

An exploration into cognitive bias in ontologies

C. Maria Keet

University of Cape Town, 18 University Avenue, Rondebosch, Cape Town 7701, South Africa

Abstract

Ontologies and similar artefacts are used in a myriad of ontology-driven information systems and increasingly also linked to data analytics. Algorithmic bias in data analytics is a well-known notion, but what does bias mean in the context of ontologies that provide a structuring mechanism for, e.g., an algorithm's or query's input? What are the sources of bias there, and cognitive bias in particular, and how do they manifest in ontologies? We examined and enumerated eight broad sources that can cause bias that may affect an ontology's content. They are illustrated with examples from extant ontologies and samples from the literature. We then assessed three concurrently developed COVID-19 ontologies on modelling bias and detected different subsets of types of bias in each one, to a greater or lesser extent. This first characterisation aims contribute to a sensitisation of bias in ontologies primarily regarding representation of the knowledge.

Keywords

Ontology development, Cognitive bias, Modeling, AI Ethics

1. Introduction

Bias in IT systems is a popular topic in the media and recent special tracks at conferences. Nearly all reports on bias in 'models' concern *statistical* or *neural* models created from some Big Data dataset by means of knowledge discovery, machine learning, and deep learning techniques. There are many more types of models, however, notably ontologies and knowledge graphs, whose success stories include, among others, the Gene Ontology [1] and representation language standards such as OWL [2], and the repository of ontology repositories OntoHub has indexed 22460 ontologies gathered from of 139 repositories¹. Ontologies for information systems provide an application-independent representation of the subject domain. Besides their use in data integration, one also can choose an ontology upfront and use it in a stand-alone application or on the Web, such as Google's Knowledge Graph that drives the creation and maintenance of its infoboxes [3]. The person who builds and controls the ontology or knowledge graph, then, is the one who has the power to control presentation and access to information and possibly also the recording of information. In the case of Google's Graph, Vang argues it "to some degree contests the autonomy of the user" [4].

It is not inconceivable that ontologies also may be biased, which would then propagate into the information systems. A recurring theme is "encoding bias" [5], which refers to merely

Cognition And OntologieS (CAOS'21)

✉ mkeet@cs.uct.ac.za (C. M. Keet)

🌐 <https://www.meteck.org> (C. M. Keet)

🆔 0000-0002-8281-0853 (C. M. Keet)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

📄 CEUR Workshop Proceedings (CEUR-WS.org)

¹<https://ontohub.org/>; last checked on 13-1-2021.

different formalisations of the same thing. There is, however, scant literature about *cognitive bias* in ontologies, with, to the best of our knowledge, three preliminary papers only [6, 7, 8]. Gomes and Bragato Barros [6] assessed bias in the Friend of a Friend (FOAF) terminology through the lens of discursive semiotics as a method, which was limited to a few well-known issues, being first & last name (vs given & family), gender, and the meaning of document. Biases are more intricate and varied than these. For instance, and what Gomes' approach cannot capture, is, among others, the menopausal hormone therapy case²: there were at least *economic* incentives that determined which properties were represented in the medical ontology with what threshold values in order to classify who was eligible for treatment. An example in conceptual data modelling concerns the "Dirty War Index" tool that aimed to inform public health policy in armed conflict settings, where a politically motivated bias of illusory superiority was observed that affected the tool's outcome due to varying levels of granularity in the model in favour of the authors' side in the conflict [8]. For software and ontology development, research into cognitive bias is focussed on the biases in development processes [9, 10] rather than the resultant models, with aforementioned exceptions of the FOAF [6] and a few anecdotes [7, 8]. We are interested in an ontology's content first, i.e., possible effects of a bias, before investigation the development processes that may have caused it: if no outcomes can be detected, then one would not have to undertake the more costly investigation into processes.

In this paper, we aim to contribute to systematising the sort of bias that may be present in ontologies, including similar artefacts such as conceptual data models and thesauri, which may help their management and prevention. In addition to extending Gomes and Bragato Barros' three sources [6] to eight, we seek to provide a preliminary answer to what bias means for ontologies, what their sources are, and how that manifests itself in ontologies, and therewith go beyond the anecdotes in [6, 7, 8]. The identified broad categories for biases are structured along three categories: high-level philosophical ones, scope or purpose, and subject domain issues. Some of them are intentional biases that insiders know very well, but outsiders and newcomers need to become aware of. Second, we assess three COVID-19 ontologies on these biases. These ontologies are under active development and relevant for data and information management of the pandemic. The assessment showed that none is free of bias and contain different subsets of cognitive and other biases rather than explicitly competing as the better alternative.

The remainder of this paper is structured as follows. We systematise and illustrate the principal sources (Sect. 2) first and then assess the COVID-19 ontologies (Sect. 3). We discuss the outcomes and touch upon automated reasoning (Sect. 4) and close with conclusions (Sect. 5).

2. Principal Sources of Bias in Ontologies

Since the notion of bias in ontologies is unclear beyond a few recorded anecdotes for knowledge graphs [6, 7], the net was cast wider to also consult various lists and taxonomies of biases, notably [11, 12], and it also required availing of the author's experience in ontology development

²In short: the range of natural variability of concentrations of key molecules was narrowed down to increase the number of 'abnormal' women that then qualified for medication, which unintentionally led to an increase in cancer incidence. An accessible account of the complicated story can be found at: <https://www.cancer.org/cancer/cancer-causes/medical-treatments/menopausal-hormone-replacement-therapy-and-cancer-risk.html>.

(among others, [13, 14]), their assessments over the years in an educational setting including for the textbook [15], and inspection for use, reuse, and reviewing and so on. Therefore, we commence with a few preliminary considerations on the cognitive biases lists, to then proceed to a description of each of the identified bias sources for ontologies.

2.1. Preliminary Considerations

Of most consequences are the cognitive biases in task and domain ontologies, since they are expected to be used most often in information systems. There may be a knock-on effect from philosophical and engineering (encoding) dimensions, so that has to be considered as well. Also, an inclusive definition for bias is adopted, as in [12], where it is *a consequence of interference with honest attempts*, rather than only the narrow scope of norm deviation and error³.

