wilhelm, Sophie Electress of Hanover and Queen Sophie Charlotte. 2011. *Leibniz and the Two Sophies: The Philosophical Correspondence*. Translated by Lloyd Strickland. Toronto: Iter Incorporated.

Leonelli, Sabina. 2008. "Bio-Ontologies as Tools for Integration in Biology." *Biological Theory* 3: 7–11. https://doi.org/10.1162/biot.2008.3.1.7

Leonelli, Sabina. 2013a. "Classificatory Theory in Biology." *Biological Theory* 7: 338–45. https://doi.org/10.1007/s13752-012-0049-z

Leonelli, Sabina. 2013b. "Global Data for Local Science: Assessing the Scale of Data Infrastructures in Biological and Biomedical Research." *BioSocieties* 8: 449–65. https://doi.org/10.1057/biosoc.2013.23

Leonelli, Sabina. 2016. *Data-Centric Biology: A Philosophical Study*. University of Chicago Press.

Levesque, H. J. 1986. "Knowledge Representation and Reasoning." *Annual Review of Computer Science* 1: 255–87. https://doi.org/10.1146/annurev.cs.01.060186.001351

Ludwig, David. 2018. "Letting Go of 'Natural Kind': Toward a Multidimensional Framework of Nonarbitrary Classification." *Philosophy of Science* 85: 31–52. https://doi.org/10.1086/694835

Ludwig, David and Charbel N. El-Hani. 2020. "Philosophy of Ethnobiology: Understanding Knowledge Integration and Its Limitations." *Journal of Ethnobiology* 40: 3–20. https://doi.org/10.2993/0278-0771-40.1.3

Marcus, Gary and Ernest Davis. 2019. *Rebooting AI: Building Artificial Intelligence We Can Trust*. Knopf Doubleday.

Martínez-Ávila, Daniel and John M. Budd. 2017. "Epistemic Warrant for Categorizational Activities and the Development of Controlled Vocabularies." *Journal of Documentation* 73: 700-15. https://doi.org/10.1108/JD-10-2016-0129

Mattu, Jaipreet, and Jacqueline A. Sullivan. 2020. "Classification, Kinds, Taxonomic Stability and Conceptual Change." *Aggression and Violent Behavior* July 101477. https://doi.org/10.1016/j.avb.2020.101477

Matuszek, Cynthia, Michael Witbrock, John Cabral and John DeOliveira. 2006. "An Introduction to the Syntax and Content of Cyc." *UMBC Computer Science and Electrical Engineering Department Collection*.

McAllister, Tara, Jacqueline Beggs, Shaun Ogilvie, Rauru Kirikiri, Amanda Black and Priscilla Wehi. 2020. "Kua Takoto Te Mānuka: Mātauranga Māori in New Zealand Ecology." *New Zealand Journal of Ecology* 43(3): 3393. https://doi.org/10.20417/nzjecol.43.41

McManamon, Francis P., Keith W. Kintigh, Leigh Anne Ellison and Adam Brin. 2017. "TDAR: A Cultural Heritage Archive for Twenty-First-Century Public Outreach, Research, and Resource Management." *Advances in Archaeological Practice* 5: 238–49. https://doi.org/10.1017/aap.2017.18

Miller, Boaz. 2013. "When Is Consensus Knowledge Based? Distinguishing Shared Knowledge from Mere Agreement." *Synthese* 190: 1293–316. https://doi.org/10.1007/s11229-012-0225-5

Miller, Boaz. 2019. "The Social Epistemology of Consensus and Dissent." In *The Routledge Handbook of Social Epistemology*, edited by David Henderson, Peter Graham, Miranda Fricker and Nikolaj Jang Lee Linding Pedersen. New York: Routledge, 228–37.

Mitchell, Sandra D. 2003. *Biological Complexity and Integrative Pluralism*. Cambridge: Cambridge University Press.

Mitchell, Sandra D. 2009. *Unsimple Truths: Science, Complexity, and Policy*. University of Chicago Press.

Morrison, Margaret. 2011. "One Phenomenon, Many Models: Inconsistency and Complementarity." *Studies in History and Philosophy of Science Part A*, *Model-Based Representation in Scientific Practice* 42: 342–51. https://doi.org/10.1016/j.shpsa.2010.11.042

Muldoon, Ryan. 2016. *Social Contract Theory for a Diverse World: Beyond Tolerance*. New York: Taylor & Francis.

Müller-Wille, Staffan. 2013. "Systems and How Linnaeus Looked at Them in Retrospect." *Annals of Science* 70: 305–17. https://doi.org/doi:10.1080/00033790.2013.783109

Mungall, Christopher J. 2019. "Never Mind the Logix: Taming the Semantic Anarchy of Mappings in Ontologies." *Monkeying around with OWL* (blog). May 27, 2019. https://douroucouli.wordpress.com/2019/05/27/never-mind-the-logix-taming-the-semantic-anarchy-of-mappings-in-ontologie/

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

Niemeyer, Simon and John S. Dryzek. 2007. "The Ends of Deliberation: Meta-Consensus and Inter-Subjective Rationality as Ideal Outcomes." *Swiss Political Science Review* 13: 497–526. https://doi.org/10.1002/j.1662-6370.2007.tb00087.x

Niles, Ian and Adam Pease. 2001. "Towards a Standard Upper Ontology." In *Proceedings of the International Conference on Formal Ontology in Information Systems.* Volume 2001, 2–9.

