

Politics and Pragmatism in Scientific Ontology Construction

Michael Travers
SRI International
mt@alum.mit.edu

19 June 2011

for Inconsistency Robustness 2011¹

Abstract

Ontologies are supposed to define a common representational framework for a domain of knowledge, but in practice achieving commonality is not an easy task. It might seem that mutual agreement about ontology ought to be an inevitable result of the structure of the world, but in practice it is often a laborious and contentious process, filled with argument, negotiations, compromise, power, and politics. Recent debates in the ontology community between “realists” and “conceptualists” hide the true nature of ontology building, and I will examine pragmatism as an alternate approach that is less brittle and more robust. One of the roots of the pragmatic approach is to look at ontologies as a social product, with an attention to how they are constructed, promulgated, and used. I’ll look at some examples of the sociology of ontology from my personal work experience building knowledge-based systems for scientists. One of the lessons learned is that ontologies in practice are influenced by a wide variety of factors other than pure logic, such as cost of implementation, the relative status of particular users, and user interface considerations. Inconsistencies and conflicts are the norm rather than exceptions. The pragmatic view says that rather than bemoan these impurities, or marginalizing them, we should explicitly acknowledge and make use of them.

¹Copyright of this paper belongs to the author, who hereby grants permission to Stanford University to include it in the conference material and proceedings of Inconsistency Robustness 2011 and to place it on relevant websites.

1 Ontologies have politics

The process of dividing up the world into well-defined objects and categories is variously referred to as ontology, model building, or simply classification. Socrates in *Phaedrus* characterized this as “carving nature at its joints”, implying that finding boundaries in nature should be as easy as carving a chicken. In practice, there is often not an obvious way to arrive at a consensual and well-defined division of a domain, and finding one can be a messy process in which social and political forces are brought to bear along with the supposed natural order of the world.

1.1 The defense of “marriage”

Sometimes the politics behind ontological conflict is obvious and explicit. For instance, the current political controversy of what constitutes the proper boundaries of the concept of “marriage” weaves together the formal semantics of the term with attendant legal rights, social status, and material benefits. Marriages already have some formal properties that translate directly into computational terms, such as symmetry (if a is married to b , then b is married to a) and exclusivity (assuming that bigamy is just as illegal in the database as it is in real life). Gay marriage complicates things a bit more, and has a direct impact on databases and code:

So, after removing the “husband and wife” limitation, you would actually have to add in a check constraint or some new application logic to ensure that people didn’t marry themselves. It would almost never be called upon but it would have to be in there, somewhere. This minor programming challenge is actually our largest obstacle.[1]

This is largely tounge-in-cheek, but it conveys an important point – that our social institutions are intertwined with our bureaucratic and computational formalisms. A society without bureaucracies, central government, and their attendant databases doesn’t have a real problem with defining marriage – marriage is whatever individuals make of it. A society with databases, however, has to fit the messy lives of individuals into a formal structure. Computers are, from this perspective, just the latest technology of bureaucracy, in keeping with Max Weber’s dictum that “Bureaucratic administration means fundamentally domination through knowledge” [78].

Once marriage is formalized by government, bureaucracy, and computers, its properties become those which are encodable in the formalism. But

real lives tend to generate exceptions to such fixed structures, e.g., when individuals change their gender².

Because of the federal system in the US, different states have different and often incommensurable definitions of marriage. To deal with this, there are meta-level laws such as The Full Faith and Credit clause of the US Constitution (which requires states to recognize each other’s “public acts and records”) and the Defense of Marriage Act (which explicitly contradicts this clause). Ultimately this ontological issue will probably be decided by the Supreme Court. It would be hard to find a clearer example of the intersection of ontology with politics and authority.

As computer scientists, our schemas and representations both reflect and help construct social reality. What if we had databases that, instead of based on rigidly-typed tables and categories, were based on more flexible technologies of representation? Since representation and society have complex and interdependent interactions, the adoption of such technology might go some ways to making society itself more tolerant.

1.2 What constitutes a disease?

Other areas in which ontology intersects with political and social controversy include the question of what constitutes a person [44] (a pivotal issue for the abortion debate, and of earlier conflicts over slavery), and the definitions of psychiatric diseases [61] (which impact eligibility for various government benefits as well as moral culpability for crimes).

For example, post-traumatic stress disorder (PTSD) is a relatively recent diagnostic category, whose adoption involved a political struggle focused on its implications for the Vietnam war and military service in general [63]:

...not to suggest that this diagnosis [is]...”merely” a social construction, or simply the result of self-interest... the story of PTSD... help[s] us understand in detail how objective knowledge – and medical scientific knowledge in particular – is produced, secured, and subsequently used... Each new clinical diagnosis of PTSD, each new warrantable medical insurance claim, each new narrative about the disorder reaffirms its reality, its objectivity, its “just thereness”. In the story of PTSD we see again how the orderliness of the natural world is to be found in its very accounts of orderliness.

And until quite recently, homosexuality was considered a psychiatric diagnosis, but now it isn’t [5], a reclassification born of a political struggle. It may

²<http://lawblog.legalmatch.com/2011/04/29/texas-trying-to-retake-title-of-most-regressive-state-from-arizona/>

be thought that the cases cited so far are exceptional, since they all involve classifications of humans and their social relationships, and are thus naturally subject to controversy and politics. But I prefer to think of these as just the more obvious cases of a more general phenomenon, and that subtler forms of politics may be found in all representational efforts.

1.3 Contested classifications in biology

Even areas that are not so obviously fraught with political implications will often involve conflicts of interest or simply divergent but incommensurable viewpoints. Scientific knowledge representation encounters this problem at many levels. The naive view of science is that it is above politics, and that it, perhaps alone of human activities, should have ontologies that are direct representations of an objective reality. Indeed, objective and disinterested representations are the *aspiration* of science, but the process of achieving them necessarily involves representations that are localized, situated and *interested* in the sense that they are created to favor a particular point of view that is contending for acceptance. Thus designing representations even for things that ought to be value-neutral, such as the component parts of a cellular metabolic process, are often marked with human bias and interests, and thus are not objective in any interesting sense.

1.3.1 Example: mitochondria

Mitochondria are parts of cells (organelles) but have their evolutionary origins as independent organisms that were assimilated symbiotically into ancestral cells. They maintain some of the characteristics of independence to this day, such as their own genomes (mtDNA). A recent article [54] mused that we really should consider mitochondria as separate organisms still, with their own taxonomic codes, evolutionary history, and the other conceptual apparatus that comes with that status:

A wide range of molecular phylogenetic studies has been applied to macromolecules encoded within mitochondrial genomes and all support a bacterial... affiliation for mitochondria. Thus, by the logic of cladistic classification, we have no choice but to accept that mitochondria belong in the domain Bacteria, or to put it plainly, that mitochondria are bacteria.

This is a somewhat atypical example of a biologist consciously toying with ontological issues and their consequences. Are mitochondria separate

organisms or parts of another organism? In fact they can easily be seen as both, and this will tend to break ontologies that have a too-rigid distinction between organisms and parts. But this is not merely a terminological question; viewing mitochondria as organisms in their own right puts them front and center for certain types of scientific attention, e.g, it suggests that they can be independent objects of metabolic re-engineering. Pallen also raises issues of *credit*, always an important issue in the process of science:

If we accept that mitochondria are bacteria, then the record books have to be rewritten. The first bacterial genome sequence was completed not by American *arriviste* Craig Venter and his team in 1995, but instead by a team at the Medical Research Council Laboratories in Cambridge, England, which included double Nobel laureate Fred Sanger, who completed the human mitochondrial genome sequence in 1981!