The notion of bias has been interpreted in a varied manner, from clear short lists of cognitive biases to also include cognitive styles and alternate perspectives and image schemas. Cognitive biases in the literature for IT and computing more specifically, are grouped according to one's dimension of interest; e.g., by type of task (estimation, decision, recall, etc.) for information visualisation [11], by software engineering "knowledge area" (e.g., design, testing, management) [9] and as cognitive styles in the approach to ontology development [10]. Dimara et al's most recent and comprehensive list contains 154 cognitive biases [11]. Many of them are not relevant within the scope of ontologies, just like only a subset was relevant to them for information visualisation [11]. For instance, while the 'Verbatim effect'—it is easier to recall the gist of something than the sentence(s)—is indeed a bias, it does not affect the content of an ontology, and a 'Ballot names bias' on a voting sheet is completely irrelevant. Also, a large subset of that inventarisation have to do with *processes*, which may be relevant for assessing ontology development methodologies and practices. The borderline between processes and the content may be unclear and difficult to establish for individual ontologies when one has not been privy to the development process. For instance, the 'Mere-exposure effect' cognitive bias might be observed when choosing to reuse a particular ontology because of its familiarity more so than based on critical evaluation of the ontologies, and an outsider would not know when the rationale remains undocumented.

A second note of consideration is whether a bias may be implicit or explicit. Here, we take 'explicit' to mean a conscious decision for one thing or another to appear as such in the ontology. For instance, a choice for developing a realism-based ontology, or explicitly representing that Tomato \sqsubseteq Fruit, since it is so according to science and the classification criteria of biologists. Other choices may be 'implicit', in the sense of not having realised it at the time of modelling. They may, or may not, be, unintended in hindsight, but at least it was not a deliberate choice among alternatives considered at development time.

We categorise the biases by broad sources of a domain, in style with the focus on an ontology's content. We identified eight broad sources, which are summarised in Table 1. They will be elaborated on in the next section. They are, at present and for the purpose and limited length of

³An ontological analysis of cognitive bias may be of use, since definitions of bias differ across the scientific and popular literature, to go to being as inclusive as bias being an image schema or viewpoint that is different from one's own. Also, the notion of 'norm deviation' may already be difficult to operationalise for an ontology that claims to be 'universal', since norms vary widely across the globe.

Type	Subtype	[im/ex]plicit bias
Philosophical	-	explicit
Purpose	-	explicit
Subject domain	Science	explicit
	Granularity	either
	Linguistic	either
	Socio-cultural	either
	Political or religious	either
	Economics	explicit

Table 1

Summary of typical possible biases in ontologies grouped by source, with an indication whether such biases would be explicit choices or whether they may creep in unintentionally and lead to implicit bias.

this paper, high-level groupings that affect the representation of knowledge in the ontologies.

2.2. High-level Philosophical Viewpoints

Ontologies are an engineering version of the original idea of Ontology by philosophers. Most subject domain ontology developers may not be interested in the finer distinctions of core notions, but they are there. Practically, for domain ontology development, one may choose a particular foundational ontology that provides the main types of entities and relations so as to help structuring the content. There are multiple top-level and foundational ontologies, such as BFO [16], DOLCE [17], and YAMATO [18], which contain different commitments. Its developers are mostly clear about general principles and how it affects the ontology’s content, such as whether it is descriptive or realist. This has a consequent effect on the ontology’s content, such as admitting the existence of abstract entities or not [17], and what the core relations in the world would be [19]; comprehensive comparisons can be found in [20, 21]. Choosing a foundational ontology is a deliberate decision; hence, an explicit bias.

There are related debates about whether an ontology’s contents is a representation of reality or our understanding thereof, or whether there would even be a reality. This is a recurring debate (see, e.g., [22]) that has no resolve that everyone agrees on. For ontology development, the key issue to determine is whether one aims to be faithful to reality (or our best understanding of it) versus ulterior motives, be it rejecting reality or not caring (‘post-truth’) or knowingly violating it for some reason. These different stances have their main effects at the subject domain level where the bias can have most effect, as we shall see below, and could result either in an explicit or an implicit bias.

2.3. Purpose with Encoding Bias

In theory and historically, ontologies are supposed to be application-independent, so as to be a solution to the data integration problem. If they are tailored to the application nonetheless, they may become part of the problem. This application independence may not always hold. Developing an ontology for the sake of it may be an interesting endeavour, but someone has to fund it and it helps to have a use case to motivate for its development. This may affect what is

represented and how and is an explicit bias motivated by pragmatics. Uschold and Gruninger call this “encoding bias” [5], which are engineering choices rather than a cognitive bias and they may give rise to “modelling styles” when several recurring choices are taken together [23]. This is normally not considered as a bias in AI ethics and bias research in cognitive science. We illustrate it nonetheless, because it affects the actual representation of knowledge in the ontology and, upon closer inspection, may induce arguments about the ontological nature of things. Consider the following three characteristic approaches for different scenarios for the same knowledge, to represent that ventilation is a treatment for COVID-19 patients:

- A. If the aim is to be as detailed and reusable as possible, Treatment can be a type of perdurant (colloquially: a type of process) within the common 3-dimensionality viewpoint (objects persist in time) and then $\text{Ventilation} \sqsubseteq \text{Treatment}$ is the bare minimum to declare. Availing of the core *participation* relation between objects and processes, then also $\text{Patient} \sqsubseteq \exists \text{participatesIn.Treatment}$. One then may assert that our hospitalised COVID-19 patients participate in the ventilation treatment.
- B. A compact representation results in faster data processing, such as for ontology-based data access. For instance, $\text{Patient} \sqsubseteq \exists \text{isOnVentilator.Boolean}$, where one uses the ontology language to develop a conceptual data model for databases, rather than the traditional Extended Entity-Relationship language. This is a different ontological commitment from option 1, because the ventilation treatment is now a property of a person, not a relational property between two entities.
- C. If the aim is, e.g., annotation of literature to better manage it, then neither the boolean nor all those constraints and relations are needed. Instead, one casts the net wide regarding terminology by specifying preferred and alternative labels, including the alternately used (but not equivalent) terms ventilator support, ventilation therapy, mechanical ventilation, and invasive ventilation that have Broader Term (BT) Ventilation and Related Term (RT) Patient.