Oreskes, Naomi. 2007. "The Long Consensus on Climate Change," February 1, 2007. http://www.washingtonpost.com/wp-dyn/content/article/2007/01/31/AR2007013101808.html

Petzold, Charles. 2008. *The Annotated Turing: A Guided Tour Through Alan Turing's Historic Paper on Computability and the Turing Machine*. Indianapolis: Wiley.

Phillips, James, Allen Frances, Michael A. Cerullo, John Chardavoyne, Hannah S. Decker, Michael B. First, Nassir Ghaemi et al. 2012. "The Six Most Essential Questions in Psychiatric Diagnosis: A Pluralogue Part 1: Conceptual and Definitional Issues in Psychiatric Diagnosis." *Philosophy, Ethics, and Humanities in Medicine* 7: 3. https://doi.org/10.1186/1747-5341-7-3

Poland, Jeffrey and Şerife Tekin (Eds.). 2017. *Extraordinary Science and Psychiatry: Responses to the Crisis in Mental Health Research*. MIT Press. https://doi.org/10.7551/mitpress/10304.001.0001

Reichenbach, Hans. 2006. *Defending Einstein: Hans Reichenbach's Writings on Space, Time and Motion*. Cambridge, UK: Cambridge University Press.

Rendsvig, Rasmus and John Symons. 2019. "Epistemic Logic." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2019/entries/logic-epistemic/

Rescher, Nicholas. 2000. *Pluralism: Against the Demand for Consensus*. Clarendon Library of Logic and Philosophy. Oxford: Clarendon Press.

Rescher, Nicholas. 2006. *Philosophical Dialectics: An Essay on Metaphilosophy*. Ithaca, NY: State University of New York Press.

Ribes, David and Geoffrey C. Bowker. 2009. "Between Meaning and Machine: Learning to Represent the Knowledge of Communities." *Information and Organization* 19: 199–217. https://doi.org/10.1016/j.infoandorg.2009.04.001

Rubis, June Mary and Noah Theriault. 2020. "Concealing Protocols: Conservation, Indigenous Survivance, and the Dilemmas of Visibility." *Social & Cultural Geogra-*

phy 21: 962–84. https://doi.org/10.1080/14649365.2019.1574882

Rye, Ståle Angen and Nanang Indra Kurniawan. 2017. "Claiming Indigenous Rights through Participatory Mapping and the Making of Citizenship." *Political Geography* 61 (November): 148–59. https://doi.org/10.1016/j.polgeo.2017.08.008

Scerri, Eric R. 2006. *The Periodic Table: Its Story and Its Significance*. Oxford: Oxford University Press.

Searle, John. 2006. "Chinese Room Argument, The." In *Encyclopedia of Cognitive Science*. Wiley Online. https://doi.org/10.1002/0470018860.s00159

Seltmann, Katja, Matthew Yoder, Istvan Miko, Mattias Forshage, Matthew Bertone, Donat Agosti, Andrew Austin et al. 2012. "A Hymenopterists' Guide to the Hymenoptera Anatomy Ontology: Utility, Clarification, and Future Directions." *Journal of Hymenoptera Research* 27 (May): 67–88. https://doi.org/10.3897/jhr.27.2961

Sen, Atriya, Nico M. Franz, Beckett Sterner and Nate Upham. 2020. "The Automated Taxonomic Concept Reasoner." *Biodiversity Information Science and Standards* 4: e59074.

Slater, Matthew H. 2017. "Pluto and the Platypus: An Odd Ball and an Odd Duck - On Classificatory Norms." *Studies in History and Philosophy of Science Part A* 61 (February): 1–10. https://doi.org/10.1016/j.shpsa.2017.01.001

Smith, Barry. 1987. "Husserl, Language and the Ontology of the Act." In *Speculative Grammar, Universal Grammar, and Philosophical Analysis of Language,* edited by Dino Buzzetti and Maurizio Ferriani. Amsterdam: John Benjamins, 205–27.

Smith, Barry. 1992. "Characteristica Universalis." In *Language, Truth and Ontology*, edited by Kevin Mulligan. Dordrecht: Springer, 48–77.