Pallen mocks his own proposal as “phylogenetic fundamentalism” (that is, his viewpoint involves taking a strict view of biological classification as based solely on the descent relationships of organisms, as opposed to a more functional classification). For our purposes, it suffices to note that this controversy is not really about the structure of the world, since all parties acknowledge the bacterial ancestry of mitochondria. The issue is one of classification, naming, and relationship; and the consequences of such classification, which include intellectual credit, qualifications for funding, different frames of reference being applied to problems, or people from different subfields being drawn in or pushed out of working on particular problems.

1.3.2 Example: genes and pathways

The definitions of even such basic biological entities such as “gene” or “pathway” pose difficult issues: the nature of their boundaries; the nomenclature by which they are represented in databases and discourse; and their stability, unity, and even their reality. Genes by their nature have variant forms within and across species, and even within a single individual (in cancer, e.g., where particular groups of cells become mutated), making it difficult to establish a firm definitional boundary for what constitutes an individual gene. A gene may be an abstract ideal entity, a standard sequence of codons in a reference database, a variant of that standard, or a particular individualized piece of physical DNA in an individual cell. The elastic semantics of “gene” frustrates formalists, but is essential to the everyday discourse of science

A metabolic pathway is a group of functionally related biochemical reactions within a cell. Pathways have been a basic concept of biology from well

before the genomic era [47, 62], but it suffers from some of the same fuzzy definitional problems as genes (if indeed they are problems), and in addition from having no obvious well-defined boundaries in the world. Unlike genes or cells, which do have natural boundaries, pathways have no uncontroversial starting and stopping points, but are really just sections of a global reaction graph that have been scissored out of the whole by the informed judgment of scientists.

For instance, the BioPAX standard [21] is an attempt by a coalition of scientists to define a common standard for representing pathways (emphasis added):

Pathway: Definition: A set or series of interactions, often forming a network, which **biologists have found useful to group together** for organizational, historic, biophysical or other reasons.

The problem with this, of course, is that the utility of a grouping depends upon the context for analysis, which can easily be different for different scientists or projects [65]. Pathways thus have a certain fictitious or constructed quality to them, in a more obvious way than some other scientific objects.

Pathway Tools [36] is a long-standing resource for genomic biology, a collection of pathway-genome databases (PGDBs) for over 1000 organisms, and tools to view, edit, and interact with them. Pathway Tools, due in part to its ontology needs to define pathways as relatively crisp, permanent objects, uses the following definition of pathway:³

Pathway boundaries are defined heuristically, using the judgment of expert curators. Curators consider the following aspects of a pathway when defining its boundaries.

- What boundaries were defined historically for pathway?
- When possible, we prefer to define boundaries at the 13 common currency metabolites: D-glucose-6-phosphate pyruvate etc.
- Coincidence with regulatory units
- Coincidence with metabolic units that are evolutionarily conserved

The preceding philosophy toward pathway boundary definition contrasts sharply with KEGG maps. KEGG maps are on average 4.2 times larger than BioCyc pathways because KEGG tends to group into a single map multiple biological pathways that converge on a single metabolite.

³<http://metacyc.org/biocyc-guide.shtml>

Note here that (1) there is a specific and articulated philosophy of pathway definition, (2) that it depends on the informed judgment of experts, and (3) it is contrasted with the methodology of a similar rival knowledge base, KEGG [35]. More fundamentally, pathway models necessarily encode a fixed structure of function and explanation, which necessarily excludes pathology (as in the case of pathways disturbed by genetic disease or cancerous mutation) or novelty (in the case of synthetic pathways).

Another interesting issue arises from the fact that ontologies and the software that uses them have *histories*, and the changing needs of science over the years will cause tensions between established representational usage and new needs. In the case of Pathway Tools, the information has been organized around separate sets of frames for each organism, so there is a separate knowledge base for each (EcoCyc for *E. coli* [37], HumanCyc for *Homo sapiens* [57]). The database collection has recently grown to over 1000 separate organisms/knowledge bases.

But now there are new proposed applications for the software that involve things that cross organism boundaries: metabolic engineering, which involves synthesizing new pathways from fragments of multiple microorganisms (for applications like synthesizing fuels and drugs) [33, 14]; the study of metabolic pathways that involve the interaction of a host and symbiote [17]; and metagenomics, which involves sequencing the DNA of large samples without prior knowledge of what organisms they come from [64]. All these applications are possible within the structure of the software, but they put some strain on the existing structures and ontology.

2 Alignment, social construction, power

Human societies are capable of accomplishing joint ventures of vast complexity despite the conflicting individual goals of the participants. The construction of representations and ontologies is no different; it is a social process and thus subject to social forces like conflict, power, negotiation, signaling, etc. Conflicts over ontologies can take many forms, including disagreements over naming (terminological), over classification (taxonomic), over boundaries of concepts, or any of the other features of representation. In real-world representation construction, these conflicts must be negotiated and the interest groups behind them brought into at least temporary alignment.

2.1 The sociology of ontology construction

My own experience working on a variety of complex software projects has sparked a fascination with the social processes that underlie the development of ontologies. Building an ontology is a complex social-coordination problem, but ontologies have the added property that their output is something that at least aspires to be a kind of objective, interest-free model of some domain. If the structure of ontologies were inevitably determined by the structure of the phenomenon they aim to describe, then anybody could be an ontologist, object models would tend to converge rather than diverge, and the problem of representation would be much simpler than it actually is. In practice, of course it isn't easy at all.

Generally representational projects will have one person occupying a role I like to call the "alpha ontologist" who is the ultimate determiner of the right way to represent something. There are a number of problems with this type of methodology, starting with its dictatorial nature. Perhaps that's not a real problem, given that some constructed things seem better for having a strong leader (consider Steve Jobs' role at Apple, e.g.). But an ontology is supposed to be universal, and having a universal knowledge schema be the product of a single mind doesn't seem right. More practically, a centralized process like this can't scale up. For knowledge representation to succeed at a mass scale, it needs to become more accessible and more democratic.

2.2 Example: representing cellular space

This section presents a discussion among the developers of Pathway Tools about a relatively small ontological issue, and is included here as an illustration of how social forces and negotiation can appear in non-obvious places and around the smallest concerns.

Underlying Pathway Tools is a frame-based knowledge representation system called Ocelot [13], with a schema that has evolved over fifteen years to capture the important parts of genomic biology as they evolved. The discussion here is part of a longer discussion on how to represent the localization of biological reactions, including transport reactions (which move molecules from one location to another). The issue is how to solve a problem that has crept into the knowledge base, a confusion about the representation of cellular location (important because some reactions are limited to certain parts of the cell). The issue revolves around the difference between the definitions of *cytoplasm* (the liquid inside the membrane of a cell, excluding the nucleus, but including organelles), and the *cytosol*, which is an almost

identical concept except that it excludes the organelles.

The participants include software developers (nl and tq) and biologists (sd and pz). [All emphasis added]

nl: An accurate representation of processes inside of a cell needs to take into account... different compartments.. and that the processes and metabolites are partitioned between these compartments... In BioCyc, compartments need to be valid frames in the CCO (Cell Compartment Ontology), which can either be children of CCO-SPACE or CCO-MEMBRANE. The default compartment for metabolites and reactions is assumed to be CCO-CYTOPLASM.

nl: the important function (default-compartment) has been **historically hard-wired** to return CCO-CYTOSOL, but I don't understand why...as we use other locations called periplasm, using cytoplasm would look much more **consistent**. I thus would prefer to use CCO-CYTOPLASM instead.