The ontologist may complain about the latter two options as woefully underspecified or too imprecise, whereas the tool developers of options 2 and 3 may complain that the first option is too complicated due to their preference for scalability or simplicity.

2.4. Subject Domain

The list of bias sources described in this section overlaps with Gomes and Bragato Barros’ one [6], but is extended with three categories, including one from [8]. In addition, we indicate whether they concern mainly explicit or implicit biases, or both, and illustrate each one in order to demonstrate relevance.

2.4.1. Difference of Opinion on Reality and Science

Even under the assumption of a commitment to the existence of reality, one still could disagree. A common example is whether a virus is an organism or not; it is not by any extant definition of what an organism is. Bio-ontologies and medical terminologies do not agree on this matter, such as the CIDO [24] versus the SIO [25] ontologies. More broadly, it concerns either the insufficient insight or competing theories that the scientists still have to investigate, or there are delays in propagating discoveries into the ontologies. It is assumed that eventually there will be

an agreement. In other areas, there are inherently competing theories, such as capitalism and socialism, that would result in a different domain ontology of economy. They are all intended choices and, arguably, just different image schemas or perspectives, or indeed biases.

2.4.2. Required or Chosen Level of Precision/Granularity

It is a general question in ontology development how detailed it should be and how deep the taxonomy should go. Less detail therefore may be an *act of omission*, an indication of *not needed*, which both may be results of a cognitive bias, or be merely a *ran out of time* situation for it to be included in a next release. Inspecting an ontology in isolation, this is impossible to determine unless it is explicitly stated in the annotations or accompanying documentation. For instance, The Gene Ontology has three versions: a GO basic that excludes several relations between entities, the GO, and a GO plus with additional axioms⁴.

An act of omission is aggregating ex-military persons with non-involved persons as one group of Civilians, which happened with aforementioned Dirty War Index tool even though the original source had a more detailed categorisation [8]. Similar issues exist for other conflict databases, which may be intentional or unintentional. For instance, a bombing may be recorded as an instance of having targeted a Government building if that is the only class available, or more precisely only if it had a hierarchy of subclasses, such as Health facility with subclasses including State hospital and State medicine manufacture plant and a Defence facility class with, say, Military base and Homeland security torture bunker as subclasses, rather than one layer of subclasses as in [26]. Such differentiations, or absence thereof, may be intended or they may be unintended.

2.4.3. Cultural-linguistic Motivations

Anyone who has learned a second language has come across untranslatable words or at least fine semantic distinctions. The question then arises if, and if so, when, a difference is a bias in the ontology or not. For instance, English has only one term for river—all rivers are just rivers—whereas French makes a distinction between a *fleuve* and a *rivière*—one flows into another river, the other flows into the sea—that somehow has to be represented and the ontologies aligned [27]. It is likewise for observed differences in part-whole relations across languages [14]. One may argue that in both examples, the reality is the same but they have varying descriptions, or that there are different conceptualisations, or use the term reality more liberally and state that there are different realities depending on language. This is, at the core, an explicit choice between philosophical viewpoints.

A borderline case between cultural-linguistic preferences and political bias are the false friends, where a term in a language has a different meaning or connotation in different countries where the language is spoken, due to historical differences across countries. For instance, ‘herd immunity’ is a common term in American and British English, but is being rebranded as ‘population immunity’ in South African English, since the former has the connotation of non-human animals that people do not want to be associated with. It is also different in other languages; e.g., in Spanish, it is *inmunidad de grupo* and Dutch *groepsimmuniteit*, i.e., ‘group’ immunity rather than ‘herd’. Note that this semantic shift is distinct from mere synonymy

⁴<http://geneontology.org/docs/download-ontology/>; last accessed 13-1-2021.

confusion, such as the Football Ontology by name being ambiguous about whether it refers to the soccer football or American football, and orthographic differences (e.g., color vs colour) that simply can be accommodated in the ontology with labels and finer-grained language-coding schemes (e.g., @en-uk and @en-us etc.).

Pushing a monolingual ontology or one that has a ‘standard’ natural language for naming entities then at least amounts to imposing one particular viewpoint or schema and any bias that comes with it. The chance that a monolingual ontology development team from one cultural identity in one country builds in such a bias is substantial, and it can be reduced by constituting a more diverse team of ontology developers who at least speak several languages among them. Any bias built in may be intentional or unintentional.

2.4.4. Socio-cultural Factors

This concerns hows society is organised, with the assumptions that underlie it and history how it came about, and any practical effects it may have when developing the ontology. This may be organisational structures, who lives with whom, demographics, allocation of resources, or social geography that influences what is salient and what not. For instance, who can marry whom and how many is a well-known point of variation across the world, which can cause difficulties for multinational organisations to harmonise that in one system by means of an ontology of organisations. For instance, it may be a company policy that one can insure the spouse of the employee, requiring a statement alike $\text{Employee} \sqsubseteq \forall \text{marriedTo.Spouse}$, but should the model also include $\text{Employee} \sqsubseteq \leq 1 \text{ marriedTo.Spouse}$, i.e., at most one spouse? Should the gender of the spouse be recorded? Any answer has at least a perspective, if not a bias, embedded in it. For an ontology to be universal, the most permissive constraint would be represented, and any stricter constraints will have to go into the conceptual data model for the database.

A concrete example is the relatively popular GoodRelations Ontology for e-commerce [28]. It lists several payment methods, but limits the ‘on delivery’ to cash only, even though cash-less options are also possible (e.g., a pre-paid card or QR-code), which is an accepted method of payment in areas where robberies are common (although it cannot be excluded that the omission was due to ‘ran out of time’). Also, its Business requires that they are legally registered, which may well hold in Europe where the ontology was developed, but in many other countries there is a vast network of the informal economy that does trade online with their smartphones and it has no specific opening hours either.