Smith, Barry. 2006. "From Concepts to Clinical Reality: An Essay on the Benchmarking of Biomedical Terminologies." *Journal of Biomedical Informatics* 39: 288–98. https://doi.org/10.1016/j.jbi.2005.09.005

Smith, Barry, Michael Ashburner, Cornelius Rosse, Jonathan Bard, William Bug, Werner Ceusters, Louis J Goldberg et al. 2007. "The OBO Foundry: Coordinated Evolution of Ontologies to Support Biomedical Data Integration." *Nature Biotechnology* 25: 1251–5. https://doi.org/10.1038/nbt1346

Smith, Barry and Werner Ceusters. 2003. "Towards Industrial Strength Philosophy: How Analytical Ontology Can Help Medical Informatics." *Interdisciplinary Science Reviews* 28: 106–11. https://doi.org/10.1179/030801803225010421

Smith, Barry and Werner Ceusters. 2006. "HL7 RIM: An Incoherent Standard," *Studies in Health Technology and Informatics* 124: 133–8.

Star, Susan Leigh and James R Griesemer. 1989. "Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebate Zoology, 1907-39." *Social Studies of Science* 19: 387–420.

Stegenga, Jacob. 2016. "Three Criteria for Consensus Conferences." *Foundations of Science* 21: 35–49. https://doi.org/10.1007/s10699-014-9374-y

Sterner, Beckett and Nico M. Franz. 2017. "Taxonomy for Humans or Computers? Cognitive Pragmatics for Big Data." *Biological Theory* 12: 99–111. https://doi.org/10.1007/s13752-017-0259-5

Sterner, Beckett, Edward E. Gilbert, and Nico M. Franz. 2020. "Decentralized but Globally Coordinated Biodiversity Data." *Frontiers in Big Data* 3. https://doi.org/10.3389/fdata.2020.519133

Sterner, Beckett, Joeri Witteveen and Nico Franz. 2020. "Coordinating Dissent as an Alternative to Consensus Classification: Insights from Systematics for Bio-Ontologies." *History and Philosophy of the Life Sciences* 42: 8. https://doi.org/10.1007/s40656-020-0300-z

Strickland, Lloyd. 2014. *Leibniz's Monadology*. Edinburgh: Edinburgh University Press.

Sullivan, Jacqueline Anne. 2017. "Coordinated Pluralism as a Means to Facilitate Integrative Taxonomies of Cognition." *Philosophical Explorations* 20: 129–45. https://doi.org/10.1080/13869795.2017.1312497

Tabb, Kathryn. 2015. "Psychiatric Progress and the Assumption of Diagnostic Discrimination." *Philosophy of Science* 82: 1047–58. https://doi.org/10.1086/683439

Tengö, Maria, Rosemary Hill, Pernilla Malmer, Christopher M Raymond, Marja Spierenburg, Finn Danielsen, Thomas Elmqvist and Carl Folke. 2017. "Weaving Knowledge Systems in IPBES, CBD and beyond—Lessons Learned for Sustainability." *Current Opinion in Environmental Sustainability* 26–27: 17–25. https://doi.org/10.1016/j.cosust.2016.12.005

Tsou, Jonathan Y. 2016. "Natural Kinds, Psychiatric Classification and the History of the DSM." *History of Psychiatry* 27: 406–24. https://doi.org/10.1177/0957154X16656580

Tucker, Aviezer. 2003. "The Epistemic Significance of Consensus." *Inquiry* 46: 501–21. https://doi.org/10.1080/00201740310003388

Uebel, Thomas. 2020. "Vienna Circle." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2020/entries/vienna-circle/

Walter, Maggie, Tahu Kukutai, Stephanie Russo Carroll, Desi Rodriguez-Lonebear, Maggie Walter, Tahu Kukutai, Stephanie Russo Carroll and Desi Rodriguez-Lonebear (Eds.). 2020. *Indigenous Data Sovereignty and Pol-*

icy. New York: Routledge. https://doi.org/10.4324/9780429273957

Weiskopf, Daniel A. 2020. "Representing and Coordinating Ethnobiological Knowledge." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* August: 101328. https://doi.org/10.1016/j.shpsc.2020.101328

Wilkins, John and Malte Ebach. 2013. *The Nature of Classification: Relationships and Kinds in the Natural Sciences*. London: Palgrave Macmillan.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3: 1–9. https://doi.org/10.1038/sdata.2016.18

Yeboah, A. A. O., R. Hill, C. Adem, W. V. Alangui, Z. Molnar, Y. Aumeeruddy-Thomas, P. Bridgewater et al. 2020. "Working with Indigenous, Local and Scientific Knowledge in Assessments of Nature and Nature's Linkages with People." *Current Opinion in Environmental Sustainability* 43: 8–20. https://doi.org/10.1016/j.cosust.2019.12.006

Zalta, E. 1983. *Abstract Objects: An Introduction to Axiomatic Metaphysics*. Dordrecht: Springer. https://doi.org/10.1007/978-94-009-6980-3

Zeng, Marcia. 2019. "Interoperability." In *ISKO Encyclopedia of Knowledge Organization*, edited by Birger Hjørland and Claudio Gnoli. https://www.isko.org/cyclo/interoperability