I don't even see why there should be much of a distinction between CCO-CYTOPLASM and CCO-CYTOSOL anyways. to me, it looks like CCO-CYTOSOL is an unnecessary and **not useful** concept...after all, we also don't have a perisol either. We only have a periplasm.

sd: While cytosol and cytoplasm are different terms and describe different entities, I think that probably most biologists will be hard pressed to point out the differences without looking it up. So while we **lose some resolution** if we eliminate cytosol, we probably don't lose anything **particularly useful**, and if this serves to eliminate some problems, I vote for the merge.

nl: I think there is confusion on both counts. What happened in some PGDBs is that for some reactions, sometimes CCO-CYTOPLASM was used instead of the default compartment which was CCO-CYTOSOL. This will **cause needless disconnects** in the reaction network.

pz: By definition, cytoplasm and cytosol **are very different things**. But I agree that people may not distinguish them when using the terms. I am asking opinion from a group at Carnegie who conducts large-scale protein localization experiments.

tq: I'm not a biologist, so I have not so far expressed an opinion, but that doesn't mean I don't have one. :-) IMO, if you are going to eliminate either cytoplasm or cytosol, the one to keep is cytosol – that is almost certainly what is intended in nearly all transport reactions... However, I would strongly prefer to keep both, and to **eliminate data inconsistencies** by only allowing reactions to occur between compounds in spaces... Cytoplasm is not classified as a space, it is a super-component, so it should never appear in reactions,

and the user should not be given the choice to add it. That way we could keep cytoplasm as a concept (with its correct definition) but **the user would never see it**.

nl: My proposal was that cytoplasm should become the space that will replace the current cytosol. In other words, I don't want the name "cytosol" to show in transport rxn diagrams, but "cytoplasm", for **consistency**.

tq: You could accomplish that by just **changing the common name** (or whatever name used for reaction display) of CCO-CYTOSOL to "cytoplasm". ... As to whether or not the name change would be misleading or confusing, I'll leave that up to the real biologists. It's possible there could be one answer for bacteria and another answer for eukaryotes, in which case you'd have to decide whether it was worth special case-ing.

The formalist within me recoils in horror at this tangled mess of concerns – but another part of me thinks, this is great! It reflects how knowledge representation is done in real-life practice, how human minds come at it through user interfaces, how history and practical considerations blend, how people collectively decide to remold part of one layer of the system to make up for deficiencies in another.

Look at some of the diverse concerns that surface in this discussion: defaults, consistency, history, utility, user confusion, resolution, voting, semantics of part-of vs. surrounded-by, impact on computational processes, user interface, naming vs denotation, status signaling, deference, and the use of emoticons to deflect possible insult. The solution proposed in the example involves papering over a conceptual fissure by using a different name in the user interface than the internal name. The technical, the social, and domain knowledge are all mixed together in a fluid way. There is constant effort required to align the interests of the users and the "interests" of formality and inertia, and the builders will use whatever tools are at their disposal – interface tricks, clever defaults, deployment of synonyms or pseudonyms, etc. This sort of thing goes on *all the time* in the course of real-world software development, and there aren't very good tools for describing it, let alone managing it.

3 Conceptualism vs. realism

The biomedical computing community has grappled with problems of representing its exceedingly complex domain for many years. Recently, a philosophical debate about the nature of ontology has erupted [23, 68, 34, 48, 45]

between two schools: “conceptualism”, which holds that an ontology is “a formal specification of a conceptualization” [29], versus “realism”, a faction led by the philosopher Barry Smith and colleagues at SUNY Buffalo, who insists that terms in an ontology denote “real” entities in the physical world.

I put “real” in quotes because this stance seems to beg the entire question, but there are certainly some strong arguments for realism, which include: the notion of a “concept” is ill-defined; realism is a necessary stance to avoid being drawn into the excesses of idealism and radical constructivism; and that existing conceptual ontologies, because they lack a well-defined relationship with reality, are riddled with incoherence. Conceptual ontologies might be good enough for human constructs, but are inadequate for empirical science:

The influence of the concept-centered view ... has become entrenched also in virtue of the fact that much work on ontology has been concerned with representations of domains, such as commerce, law, or public administration, where we are dealing with the products of human convention and agreement – and thus with entities which are in some sense merely ‘conceptual’. Today, however, we are facing a situation where ontologies are increasingly being developed in close co-operation with those working at the interface between the informatics disciplines and the empirical sciences, and under these conditions the concept-centered view is exerting a damaging influence on the progress of ontology. [67]

This is a very clear statement of the metaphysical basis of Smith’s approach to ontology – the physical (or empirical) is real, while the social consists of “mere” concepts. This hard boundary between the real and the social is exactly what Latour’s approach is attempting to demolish (see section 4.2). However, in the knowledge representation world, the realists seem to be winning in their battle with conceptualists. A realist ontology, the BFO (Basic Formal Ontology) [55] has been adopted by leading biomedical ontology efforts such as OBO [70]. Possibly realism is winning because nobody in the sciences wants to be seen as being against reality:

Realism has chosen wisely in its choice of name. Most scientists believe in reality so, when faced with realism vs conceptualism, their gut feeling is that the former will be right. They believe in a mind-independent reality so, therefore, conceptualism must be wrong. [46]

This section explores a variety of critiques of the realist position, which I believe to be exerting a pernicious influence on the practice of representation. For a correspondingly impassioned critique of the conceptualist position, see

[67]. The following section on pragmatism attempts to sketch out a view of representation that is different from both of these approaches and aspires to transcend some of these controversies.

3.1 The poverty of realism

Realism runs into problems even when dealing with quite elementary use cases for scientific knowledge representation. For instance, realists insist that any concept in an ontology has to have a physical instantiation (which seems to be the operational definition of “real”). But consider the case of software to support pharmaceutical drug design⁴. Synthetic chemists, in pharmaceutical research and elsewhere, spend their professional lives imagining molecules that do not exist yet, then figuring out ways to make them exist. Today this is often done with the aid of software that can generate new molecular structures from libraries of parts, along with plans for synthesising them. Are such compounds real-world objects? Are they unreal before synthesis, but real after? These are not questions that are germane to a chemist, or to a designer of chemical software. Chemists have no ontological or epistemological distinction between a molecular structure before and after synthesis. Yet realism insists they are different and outlaws talking about the yet-to-exist ones, just as it outlaws talking about unicorns (which you might also want to represent, despite their unreality) and Higgs bosons (which could well be real but have not yet been demonstrated to be so):

Descriptions and plans do, after all, exist... [but] it would be an error to include in a scientific ontology of drugs terms referring to pharmaceutical products which do not yet (and may never) exist, solely on the basis of plans and descriptions. Rather, such terms should be included precisely at the point where the corresponding instances do indeed exist in reality... [69]

Genes too were hypothesized abstract entities before their physical basis in DNA was discovered by Watson and Crick, and thus would be problematic to represent if people were doing formal knowledge representation back then – and we can assume that there will be similar problems today and in the future. In practice, advocates of the realist approach may have work-arounds for these problems (such as considering an informational entity, such as a chemist’s sketch of a molecule, to be “real”), but the point is that this is awkward, unnatural, and a distraction from more important issues. In

⁴Here I draw from experience as one of the developers of Afferent, a commercial software system to support combinatorial synthetic chemistry.

practice, it is a barrier to the adoption of ontologies in practical scientific work.

3.2 Complexity and barrier to entry

Another common critique of realism is that by forcing representation into a rigid and inexpressive framework, the result is useless and forbidding complexity. I think this is one reason that semantic-web style knowledge representation has not yet had the success and mass uptake that drove the web to its current universal deployment. The barriers to entry to formal representation are forbiddingly high, and realism makes them worse. For instance, to solve problems like the above (representing an abstract informational entity), the BFO apparently has been patched to handle such cases with an abstraction called “generically dependent continuants” [52]. Here, for example, is the definition of the relationship **is-concretization-of**, which links informational entities with their realizations:

is-concretization-of is a relationship between a generically dependent continuant and a specifically dependent continuant. A generically dependent continuant may inhere in more than one entity. It does so by virtue of the fact that there is, for each entity that it inheres, a specifically dependent **concretization** of the generically dependent continuant that is specifically dependent. For instance, consider a story, which is an information artifact that inheres in some number of books. Each book bears some quality that carries the story. The relation between this quality and the generically dependent continuant is that the former is the concretization of the latter.⁵

Whether or not this is a good representation is hard to say, but the point is, very few lay persons not conversant with intricate formal ontology can understand it at all. In some ways realism is not a natural fit to human cognition (for alternatives that might alleviate this problem, see section 5.2 on prototype-based representation). This may have nothing to do with an ontology’s *accuracy*, but will affect its *utility*. There is some question here about what an ontology is for; if it is purely for machines to reason with; or also a tool for humans to organize their knowledge. In practice, humans must still be involved at least in the construction of ontologies, if not their deployment and use, so fitting ontologies to the way humans actually think is something of a necessity.