Socio-cultural factors may also influence the content of medical terminologies, such as the perception of alcohol use across cultures, and what qualifies as having a drinking problem. A recent example is demonstrated by a comparison between the Diagnostic and Statistical Manual of Mental Disorders and the International Classification of Diseases, and their versions DSM-IV, DSM-V, and ICD-10 in particular, on issues with alcohol intake, where the criteria were changed. Based on the same data, it resulted in an increase in Alcohol Use Disorder using DSM-V compared to the DSM-IV criteria. This was primarily due to lowering the threshold for the number of diagnostic criteria (properties) required for it and increasing the number of criteria through replacing one class with four new classes that were arguably features of it [29]. This modification in the lightweight ontology has been blamed on a combination of socio-cultural factors and scientific disagreement [30].

2.4.5. Political and Religious Motivations

The line between societal bias, political, and religious may be difficult to draw depending on the case. Aforementioned DSM, which ought to be based on science, was not entirely and likely was influenced by religious viewpoints at least in some instances. Since the separation between state and church may not be all that strict, it practically may not be possible to disentangle the two. A clear-cut case, however, is where the entity type Aggrieved group, as a neutral term, enters the ontology as Terrorist organisation as preferred label. Concretely, there are terrorist and terroristgroup in the terrorism ontology of [31], compared to an ActorEntity with various types of Insiders and Protestors in the Cyberterrorism ontology of [26].

As with society and language matters, these issues more easily come to light if the team of ontology developers is diverse or at least has diverse knowledge to bring in. Also here, such differences may be intentional or not.

2.4.6. Economic Motivations

The, perhaps, most well-known arena where economic motivations play a role, is the recognition of something as a disorder or disease, from which follows whether it deserves at least funding of a treatment if there is one, as well as resources for prevention and research. The Obesity Society's panel of experts even stated this bluntly as the main reason in favour of classifying obesity as a subclass of disease [32]. Its recognition is good for big pharma and possibly also the patients, but costly for insurers, which results in tension. For the ontology, it means that it is either included or not, whose decision propagates to electronic health records that make use of the ontology to annotate findings and propose treatments, and when they are linked to the pharmacy and the health insurer's databases. These issues are well-known and therefore are classified as explicit biases, as deviations to the norm.

3. Cognitive Biases Assessment: the COVID-19 Ontologies

The aim of the evaluation is to assess bias in ontologies systematically beyond the selected examples in the previous section. In particular, to examine it on a set of ontologies in the same subject domain so that consequences of cognitive bias on the same task might be assessed. Also, since there are several ontologies in that given domain, at least differences in perspective may have been a reason to develop another one. In narrowing down the choice for selection of core or domain ontologies to assess, we note that there are a few on time and measurements, many on health and medicine (e.g., 37 are contextualised in [33]), data mining, organisations and government, and others, which are more or less stable and more or less maintained. Among the sets of same domain, it is narrowed down to ontologies that are under active development and, ideally, from the same timeframe so that there has been less chance of mutual influence (and setting aside the caveat that any biases observed may have been resolved in the meantime). And, as last criterion, in a domain that the author has sufficient knowledge about. Applying these criteria, it resulted in the selection of the three COVID-19 ontologies that were developed in 2020. The next section contextualises each ontology and thereafter they are assessed.

3.1. Ontology Descriptions

The Coronavirus Infectious Disease Ontology (CIDO) [24] is an ontology that was developed within the OBO Foundry approach [34]. This entails community-based development and reuse of OBO Foundry ontologies, such as the Infections Disease Ontology that in turn is linked to the top-level ontology BFO [16]. The scope of the ontology was aimed at knowledge and information about the SARS-CoV-2 virus and host taxonomy data, its phenotype, and drugs and vaccines to foster data integration. CIDO v1.0.109 was used for the assessment, to keep with the time frame where all ontologies were released around July-August 2020; specifically, `cido-base.owl` (downloaded on 20-7-2020) with the relevant imports was assessed. It contains 82 classes, 15 object properties, no data properties, and one individual, and 90 logical axioms and is within the OWL Full profile due to undeclared annotation properties and a few undeclared classes; logically, it is expressible in the Description Logic $\mathcal{AL}\mathcal{E}\mathcal{H}\mathcal{O}$, i.e., it is a basic hierarchy with existentially quantified properties and an occasional nominal.

The COviD-19 Ontology (CODO) [35] has as purpose to assist in representing and publishing of COVID-19 data from the disease course perspective and has subject domain scope COVID-19 cases and patient information. That is, it aims to be a component in IT systems for healthcare. The CODO `v1.2-16July2020.owl` was used for the assessment, which contains 51 classes, 61 object properties, 45 data properties, 56 individuals, and 463 logical axioms. It is within the OWL 2 DL profile, and $\mathcal{SHOIQ}(D)$ more specifically, i.e., it is an expressive ontology that uses many of the OWL 2 DL constructs available in the language.

The Coronavirus Vocabulary (COVoc) was developed by the European Bioinformatics Institute and has as purpose to support navigating and curating the literature on COVID-19, and in particular the scientific research of it. Documentation of its rationale is available as a workshop presentation [36]. Its first, and latest, released version is slightly later than that of CIDO and CODO, although all had their drafts in June 2020, which did not affect its contents. The `covoc.owl` was used for the assessment (d.d. 28-8-2020), which contains 541 classes, 179 object properties, no data properties or individuals, and 672 logical axioms. It is within the OWL Full profile due to a subset property issue with the annotation properties; just the logical theory is expressible in $\mathcal{ALC}\mathcal{HL}$. In practice, it means it consists of a basic hierarchy with existentially quantified properties and a few subproperties and inverses.

3.2. Bias Assessment

The presence and absence of the different types of bias are summarised in Tables 2 and 3, and will be discussed in the remainder of this section.

3.2.1. CIDO

There are two *socio-cultural* biases in the CIDO. First, there is a COVID-19 diagnosis class with three subclasses: negative, positive, and presumptive positive. There are two aspects that stem from the *negative interpretation cognitive bias*. First, the [disease]-positive/negative labeling has clear HIV connotations with stigmatisation and ostracisation from its worst times when little of it was understood. Test outcomes and diagnoses easily could have been the more common ‘infected’, ‘detected’, or ‘present’ and ‘not infected/detected’ or ‘absent’, alike a patient has a flu

Bias (Source/type)	CIDO	CODO	COVoc
Philosophical	+	-	+
Purpose	-	+	+
Science	-	-	+
Granularity	±	+	+
Linguistic	+	-	-
Socio-cultural	+	+	+
Political or religious	+	+	+
Economics	-	-	±

Table 2

Presence or absence of bias in the three COVID-19 ontologies examined, by source; see text for details.