⁵<http://berkeleybop.org/obo/OBI:0000294>

3.3 Mental illness and realism

A fine illustration of the poverty of the realist approach may be found in a paper from that school that applies their techniques to the problem of building an ontology for mental disease [12]. The amazing thing about this paper is that, aside from an offhand quotation of Szasz’s theory that mental illness does not exist, it makes no mention whatsoever of the controversies about the reality of mental disease, the evolution of the concept over time, the political and social forces involved in its definition, or the constant revision of the standard references (DSM-III and its successors) to reflect evolving social standards. In short, *none of the interesting issues* involved in the ontology of mental illness are dealt with. Instead, they are reduced to a single link in a graph (see Figure 1).

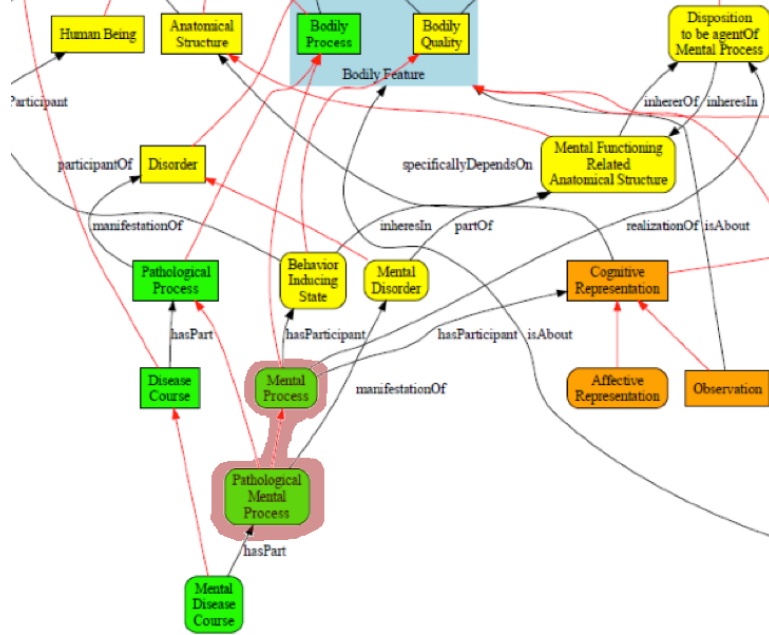


Figure 1: Ontology of mental illness from [12], highlight added

It may be that I’m missing something in realism or otherwise being unfair to it. In *Against Fantology* [56] and in a reply to Merrill [68], Smith sets himself against a certain form of logicist representation, implying a willingness and ability to deal with the world’s messiness:

... we ourselves are interested precisely in really existing scientific theories, and in the associated really existing ontologies, which in nor-

mal circumstances are not associated with any claim to completeness. Really existing scientific theories are marked, rather, by messy and inconvenient processes of change and of correction of error, including ontological error, and our formulation of the realist methodology is designed precisely to do justice to this fact.

On the other hand, he doesn't believe conflict is important:

Where conflicts do arise in the course of scientific development, these are highly localized, and pertain to specific mechanisms, for example of drug action or disease development, which can serve as the targets of conflicting beliefs only because researchers share a huge body of presuppositions.

As best as I can interpret Smith's idea of realism, it is more of an ideal that science and scientific representation should aspire to, rather than a model or tool for science as practiced (despite the first quote above). In that sense, I find no problem with it. But it ignores the need for transient, situated, tentative representations that are the basis of everyday scientific practice, and computing practice as well. As Philip Lord put it, "the choice is between representing reality or representing how we practice science" [45].

4 Representational pragmatism

The debate between conceptualism and realism threatens to become as contentious as the so-called "science wars" that took place in the larger scientific and intellectual community in the 1990s [28, 30], and it revolves around some of the same philosophical dichotomies and exaggerations of opposing positions. While I am pretty firmly on the conceptualist side, I find the whole debate somewhat misguided. In general when faced with a hopeless philosophical quarrel, a good tactic is to try to transcend it or find a dimension of the issue that is orthogonal to the debate.

This section explores a set of ideas from philosophy and sociology of science in an attempt to sketch out a view of representation that avoids the pitfalls of both realism and conceptualism.

4.1 Philosophical Pragmatism

Just as the realists appropriated the term "realism", I would like to do the same with "pragmatism", which conflates a philosophical school (exemplified by William James, John Dewey, and Richard Rorty) with a common desirable attitude. Neither biologists nor computer scientists want to get mired

in fruitless philosophical debates; they want to produce insightful models of the world and build tools that get used. But sometimes descending into philosophy is necessary if bad philosophical assumptions are undermining your efforts.

Philosophical pragmatism was introduced by William James as a corrective to a debate of his own day between materialism and idealism; one that roughly parallels the science wars of today. The essence of pragmatism is to cease looking for the essences of things, but rather to judge them by their consequences: how they are used and how they affect other things. As James put it, “It is astonishing how many philosophical disputes collapse into insignificance the moment that you subject them to this simple test of tracing a concrete consequence.” He further writes:

Pragmatism represents a perfectly familiar attitude in philosophy, the empiricist attitude, but... in a more radical and in a less objectionable form than it has ever yet assumed. A pragmatist ... turns away from abstraction and insufficiency...from fixed principles, closed systems, and pretended absolutes and origins. He turns toward concreteness and adequacy, and towards power...It means the open air and the possibilities of nature, as against dogma, artificiality, and the pretense of finality in truth.

...pragmatism [is] a mediator and reconciler...that ‘unstiffens’ our theories. ...pragmatism is willing to take anything, to follow either logic or the senses and to count the humblest and most personal experiences.

– William James, [32]

Richard Rorty extended pragmatism into the late 20th century and emphasized its antifoundational character [58], and Paul Feyerabend stretched it even further (perhaps too far) with his notion of “epistemological anarchism” [25], an attitude that all forms of representation and inquiry should be allowable. It is just this tendency of the radical pragmatists to take things too far that inspired the realist reaction.

The motivation for the hard-line realist position of Smith is that anything less than realism risks being pulled into a fuzzy world of uncertainty, conflict, and unanchored reference. Feyerabend, for his part, believed he was defending the practices of scientists from philosophers who wished to impose artificial restrictions on their activity. Is there a happy medium, a path that threads between the dangers of rigidity on one hand and losing the mooring of reality on the other? I maintain that looking at the actual, situated, bottom-up representations of science as actually practiced provides a guide.

4.2 Latour: circulating reference

Bruno Latour is a sociologist of science whose work has generated a good deal of controversy [72]. While he is often taken as an exemplar of the much-maligned “social construction of science” and thus an enemy of realism and even of science itself, he vehemently denies this (recently in *Pandora’s Hope* [?], Ch 1, for instance, titled *Do you Believe in Reality?*):

If science studies has achieved anything, I thought, surely it has *added* reality to science, not withdrawn any from it...Who loves the sciences...more than this tiny scientific tribe that has learned to open up facts, machines, and theories with all their roots, blood vessels, networks, rhizomes, and tendrils? Who believes more in the objectivity of science than those who believe it can be turned into an object of inquiry?