Bias (Cognitive biases from Dimara et al's list)	CIDO	CODO	COVoc
Mere exposure/familiarity (choice is influenced by exposure to it and thus familiarity with it)	+		+
Negative interpretation (judgement is affected more by negative information than positive)	+		
Optimism (more positive predictions for oneself than for others)	+		
Naive realism (the belief that you experience objects in your world objectively)	+		
False Consensus (Overestimating that other people are and behave like you and agree with your opinion)		+	
Illusory truth effect (a statement is considered to be true after repeated exposure to it)			+

Table 3

Tentative presence of bias in the three COVID-19 ontologies, by cognitive bias; see text for details.

or meningitis infection but is not meningitis positive or flu negative. Second, the presumptive positive portrays a negativity bias as well, by playing into people's fears and would brand people that are statistically unlikely to have been infected, since the WHO guideline is at most 5% positivity rate and countries aim for that. Neutral and accurate terminology would be, e.g., 'pending', 'awaiting test outcome', or 'under investigation'.

Conversely, an unwarranted *optimism bias* is reflected in COVID-19 experimental drug in clinical trial \sqsubseteq COVID-19 drug, noting that COVID-19 drug $\sqsubseteq \exists$ treatment for.COVID-19 disease process is asserted in the ontology, and thus entails that COVID-19 experimental drug in clinical trial *is* a drug already and *is* being part of regular treatment processes of COVID-19, since the property of \exists treatment for.COVID-19 disease process is inherited down into the hierarchy. This is wishful thinking. A substance under investigation that is being evaluated is not necessarily effective or safe and for it to be a drug, it has to be effective and approved by the regulatory body.

A minor *language* note is drive-thru instead of drive-through for testing stations, but this can easily be addressed by providing alternative labels. The naming of SARS-CoV-2 also as the Wuhan virus, however, reflects a *political* stance, as the term was rarely used outside the USA since it was advocated by former President Trump and his policies toward China. The FDA EUA-

authorized organization as the only other organisation as sibling of drive-thru COVID-19 testing facility may, in a lenient reading, be argued to be an instance of ran-out-of-time, considering that the authors of the accompanying paper [24] have affiliations from different countries who may intend to add their FDA counterpart, or be a *familiarity/mere-exposure cognitive bias* and a *granularity* issue.

The *philosophical* view is evident by its embedding in the OBO Foundry [34], its reuse of ontologies within that framework, such as OBI and IAO, and the organisational principles how the ontology is structured, which follows the BFO foundational ontology design principles [24]. BFO being founded on the realism stance, content in the ontology may be subject to the *naïve realism cognitive bias*. The reuse may be categorised as a *mere-exposure cognitive bias*, which is by design and Foundry principles, and thus explicit.

In sum, CIDO takes the science angle to representing knowledge about COVID-19, but with a few biases toward USA-centrism, which reduces its off-the-shelf potential. Or: possible CIDO adoption in Europe or any of the key Global South countries with ample research, testing, or production capacities, such as India and South Africa, requires modifications to CIDO first.

3.2.2. CODO

The CODO fares slightly better than CIDO on the Laboratory test finding, which can be negative, positive, or pending. It does have the well-known issue with representing gender or sex, which is represented as Gender type $\equiv \{\text{Female, Male}\}$ in CODO. A clear *socio-cultural bias* axiom in the ontology is $\text{InfectedSpouse} \sqsubseteq \text{InfectedFamilyMember}$ and $\text{InfectedFamilyMember} \sqsubseteq \text{Exposure to COVID-19}$. This may indicate omissions or time constraints, since the only family member that can be infected is the spouse according to CODO. It does allow for more family members by means of subproperties of *hasRelationships*, but it is not linked to the Exposure class. More importantly for epidemiological control, is the cultural viewpoint jointly with the *consensus bias* by centering on the concept of the (nuclear) family with parents and their children. Globally more broadly applicable is the notion of household that admits a wider range of composition, such as live-in grandparents, cousins, nannies, domestic workers, and so on, but where spouses may not live together in one household due to being a migrant worker. Such complexities in the context of COVID-19 are recorded and investigated [37], so if CODO were to be used elsewhere, then this branch needs revision.

CODO's *purpose* is implicated by its abundant use of data properties; hence, it is more alike a model for recording data than for representing the science of COVID-19 or SARS-CoV-2; e.g., in shorthand notation for domain and range axiom, $\text{heartrate} \sqsubseteq \text{Vital signs} \times \text{xsd : integer}$, which is alike Option B in Section 2.3. A substantial amount of information may be usable across countries trying to record data about patients. One class is specific to the country of its developers, India, which is the Mild and very mild COVID-19, which is one of the three categories mandated by its *government* rather than the modellers' *granularity bias*, which the ontology developers noted in the annotations of Patient.

3.2.3. COVoc

COVoc clearly states that its *purpose* is COVID-19 “scientific literature triage”. Knowledge organisation systems for literature annotation prioritise facilitating that process over ontological

precision or correctness. COVoc's contents are not clearly structured as a result, in the sense that there are many top-level terms and mixing of classes and instances. Some aspects, such as the use of the IAO and import of the RO, may indicate some leaning to the OBO Foundry stack as *philosophical bias* or a *mere-exposure cognitive bias*, since one of the COVoc authors is also an IAO contributor.

Its contents regarding cognitive bias raises several questions. One is of *granularity*, and perhaps also focus or time, which are straightforward omissions, such as listing only two continents, Asia and Europe, even though there are 4-7 (depending on how one categories). It is a mixture of omission or *granularity* and *politics* regarding the countries, since there are 10 subclasses of Country, of which two are disputed (Hong Kong and Taiwan) and one is definitely an error, since West Africa is not a country but a region on the African continent.