Here, we can ignore the metaphysical implications of Latour’s viewpoint and concentrate on his picture of science as a process of mobilizing representations. A scientist will examine an object in the field (e.g., in Ch 2 of *Pandora’s Hope*, the structure of the border of a forest in Brazil), turn that phenomenon into inscriptions in a field notebook, bring those inscriptions back into a laboratory where they will be further condensed, abstracted, computed with, mixed with other representations that strengthen the scientists position. While Latour does not deny truth any more than he denies reality, he focuses not so much on the truth or veracity of representations but on the ways they circulate between scientists, instruments, laboratories, and publications, and the ways they enable a scientist in a laboratory to make statements that are true of the larger world. Science in this view is not only a social process, but one that unites humans, objects, and representations into seamless networks:

The same article [on global warming and CFC emissions] mixes together chemical reactions and political reactions. A single thread links the most esoteric sciences and the most sordid politics, the most distant sky and some factory in the Lyon suburbs, dangers on a global scale and the impending local elections or the next board meeting. The horizons, the stakes, the time frames, the actors – none of these is commensurable, yet there they are, caught up in the same story.

– We Have Never Been Modern [42], p1.

Latour views his position as an *enriched* realism, one that considers scientists, representations, social institutions, and politics just as real as rocks and atoms, with the implication that they need to be analyzed together.

It seems to me that this should be a very *congenial* attitude for computationalists; a much better model of both reality and representation than that offered by the Smith school of realism. The essence of computer science is building representations that are at the same time real things, not only encoded in bits in a particular physical location in memory, but also potentially dynamic entities (actants in Latour’s vocabulary), participants in an ecosystem made of software, humans, social relationships and institutions. Furthermore, we do not build representations alone but as part of systems for creating, encoding, and transmitting them. Computational representations are not mere mirrors of nature but are themselves parts of networks, of causal systems that connect humans, computers, and the domains of study, and need to be designed as such.

4.3 Representation from the bottom up

Representational pragmatism is partly motivated by observations of the actual work processes of scientists in the laboratory, who regularly invent representational schemes that are deeply situated in the phenomena they are tracking. The job of a scientist, as described by sociological observers of science, is to turn phenomena into inscriptions (“the transformation of rats and chemicals into paper” – Latour, [41]). The inscriptions of science start out as radically concrete, situated things, in somebody’s notebook denoting concrete material entities and observations. But somehow scientists can lever themselves and their representations out of their situated perspective into something closer to *objectivity*, where they can issue abstract truths about the world that are deserving of publication. The sheer variety, inventiveness, and richness of the ways in which this happens is itself a rewarding object of study. Jeff Shrager described some of the realities of lab-notebook based representation, gleaned from his experience training to be a biologist [66]:

But if you lose your lab notebook, you’re hosed, mainly because you’ll never figure out what the hell is in the hundreds of obscurely-labeled tubes in the various freezers in the various boxes with the various obscure markings on them. I haven’t come upon a perfect scheme for organizing all this yet. The protocols I just put in date order and then I have an index in the front that tells me what date to look at for what protocols...These are labels that I’ll put on tubes that have reached that stage, so that I know where they are in their chemical careers.

This passage is interesting because it illustrates one way in which the initial representations of science are connected (in this case, a direct physical

connection) to the phenomenon they describe. Scientists will invent labeling schemes that they translate into spreadsheets, then into tables in articles. The reality of scientific representation is just as Latour describes, a bottom-up matter of turning reality into inscriptions that are progressively refined. If logically formal representation is here, it comes at the endpoints of the process. In order for this process to happen, scientists require the sort of representational flexibility on display here.

4.4 Convergence and objectivity

The goal of scientific representation is *convergence* onto some formal structure that accurately captures and reflects an underlying reality. Science under construction must pursue divergent and rivalrous paths, but is guided by the faith (which so far seems well-placed) that the paths will ultimately converge on an a common understanding. Unlike the more radical epistemologists of science like Feyerabend and Rorty, I believe that the structure of reality does in fact generate this kind of convergence – at least some of the time. Some areas of study are more strongly convergent than others – for instance, it seems likely that mathematical truths are the same for all cultures. However, even in mathematics there are different approaches to truth [18], including logics that deemphasize standard notions of truth by permitting inconsistency [31]. The further knowledge gets from mathematics, the more it will be subject to divergent approaches. Whether science as a whole can reach convergence is a controversial question in the philosophy of science, with leading figures (Kuhn, Feyerabend) maintaining that science inevitably develops incommensurable (nonconvergent) theories.

The difference between realism and pragmatism as approaches to ontology is that realism thinks that convergence can happen by decree, while pragmatism thinks that it happens more as an asymptotic endpoint of a bottom-up social process, and that the semantic structures of that imagined end is not knowable in advance. The beautiful book *Objectivity* [19] demonstrates that even the concept of objectivity as an “epistemic virtue” has a history, and the modern idea of objectivity as a mechanical reproduction of nature was a product of a series of revolutions in scientific visualization. To the extent representations are objective, that is an *achievement*, and this applies equally to individuals, scientists, and modelers. Science is a somewhat magical tool for allowing us to escape from our lowly subjective viewpoints and getting a view of the world from above it as objectivity demands. But it’s a process, and our tools need to support the process as it happens, not merely its endpoints.

4.5 Summary: What is Representational Pragmatism?

Representational pragmatism is an effort to get beyond both the conceptualist and realist factions by positing a different philosophy of what representation is and how it is used⁶. It is not (yet) a concrete technical architecture, although there are many technical systems that in various ways are suggestive (see the next section). Rather than imposing an ontology or even a framework of representation, it takes a computational “anything goes” approach. It encourages lightweight standards, and tools that make it easy to integrate divergent representations.

Representational pragmatism is based on what people actually do with representation, in the everyday world and in the real practices of science. It encompasses actual representations found in lab notebooks, articles, whiteboards and everyday discourse as well as the formal structures of ontologists. It is to be found in the processes of formalization rather than in the final formalisms themselves. The role of inconsistency and conflict in this process should be made a strength rather than a weakness. These factors are an inescapable fact of actual representation use, so computational tools should be designed to serve those actual uses, rather than attempt to dictate particular forms of thought. An overly rigid framework for ontology is not robust to the actual structures of knowledge that people need to work with. The next section explores some possible strategies for achieving robustness and flexibility.

5 Computational alternatives

But logic also has its disadvantages... No exceptions are allowed, no matter how closely they match. This approach permits you to use no near misses, no suggestive clues, no compromises, no analogies, and no metaphors. To shackle yourself so inflexibly is to shoot your own mind in the foot—if you know what I mean.

– Marvin Minsky [50]

Knowledge representation should be a matter of making representations that capture and augment human knowledge and reasoning, rather than attempting to force it into narrow and unproductive channels. Convergence may be the end goal, but the journey towards that goal must take into account the divergent representations used in daily life and the ordinary practices of science. This means that representational systems must be fluid, sit-

⁶for an earlier and inspirational effort to unpack the nature of representation, see [2]

uated, flexible, robust, and social. A variety of techniques suggest themselves to achieve these goals. These include alternative forms of representation such as prototype-based formalisms that eliminate class/instance distinctions inherited from formal logic and replace them with structures more attuned to actual human thought. Another class of techniques are those borrowed from the web, open source and agile software communities: versioning, forking, branching, wikis, and explicit technological and institutional support for use cases, unit testing, and refactoring.

5.1 Is the semantic web robust?

The Semantic Web has ambitions of being a universal framework for knowledge representation; to be as successful as a *lingua franca* in the formal representation domain as the WWW standards have been for documents and human communication. It has had at best limited success to date. In part this is due to the objectivist and logicist influences that have resulted in methodologies that are difficult to use, with extremely steep learning curves. This is in stark contrast with the web, where readily understandable and visible standards have resulted in a vast proliferation of content and applications.

One thing that the semantic web community has done right is to design it as a layered architecture, with separable standards for assigning names to entities (using URIs and XML namespaces), for encoding basic relations (RDF), and for more formalized ontologies (OWL). So even if OWL suffers from the rigidities of formalization and realist methodologies, it is still possible to utilize the other layers of the standards to achieve representational interoperability. This seems to be happening; the most recent incarnation of the semantic web has a different name (Linked Data⁷) and more emphasis on lightweight representation. Unfortunately RDF itself is somewhat limited in that it makes it difficult to represent relations between more than two objects.

The semantic web is a much more distributed effort than a typical knowledge representation project, but there too we have seen the emergence of a cadre of alpha ontologists who rove around to various user communities helping them get their concepts in order. There isn't anything wrong with this, but again, it doesn't seem to scale, and may be part of the reason the semantic web has not quite taken off the way successful web efforts do. And it seems also that the semantic web effort has been captured by the realists,

⁷<http://linkeddata.org/>

with a resultant rigidity and high barrier to entry. This is extremely unfortunate, because the world needs some way to interconnect all its knowledge. A universal standard for sharing formalized knowledge would be a great asset to the world, but only if it's used.

A full analysis of what works and what doesn't in the semantic web world is beyond the scope of this paper. My opinion is that the world very much needs a universal framework for knowledge but the semantic web has so far failed to supply that because of some unfortunate commitments, but may yet find its way. I believe the ideas outlined here could serve as an alternative pathway forward.

5.2 Prototype-based representation

Representation based on prototypes rather than classes is an old but currently out-of-favor idea in computation [73, 43, 7, 77, 71]. It was implicit in some early AI work in knowledge representation [49, 8], but has faded in popularity in favor of class-based schemes which are easier to formalize and thus more amenable to mathematical treatment. Prototypes have several advantages over class-based representation, including:

- Prototypes are a more natural fit with human cognition, according to the branch of cognitive science exemplified by Eleanor Rosch [59] and George Lakoff [40];
- Prototype-based systems permit the representation of generalizations, defaults, and exceptions (non-monotonic reasoning), e.g., being able to say that birds fly while permitting exceptions for penguins and ostriches, or that dogs have four legs while permitting an exception for amputees; or that cats are mammals and mammals have hair, except those that don't (see Figure 2);
- Prototype-based systems suggest a different style of inference than strict logical deduction, such as reasoning by analogy or by application of concrete cases.

Prototype-based methodologies may be particularly useful for biological representation, since biology is rich with exceptions to general rules. Every biological feature started out as an exception (aka mutation) which then became fixated as a new norm. I suggest that our representations should be capable of evolving in a similar fashion. To return to the example we started with: under a prototype-based representational scheme, a marriage would be



Figure 2: Ripley the exceptional hairless cat

between a man and a woman, except when it wasn't. This might not make activists completely happy, but it would more accurately match the mind's cognitive structures. More importantly, such a representation is robust to the variety of the natural world. Classical logic can't handle contradictions, but natural cognition handles them all the time. It should be noted that nonmonotonic reasoning systems and other logics that admit inconsistency [31] can also handle exceptions while retaining a class-based structure.

5.3 Discourse, argumentation, provenance

Efforts to explicitly represent the discourse and argumentation structure of knowledge are one constructive path towards a more pragmatic model of representation. Some examples of this approach include the SWAN ontology for scientific discourse [15], the ScholOnto project [10] for representing scholarly discourse and interpretation, and a number of projects for representing and visualizing argument structure [38, 60, 27, 51]. Perhaps closest in spirit to this paper is the DILIGENT system [74] which explicitly supports a role for argumentation in ontology development.

5.4 Wikipedia and socially-constructed knowledge bases

Wikipedia is a phenomenally successful collaborative knowledge building effort. Obviously the distributed Wikipedia curation community is not defining a *formalized* knowledge representation, but it is, nonetheless forming informal ontologies, with semi-structured data being the norm (there are a variety of variants on Wikipedia that are closer to structured knowledge bases, including DBpedia [3] and Freebase⁸).

One of the more interesting features of Wikipedia is its social structure and meta-level ontology of problems, bugs, and conflicts about how to write articles. Every article has a corresponding discussion page for these issues to be hashed out. There is also a semi-formalized ontology of issues, each of which translates into a cleanup template⁹ or banner on articles. Typical examples warn of copyright violation, bias, problems with style, or consist of boilerplate notations such as “This article contains weasel words, vague phrasing that often accompanies biased or unverifiable information. Such statements should be clarified or removed.” The result is that Wikipedia is an ongoing laboratory for policy and ontology negotiation [6, 11]. The informal nature of Wiki-based knowledge makes this process extremely fluid. The convergence of wikis and more formalized representations in Semantic Wikis [39] suggests that similar processes might be possible for more formalized representations [22]. Wikipedia is a very interesting example of the slow emergence of semi-formal structures from a largely informal beginning, both in terms of representation and social structure.

6 Conclusion: Knowledge Systems and Open Science

Science is in the business of producing objective representations of reality, but objectivity is the end goal, not the starting point or the road to the destination. Scientists in practice use representations that are local, tentative, situated, subjective, semi-formalized (at best) and subject to the pull of conflicting interest. So, computationalists need to acknowledge the way representations are actually used, by scientists and others. If computational systems are to support the process of science they must take into account the way science-in-progress works, and provide a way for these messy sorts of representations to be generated, shared, and processed.

⁸<http://freebase.com>

⁹http://en.wikipedia.org/wiki/Wikipedia:Template_messages/Cleanup

Science is almost defined by its values of open publication of results. Science invented open knowledge centuries before the web existed [20], but in this age of the web there is a movement to get science to utilize the tools of the web to share knowledge with greater speed and flexibility than it has been able to do in the past. Such efforts involve any or all of: standard representations for data, public repositories for data (e.g Geo [4]) and workflows [53], open access publications like the Public Library of Science¹⁰, and arXiv¹¹ and their convergence in web-scale science [75].

The movement for Open Notebook Science [9] aims to make these representations more public, promising a more collaborative, accelerated, and open form of scientific knowledge management than the traditional publishing models. But such efforts are likely to founder without some kind of shared basis of representation. Between the private, radically situated scribbblings in individual lab notebooks and the crystalline, formal, supposedly objective status of realist ontologies lies a vast space of possible representational schemes and practices to be explored.

The fact that ontologies, like anything else, are subject to conflicting social forces should not come as a surprise. Nor should it be something that is swept under the rug. Ontologists are trying to put scientific representation on a sounder footing, but it seems to me they are going about it in the wrong way. There is a vast and largely unexplored design space in between pure chaos and rigid formalism.

6.1 A Platform for Knowledge

What would a scalable, non-realist, distributed knowledge representation regime look like? An earlier paper described the concept of a Knowledge Operating System (or KnowOS) [76]. As a traditional operating system provides an abstraction layer between users and applications on one side and computer hardware on the other, a KnowOS provides a knowledge abstraction layer. Just as an operating system does not dictate what a user or programmer can do with a computer, a knowledge operating system doesn't dictate how a community represents knowledge – it provides a platform, a framework, a common *lingua franca* by which knowledge and the users and creators of knowledge (both human and computational) can come together. And a truly web-scale knowledge representation scheme has to be a universal platform, one capable of encompassing the vast diversity of knowledge as it

¹⁰<http://www.plos.org>

¹¹<http://arxiv.org>

is actually used. How can we, as computationalists, design tools that make it easier for this magic to happen?

One early realization of the KnowOS concept, BioBike [24], provided a knowledge abstraction in the form of a frame system, along with a web-based interface for collaborative end-user programming by scientists. BioBike serves a number of scientific communities, provided a platform for the development of more complex tools [?, 16], and has been used as a teaching tool in several bioinformatics courses. Later efforts were made to integrate its knowledge representation system with semantic web standards.

BioBike differs in two important ways with typical knowledge platforms (eg, Protege [26]). First, it imposed very little in the way of ontological method, preferring to import a multitude of existing knowledge bases with their structures intact. Second, it emphasized computing with knowledge – the end-goal of a BioBike user was not the ontology itself but a knowledge-based scientific computation. These differences of emphasis are what makes BioBike a true knowledge-based computational medium rather than merely a user interface to a knowledge base.