The low-hanging fruit for cognitive bias detection with *science* as source is Virus \sqsubseteq Organism, because a virus is not an organism no matter its repetition in popular media, i.e., this would be an *illusory truth effect bias*. Medically, the easy bias observation are that there are several disorders that are represented as subclasses of Disease, such as headache disorder \sqsubseteq Disease, whereas they are distinct medically, although some organisations would like to see certain disorders classified as a disease. If there were *economic* motivations, then the latter is a good candidate for bias source. One may be tempted to brush them off as mere modelling mistakes based on layperson commonsense assumptions and ran-out-of-time, but that is precisely where cognitive bias operates, and this is problematic especially for an ontology for scientific literature. Further scientific perspectives are built in by recording symptoms, such as Cough and Diarrhea as subclasses of phenotype, with phenotype defined as "The detectable outward manifestations of a specific genotype.". This a very gene-centric view on the body.

Gender is not present, but biological sex is. The only biological sex recorded in COVoc is male. Published literature on women and COVID-19 easily dates back to March 2020 (e.g., [38]), however, which is well before COVoc's development. Hence, it is, at best, an issue of *granularity* and possibly also *socio-cultural* since the gender bias in medical research is well-known [39].

Since many terms are plain science terms, like replicase polyprotein 1a (BtCoV) and cryogenic electron microscopy, there are no obvious language or linguistic issues in the sense of bias, other than an English language bias that nearly all extant ontologies have.

4. Discussion: Consequences of Bias in Ontologies

Having observed that there are indeed cognitive biases in ontologies, does it really matter? There are several ways where it can affect outcomes in information systems, with the three principle ones being due to *omissions*, *incorrect attributions*, and *undesirable deductions* that are logically correct but not ontologically or not according to the other bias.

Omissions and incorrect attributions have a direct effect on data analysis, since they increase the amount of noise (technically speaking) when the ontology is used for ontology-based data access and literature annotation and search. For instance, while mortality rates of men are higher for COVID-19, relatively more women get infected. If that cannot be annotated, since absent in COVoc, then searches on the emerging literature will obtain poorer query answers for possible causes as to why women are tested positive more often than men. Similarly, the lack of

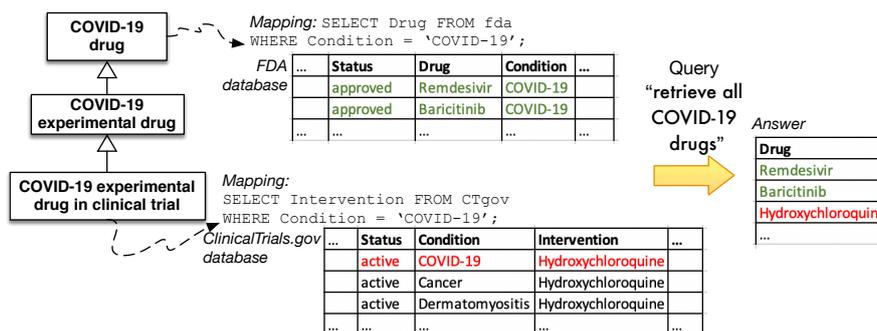


Figure 1: Ontology-based data access and integration scenario with CIDO and two database tables, from the ClinicalTrials.gov and FDA (selection shown, and mappings to the OWL classes are abbreviated). Retrieving hydroxychloroquine is the undesirable deduction from a scientific, medical, and regulatory standpoint.

the concept of household, or at least more family members, in CODO, prohibits finer-grained recording of the chain of infection and thus more likely to lose control of the spread of the virus.

Incorrect attributions happen when an annotator cannot find the desired knowledge in the ontology and then uses something else for it. For instance, if Ireland were to use CIDO, then the *walk-through* testing facility at Dublin Airport can be approximated by CIDO's drive-thru in the sense of passing by, or FDA authorised in the sense of being an official test location. More generally: annotators choose approximations based on different criteria, so any data analysis then will both miss instances and have false positives. Also, a presumptive positive annotation is, on the whole, an incorrect categorisation about 95% of the test cases (with the aim of $\leq 5\%$ test positivity rate) and it would seriously distort epidemiological investigations and overload tracking and tracing efforts if all presumptive positives had to be followed up. As long as an ontology does not fully characterise all required properties of an entity type so as to be sufficiently semantically precise, there is a heavier reliance on the mere term, which is an easier target open to multiple interpretations, image schemas, or bias.

An example of an *undesirable deduction* resulting from a cognitive bias built into an ontology can be readily seen with CIDO's experimental drug (recall Section 3.2.1). Take a data integration scenario with ontology-based data access, as illustrated in Fig. 1, where each class in the ontology is mapped to a query over the associated databases. A query over the ontology then avails of those mappings to retrieve the answer, together with the knowledge represented in the ontology. Hydroxychloroquine is still used as an experimental drug in COVID-19 clinical trials⁵, so then the query "retrieve all COVID-19 drugs" will include in the query answer hydroxychloroquine, since it recursively retrieves the instances down in the class hierarchy for all COVID-19 drug subclasses. It is definitely not a drug to treat COVID-19, however, nor has it been approved for that purpose in any country.

None of the COVID-19 ontologies have any meaningful deductions similar to the protein phosphatase experiment that deduced a novelty for human understanding of it [40], nor are

⁵There were 24 active trials with it for COVID-19 (of the 47 in total) https://clinicaltrials.gov/ct2/results?term=Hydroxychloroquine&Search=Apply&recrs=d&age_v=&gndr=&type=&rslt=; last accessed on 15-1-2021.

they aimed at achieving that at present. Conversely, on detecting inconsistencies, any issues with typical sources of problems would likely surface during ontology development already rather than at runtime in applications, since the reasoner will return an error or an undesirable deduction. For instance, if a developer adds that sentient beings are either human or other-animal and someone else wants to add plant, then this is caught during that development step already before deployment. Alternatively, a light-weight ontology language is used from the start so that such disagreements do not surface due to lack of language expressiveness, notably because of the absence of disjointness and qualified cardinality constraints. Therefore, our expectation is that the effects of bias with respect to reasoning consequences may be more salient in data management and retrieving information rather than in reasoning over the TBox.