BioBike really just scratches the surface of the problem, but it suggests the possibility of a larger-scale infrastructure for distributed knowledge-based computing, with a model that supports and redxxintegrates both human and computational use of representations. Such an infrastructure would need to be robust to the vagaries of real-world knowledge. At minimum, that means it would need to be capable of encompassing the variety of phenomena covered above, including: disagreements and their resolutions, prototype effects in categorization, and the evolution of categories due to interest groups. Designing a framework with these capabilities would not be an easy task, but is necessary if the evolving world-wide computational knowledge infrastructure is going to enhance our thinking, rather than constrain it.

Acknowledgements

Carl Hewitt, Linda Hershenson, Jeff Shrager, and David Chapman read drafts and provided valuable feedback. Partial funding provided by SRI International.

References

- [1] Anonymous. Gay marriage: the database engineering perspective. <http://qntm.org/gay>

- [2] PE Agre. Writing and Representation. In Phoebe Sengers, editor, *Narrative Intelligence*. John Benjamins Publishing Company, 2003.
- [3] S Auer, C Bizer, G Kobilarov, and J Lehmann. Dbpedia: A nucleus for a web of open data. *The Semantic Web*, 2007.
- [4] T Barrett, D B Troup, S E Wilhite, P Ledoux, D Rudnev, C Evangelista, I F Kim, A Soboleva, M Tomashevsky, K A Marshall, K H Phillippy, P M Sherman, R N Muerter, and R Edgar. NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Research*, 37(Database):D885–D890, January 2009.
- [5] R Bayer. *Homosexuality and American psychiatry: The politics of diagnosis*. the politics of diagnosis. Princeton Univ Pr, 1987.
- [6] I Beschastnikh and T Kriplean. Wikipedian self-governance in action: Motivating the policy lens. *Proc ICWSM 2008*, 2008.
- [7] AH Borning. Classes versus prototypes in object-oriented languages. In *Proceedings of 1986 ACM Fall joint computer ...*, 1986.
- [8] Ronald J Brachman. I Lied About the Trees, Or, Defaults and Definitions in Knowledge Representation. *AI Magazine*, 6:1–14, February 1985.
- [9] Jean-Claude Bradley. Open Notebook Science Using Blogs and Wikis. *Nature Precedings*, (713), June 2007.
- [10] S Buckingham Shum, E Motta, and J Domingue. ScholOnto: an ontology-based digital library server for research documents and discourse. *International Journal on Digital Libraries*, 3(3):237–248, 2000.
- [11] B Butler and E Joyce. Don’t look now, but we’ve created a bureaucracy: the nature and roles of policies and rules in wikipedia. In *Proceeding of the twenty-sixth annual CHI Conference*, 2008.
- [12] Werner Ceusters and Barry Smith. Foundations for a realist ontology of mental disease. *Journal of Biomedical Semantics*, 1(1):10, September 2010.
- [13] VK Chaudhri, A Farquhar, R Fikes, Peter Karp, and James Rice. OKBC: A programmatic foundation for knowledge base interoperability. In *Proceedings of the ...*, 1998.

- [14] Ayoun Cho, Hongseok Yun, Jin Hwan Park, Sang Yup Lee, and Sunwon Park. Prediction of novel synthetic pathways for the production of desired chemicals. *BMC Systems Biology*, 4:35, 2010.
- [15] T Clark and J Kinoshita. Alzforum and SWAN: the present and future of scientific web communities. *Briefings in Bioinformatics*, 8(3):163–171, May 2007.
- [16] R Cornel, RS Amant, and Jeff Shrager. Collaboration and modeling support in CogLaborate. In *19th Conference on Behavior Representation in Modeling Simulation*, 2012.
- [17] L Cottret, P Milreu, V Acuña, and A Marchetti. Graph-Based Analysis of the Metabolic Exchanges between Two Co-Resident Intracellular Symbionts, *Baumannia cicadellinicola* and *Sulcia muelleri*, with *PLoS Computational . . .*, January 2010.
- [18] Ubiratan d’Ambrosio and Ubiratan d’Ambrosio. *Ethnomathematics and Its Place in the History and Pedagogy of Mathematics*. 1985.
- [19] Lorraine Daston and Peter Galison. *Objectivity*. Zone Books (NY), October 2010.
- [20] PA David. From Keeping "Nature’s Secrets" to the Institutionalization of "Open Science". *Code: Collaborative Ownership and the Digital . . .*, 2005.
- [21] Emek Demir, Michael P Cary, Suzanne Paley, Ken Fukuda, Christian Lemer, Imre Vastrik, Guanming Wu, Peter D’Eustachio, Carl Schaefer, Joanne Luciano, Frank Schacherer, Irma Martinez-Flores, Zhenjun Hu, Veronica Jimenez-Jacinto, Geeta Joshi-Tope, Kumaran Kandasamy, Alejandra C Lopez-Fuentes, Huaiyu Mi, Elgar Pichler, Igor Rodchenkov, Andrea Splendiani, Sasha Tkachev, Jeremy Zucker, Gopal Gopinath, Harsha Rajasimha, Ranjani Ramakrishnan, Imran Shah, Mustafa Syed, Nadia Anwar, Özgün Babur, Michael Blinov, Erik Brauner, Dan Corwin, Sylva Donaldson, Frank Gibbons, Robert Goldberg, Peter Hornbeck, Augustin Luna, Peter Murray-Rust, Eric Neumann, Oliver Reubenacker, Matthias Samwald, Martijn van Iersel, Sarala Wimalaratne, Keith Allen, Burk Braun, Michelle Whirl-Carrillo, Kei-Hoi Cheung, Kam Dahlquist, Andrew Finney, Marc Gillespie, Elizabeth Glass, Li Gong, Robin Haw, Michael Honig, Olivier Hubaut, David Kane, Shiva Krupa, Martina Kutmon, Julie Leonard, Debbie Marks,

- David Merberg, Victoria Petri, Alex Pico, Dean Ravenscroft, Liya Ren, Nigam Shah, Margot Sunshine, Rebecca Tang, Ryan Whaley, Stan Letovksy, Kenneth H Buetow, Andrey Rzhetsky, Vincent Schachter, Bruno S Sobral, Ugur Dogrusoz, Shannon McWeeney, Mirit Aladjem, Ewan Birney, Julio Collado-Vides, Susumu Goto, Michael Hucka, Nicolas Le Novère, Natalia Maltsev, Akhilesh Pandey, Paul Thomas, Edgar Wingender, Peter D Karp, Chris Sander, and Gary D Bader. The BioPAX community standard for pathway data sharing. *Nature Biotechnology*, 28(9):935–942, September 2010.
- [22] F Dengler, S Lamparter, and M Hefke. Collaborative process development using semantic mediawiki. In *Proceedings of the 5th ...*, 2009.
 - [23] M Dumontier and R Hoehndorf. Realism for scientific ontologies. *Proceeding of the 2010 conference on Formal Ontology in Information Systems: Proceedings of the Sixth International Conference (FOIS 2010)*, pages 387–399, 2010.
 - [24] Jeff Elhai, Arnaud Taton, J P Massar, John K Myers, Mike Travers, Johnny Casey, Mark Slupesky, and Jeff Shrager. BioBIKE: a Web-based, programmable, integrated biological knowledge base. *Nucleic Acids Research*, 37(Web Server issue):W28–32, July 2009.
 - [25] P Feyerabend. Against method: Outline of an anarchistic theory of knowledge. Verso, 1975.
 - [26] JH Gennari, MA Musen, and RW Fergerson. The evolution of Protégé: an environment for knowledge-based systems development. *International Journal of ...*, 2003.
 - [27] R Glenn and R Chris. Translating Wigmore Diagrams. *Computational models of argument: proceedings of COMMA 2006*, page 171, 2006.
 - [28] PR Gross and Norman Levitt. *Higher superstition: The academic left and its quarrels with science*. Johns Hopkins University Press, 1994.
 - [29] TR Gruber. A translation approach to portable ontology specifications. *Knowledge acquisition*, 1993.
 - [30] Ian Hacking. *The social construction of what?* Harvard Univ Pr, 1999.
 - [31] C Hewitt. Common sense for inconsistency robust information integration using Direct Logic (TM) and the Actor Model. 2008.