5. Conclusions and Future Work

Bias may be present in an ontology, a number of which can be categorised as cognitive biases. Eight categories of sources of bias for ontologies were identified and illustrated: philosophical, purpose, science, granularity, linguistic, socio-cultural, political or religious, and economic motives. Four of them are explicit, and the other four may be either explicit or implicit. Three COVID-19 ontologies that were developed at the same time by different groups were assessed on these types of bias, which showed that each one exhibited a different subset of the sources of bias. This first characterisation and comparative assessment may contribute to further research into cognitive bias, and therewith also potential ethical aspects of ontologies, both regarding the modelling component and how it affects their use in applications.

The work presented in this paper aimed to offer initial steps to explore what bias in ontologies may amount to, in a more than anecdotal manner. Besides a sensitisation to the topic, it, perhaps, raised more questions than it offered answers. A relatively easy addition would be methodical and software support to note any explicit biases either as annotation in the ontology or in its documentation. A rigorous method involving interrogation of modelling choices during the ontology authoring stage may be beneficial, perhaps guided also by a relevant subset of Dimara et al's [11] list of cognitive biases. Another interesting avenue for future work is to disentangle cognitive bias from a case of innocuous ran-out-of-time and from a modelling mistake and attendant ontology quality issues that have other causes. Further ontological investigation into cognitive bias and its definition would also be useful, since a narrow definition with 'norm deviation' may not be operationalisable for ontologies outside science and engineering.

References

- [1] Gene Ontology Consortium, Gene Ontology: tool for the unification of biology, *Nature Genetics* 25 (2000) 25–29.
- [2] B. Motik, P. F. Patel-Schneider, B. Parsia, OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax, W3C Recommendation, W3C, 2009. <http://www.w3.org/TR/owl2-syntax/>.
- [3] N. Noy, Y. Gao, A. Jain, A. Narayanan, A. Patterson, J. Taylor, Industry-scale knowledge graphs: Lessons and challenges, *Queue* 17 (2019) 20:48–20:75.

- [4] K. Juel Vang, Ethics of Google's Knowledge Graph: some considerations, *Journal of Information, Communication and Ethics in Society* 11 (2013) 245–260.
- [5] M. Uschold, M. Gruninger, Ontologies: principles, methods and applications, *Knowledge Engineering Review* 11 (1996) 93–136.
- [6] D. L. Gomes, T. H. Bragato Barros, The bias in ontologies: An analysis of the foaf ontology, in: M. Lykke, T. Svarre, M. Skov, D. Martínez-Ávila (Eds.), *Proceedings of the Sixteenth International ISKO Conference*, Ergon-Verlag, 2020, pp. 236 – 244.
- [7] K. Janowicz, B. Yan, B. Regalia, R. Zhu, G. Mai, Debiasing knowledge graphs: Why female presidents are not like female popes, in: M. van Erp, M. Atre, V. Lopez, K. Srinivas, C. Fortuna (Eds.), *Proceeding of ISWC 2018 Posters & Demonstrations, Industry and Blue Sky Ideas Tracks*, volume 2180 of *CEUR-WS*, 2017.
- [8] C. M. Keet, Dirty wars, databases, and indices, *Peace & Conflict Review* 4 (2009) 75–78.
- [9] R. Mohanani, I. Salman, B. Turhan, P. Rodríguez, P. Ralph, Cognitive biases in software engineering: A systematic mapping study, *IEEE Transactions on Software Engineering* 46 (2020) 1318–1339.
- [10] T. A. Gavrilova, I. A. Leshcheva, Ontology design and individual cognitive peculiarities: A pilot study, *Expert Systems with Applications* 42 (2015) 3883–3892.
- [11] E. Dimara, S. Franconeri, C. Plaisant, A. Bezerianos, P. Dragicevic, A task-based taxonomy of cognitive biases for information visualization, *IEEE Transactions on Visualization and Computer Graphics* 26 (2020) 1413–1432.
- [12] S. Oreg, M. Bayazit, Prone to bias: Development of a bias taxonomy from an individual differences perspective, *Review of General Psychology* 13 (2009) 175–193.
- [13] C. M. Keet, A. Lawrynowicz, C. d'Amato, A. Kalousis, P. Nguyen, R. Palma, R. Stevens, M. Hilario, The data mining optimization ontology, *Web Semantics: Science, Services and Agents on the World Wide Web* 32 (2015) 43–53.
- [14] C. M. Keet, L. Khumalo, On the ontology of part-whole relations in Zulu language and culture, in: S. Borgo, P. Hitzler (Eds.), *10th International Conference on Formal Ontology in Information Systems 2018 (FOIS'18)*, volume 306 of *FAIA*, IOS Press, 2018, pp. 225–238. 17-21 September, 2018, Cape Town, South Africa.
- [15] C. M. Keet, An introduction to ontology engineering, volume 20 of *Computing*, College Publications, UK, 2018. 334p.
- [16] R. Arp, B. Smith, A. D. Spear, *Building Ontologies with Basic Formal Ontology*, The MIT Press, USA, 2015.
- [17] C. Masolo, S. Borgo, A. Gangemi, N. Guarino, A. Oltramari, *Ontology library*, WonderWeb Deliverable D18 (ver. 1.0, 31-12-2003)., 2003. [Http://wonderweb.semanticweb.org](http://wonderweb.semanticweb.org).
- [18] R. Mizoguchi, YAMATO: Yet Another More Advanced Top-level Ontology, in: *Proceedings of the Sixth Australasian Ontology Workshop, Conferences in Research and Practice in Information, CRPIT*, 2010, pp. 1–16. Sydney : ACS.
- [19] B. Smith, W. Ceusters, B. Klagges, J. Köhler, A. Kumar, J. Lomax, C. Mungall, F. Neuhaus, A. L. Rector, C. Rosse, Relations in biomedical ontologies, *Genome Biology* 6 (2005) R46.
- [20] Z. Khan, C. M. Keet, ONSET: Automated foundational ontology selection and explanation, in: A. ten Teije, et al. (Eds.), *18th International Conference on Knowledge Engineering and Knowledge Management (EKAW'12)*, volume 7603 of *LNAI*, Springer, 2012, pp. 237–251. Oct 8-12, Galway, Ireland.