- [32] W James. What pragmatism means. *Essays in pragmatism*, 1948.
- [33] Laura R Jarboe, Xueli Zhang, Xuan Wang, Jonathan C Moore, K T Shanmugam, and Lonnie O Ingram. Metabolic Engineering for Production of Biorenewable Fuels and Chemicals: Contributions of Synthetic Biology. *Journal of Biomedicine and Biotechnology*, 2010:1–19, 2010.
- [34] Ingvar Johansson. Bioinformatics and biological reality. *Journal of Biomedical Informatics*, 39(3):274–287, June 2006.
- [35] M Kanehisa. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 2000.
- [36] P D Karp, S M Paley, M Krummenacker, M Latendresse, J M Dale, T J Lee, P Kaipa, F Gilham, A Spaulding, L Popescu, T Altman, I Paulsen, I M Keseler, and R Caspi. Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Briefings in Bioinformatics*, 11(1):40–79, January 2010.
- [37] IM Keseler and C Bonavides-Martínez. EcoCyc: a comprehensive view of Escherichia coli biology. *Nucleic acids ...*, 2009.
- [38] M Klein and L Iandoli. Supporting Collaborative Deliberation Using a Large-Scale Argumentation System: The Mit Collaboratorium. *the proceedings of the 11th Symposim on ...*, 2008.
- [39] M Krötzsch, D Vrandečić, and M Völkel. Semantic mediawiki. *The Semantic Web-ISWC 2006*, pages 935–942, 2006.
- [40] George Lakoff. *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. University of Chicago Press, 1990.
- [41] B Latour. Drawing things together. In Michael Lynch and Steve Woolgar, editors, *Representation in Scientific Practice*. MIT Press, 1990.
- [42] B Latour. *We have never been modern*. Harvard Univ Pr, 1993.
- [43] H Lieberman. Using prototypical objects to implement shared behavior in object-oriented systems. *ACM SIGPLAN Notices*, 21(11):214–223, 1986.
- [44] MO Little. Abortion and the Margins of Personhood. *Rutgers Law Journal*, 2007.

- [45] P Lord and R Stevens. Adding a Little Reality to Building Ontologies for Biology. *PLoS ONE*, 5(9):e12258, 2010.
- [46] Philip Lord. Why Realism is Wrong, September 2010. <http://www.russet.org.uk/blog/2010/09/why-realism-is-wrong/>
- [47] W Malisoff. What Is a Gene? *Philosophy of Science*, January 1939.
- [48] G Merrill. Realism and reference ontologies: Considerations, reflections and problems. *Applied Ontology*, January 2010.
- [49] Marvin Minsky. A Framework for Representing Knowledge (MIT-AI Memo 306). MIT AI Laboratory, 1974.
- [50] Marvin Minsky. Logical Versus Analogical or Symbolic Versus Connectionist or Neat Versus Scruffy. *AI Magazine*, 12(2):1–18, January 1996.
- [51] A Moor and M Aakhus. Argumentation support: from technologies to tools. *Communications of the ACM*, 49(3):98, 2006.
- [52] F Neuhaus and P Grenon. A formal theory of substances, qualities, and universals. . . . of the 3rd International Conference on . . . , 2004.
- [53] D Newman and S Bechhofer. myExperiment: An ontology for e-Research. page 408, 2009.
- [54] Mark J Pallen. Time to recognise that mitochondria are bacteria? *Trends in Microbiology*, 19(2):58–64, February 2011.
- [55] DM Pisanelli. Biodynamic Ontology: Applying BFO in the Biomedical Domain. *Ontologies in medicine*, page 20, 2004.
- [56] ME Reicher and JC Marek. Against Fantology. *Erfahrung und Analyse*, page 153, 2005.
- [57] Pedro Romero, Jonathan Wagg, Michelle L Green, Dale Kaiser, Markus Krummenacker, and Peter D Karp. Computational prediction of human metabolic pathways from the complete human genome. *Genome Biology*, 6(1):R2, December 2004.
- [58] R Rorty. *Objectivity, relativism, and truth*. Cambridge University Press, 1991.

- [59] E Rosch. Prototype classification and logical classification: The two systems. In EK Scholnick, editor, *New Trends in Conceptual Representation: Challenges to Piaget's Theory?*, page 109. Lawrence Erlbaum Associates, 1985.
- [60] Warren Sack. *Design for very Large-scale Conversations*. PhD thesis, September 2009.
- [61] Thomas E Schacht. DSM-III and the politics of truth. *American Psychologist*, 40(5):513–521, 1985.
- [62] E Schrödinger. What is life?: the physical aspect of the living cell: with Mind and matter & Autobiographical sketches. Cambridge Univ Pr, 1992.
- [63] Wilbur J. Scott. PTSD in DSM-III: A Case in the Politics of Diagnosis and Disease. *Social Problems*, 37:1–18, March 1990.
- [64] Rekha Seshadri, Marvin Frazier, Saul A Kravitz, Larry Smarr, and Paul Gilna. Oceanic Metagenomics: CAMERA: A Community Resource for Metagenomics. *PLoS Biology*, 5(3), March 2007.
- [65] J Shrager. The fiction of function. *Bioinformatics*, 19(15):1934–1936, October 2003.
- [66] J Shrager. *Diary of an insane cell mechanic*. Scientific and technological thinking, 2005.
- [67] B Smith. Beyond concepts: ontology as reality representation. *Formal Ontology In Information Systems: Proceedings of the Third International Conference (FOIS-2004)*, pages 73–84, 2004.
- [68] B Smith. Ontological realism: A methodology for coordinated evolution of scientific ontologies. *Applied Ontology*, January 2010.
- [69] B Smith, W Kusnierczyk, D Schober, and W Ceusters. Towards a reference terminology for ontology research and development in the biomedical domain. *Proceedings of KR-MED*, 2006:57–65, 2006.
- [70] Barry Smith, Michael Ashburner, Cornelius Rosse, Jonathan Bard, William Bug, Werner Ceusters, Louis J Goldberg, Karen Eilbeck, Amelia Ireland, Christopher J Mungall, Neocles Leontis, Philippe Rocca-Serra, Alan Ruttenberg, Susanna-Assunta Sansone, Richard H Scheuermann, Nigam Shah, Patricia L Whetzel, and Suzanna E Lewis.

- The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, 25(11):1251–1255, November 2007.
- [71] WR Smith. Using a prototype-based language for user interface: the Newton project’s experience. *ACM SIGPLAN Notices*, 30(10):61–72, 1995.
 - [72] AD Sokal. *Fashionable Nonsense*. 1998.
 - [73] IE Sutherland. Sketch pad a man-machine graphical communication system. In *Proceedings of the SHARE design automation ...*, 1964.
 - [74] Christoph Tempich, Elena Simperl, Markus Luczak, Rudi Studer, and H Sofia Pinto. Argumentation-Based Ontology Engineering. November 2007.
 - [75] Jay M Tenenbaum. AI Meets Web 2.0: Building the Web of Tomorrow, Today. pages 1–22, December 2006.
 - [76] M Travers and JP Massar. KnowOS: The (re) birth of the knowledge operating system. *The International Lisp Conference*, 2005.
 - [77] D Ungar and RB Smith. Self: The power of simplicity. *Conference proceedings on Object-oriented programming systems, languages and applications*, pages 227–242, 1987.
 - [78] M Weber. Economy and society. *New York*, 1968.