- [21] C. Partridge, A. Mitchell, A. Cook, D. Leal, J. Sullivan, M. West, A Survey of Top-Level Ontologies - to inform the ontological choices for a Foundation Data Model, Technical Report, The Construction Innovation Hub, Centre for Digital Built Britain, 2020.
- [22] G. H. Merrill, Ontological realism: Methodology or misdirection?, *Applied Ontology* 5 (2010) 79–108.
- [23] P. R. Fillottrani, C. M. Keet, Dimensions affecting representation styles in ontologies, in: *Proceedings of the 1st Iberoamerican conference on Knowledge Graphs and Semantic Web (KGSWC'19)*, volume 1029 of *CCIS*, Springer, 2019, pp. 186–200. 24–28 June 2019, Villa Clara, Cuba.
- [24] Y. He, H. Yu, E. Ong, Y. Wang, Y. Liu, A. Huffman, H. hui Huang, J. Beverley, A. Y. Lin, W. D. Duncan, S. Arabandi, J. Xie, J. Hur, X. Yang, L. Chen, G. S. Omenn, B. Athey, B. Smith, Cido: The community-based coronavirus infectious disease ontology, in: J. Hastings, F. Loebe (Eds.), *Proceedings of the International Conference on Biomedical Ontology (ICBO'20)*, volume 2807, CEUR-WS, 2020.
- [25] M. Dumontier, C. Baker, J. Baran, A. Callahan, L. Chepelev, J. Cruz-Toledo, N. Del Rio, G. Duck, L. Furlong, N. Keath, D. Klassen, J. McCusker, N. Queralt-Rosinach, M. Samwald, N. Villanueva-Rosales, M. Wilkinson, R. Hoehndorf, The semanticscience integrated ontology (SIO) for biomedical research and knowledge discovery, *Journal of Biomedical Semantics* 5 (2014) 14.
- [26] N. Veerasamy, M. Grobler, B. V. Solms, Building an ontology for cyberterrorism, in: E. Filiol, R. Erra (Eds.), *Proc. 11th European Conference on Information Warfare and Security*, Academic Publishing International, 2012, pp. 286–295.
- [27] J. McCrae, G. A. de Cea, P. Buitelaar, P. Cimiano, T. Declerck, A. Gómez-Pérez, J. Gracia, L. Hollink, E. Montiel-Ponsoda, D. Spohr, T. Wunner, The Lemon Cookbook, Technical Report, Monnet Project, 2012. www.lemon-model.net.
- [28] M. Hepp, Goodrelations: An ontology for describing products and services offers on the web, in: *Proceedings of the International Conference on Knowledge Engineering and Knowledge Management (EKAW'08)*, volume 5268 of *LNCS*, Springer, 2008, pp. 332–347.
- [29] A. Lundin, M. Hallgren, M. Forsman, Y. Forsell, Comparison of DSM-5 classifications of alcohol use disorders with those of DSM-IV, DSM-III-R, and ICD-10 in a general population sample in sweden, *J Stud Alcohol Drugs* 76 (2015) 773–780.
- [30] J. C. Wakefield, DSM-5 substance use disorder: How conceptual missteps weakened the foundations of the addictive disorders field, *Acta Psychiatrica Scandinavica* 132 (2015) 327–334.
- [31] R. Jindal, K. Seeja, S. Jain, Construction of domain ontology utilizing formal concept analysis and social media analytics, *International Journal of Cognitive Computing in Engineering* 1 (2020) 62 – 69.
- [32] TOS Obesity as a Disease Writing Group, et al., Obesity as a disease: A white paper on evidence and arguments commissioned by the council of the obesity society, *Obesity* 16 (2008) 1161–1177.
- [33] M. A. Haendel, J. A. McMurry, R. Relevo, C. J. Mungall, P. N. Robinson, C. G. Chute, A census of disease ontologies, *Annual Review of Biomedical Data Science* 1 (2018) 305–331.
- [34] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. Goldberg, K. Eilbeck, A. Ireland, C. Mungall, T. OBI Consortium, N. Leontis, A. Rocca-Serra, A. Ruttenberg,

- S.-A. Sansone, M. Shah, P. Whetzel, S. Lewis, The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration, *Nature Biotechnology* 25 (2007) 1251–1255.
- [35] B. Dutta, M. DeBellis, CODO: an ontology for collection and analysis of COVID-19 data, in: *Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2020)*, INSTICC, 2020.
- [36] Z. M. Pendlington, P. Roncaglia, N. Matentzoglou, D. Osumi-Sutherland, D. Caucheteur, J. Gobeill, L. Mottin, D. Agosti, P. Ruch, H. Parkinson, COVoc: a COVID-19 ontology to support literature triage, 2020. URL: https://raw.githubusercontent.com/CIDO-ontology/WCO/master/day-1/Zoe_COVoc.pdf, wCO-2020: Workshop on COVID-19 Ontologies.
- [37] A. Parker, J. de Kadt, Household characteristics in relation to COVID-19 risks in Gauteng, 2020. URL: <https://gcro.ac.za/data-gallery/interactive-data-visualisations/detail/household-characteristics-relation-covid-19-risks-gauteng/>.
- [38] N. Li, L. Han, M. Peng, Y. Lv, Y. Ouyang, K. Liu, L. Yue, Q. Li, G. Sun, L. Chen, L. Yang, Maternal and Neonatal Outcomes of Pregnant Women With Coronavirus Disease 2019 (COVID-19) Pneumonia: A Case-Control Study, *Clinical Infectious Diseases* 71 (2020) 2035–2041.
- [39] A. Holdcroft, Gender bias in research: how does it affect evidence based medicine?, *Journal of the Royal Society of Medicine* 100 (2007) 2–3.
- [40] K. Wolstencroft, R. Stevens, V. Haarslev, Applying OWL reasoning to genomic data, in: C. Baker, H. Cheung (Eds.), *Semantic Web: revolutionizing knowledge discovery in the life sciences*, Springer: New York, 2007, pp. 225–248.