SOURCES OF BIAS IN ONTOLOGIES AND SOME CONSEQUENCES IN ONTOLOGY-DRIVEN INFORMATION SYSTEMS

C. Maria Keet
Department of Computer Science
University of Cape Town, South Africa

ONTOBRAS 2022 Keynote
Brazil (online), 24 November 2022

MUSHROOMING OF “AI ETHICS INITIATIVES”

AI Ethics

- Multidisciplinary field of study
- Main goal: how to optimize AI’s beneficial impact while reducing risks and adverse outcomes

Tech solutions: How to design and build AI systems that are aware of the values and commitments that are important in the deployment scenarios

- Social-technical approach: To identify, study, and propose technical and nontechnical solutions for ethics that lead to socially acceptable use of AI in life and society

AI Ethics in practice

Research
- Fairness
- Explainability
- Interpretability
- Robustness
- Privacy
- Value alignment

AI companies
- Governance
- Internal processes
- Tools
- Risk assessment
- Training

Standard bodies
- EU AI Act
- OECD AI Principles
- ISO/IEC 29134
- NIST 
- Indian Government

Educational institutions
- Bachelor’s programs
- Master’s programs
- Doctoral programs
- Short courses

Governments
- Risk-based approach
- Four levels of risk
- Focus on AI systems
- Obligations for high-risk applications, providers and users
WHAT ABOUT ONTOLOGIES AND KNOWLEDGE GRAPHS?

- Semantics to the rescue [ReyeroLobo22]
- Very little investigation into it
- Data aggregation for the Dirty War index [Keet09]
- Exploratory notes on knowledge graphs [Janowicz17]
- Friend of a friend vocabulary [Gomes20]
- “Gaps” in clinical terminologies [Geller21]
- Surveillance example, OE processes [Paparidis21]
- Exploration of cognitive bias [Keet21]

Google’s Knowledge Graph

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. "to some degree contests the autonomy of the user" [Vang13]
WHAT ABOUT ONTOLOGIES AND KNOWLEDGE GRAPHS?

- Google's Knowledge Graph!
  - 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.
  - "to some degree contests the autonomy of the user" [Vang13]

MORE TERMS…

- Thomas Eiter
  - Computer researcher
- Asunción Gómez Pérez
  - Spanish computer scientist
- Deborah L. McGuinness
  - American computer scientist
- Ian Horrocks
  - Computer science researcher

(FAST'S GRAPH ISN'T WORKING MUCH BETTER)

- Franz Baader
  - German computer scientist
  - Wikipedia, Citations, Bibtex

• It gets data off Wikipedia – no wiki page, no infobox
• It somehow infers different people to be the same, with
  - Franz to be dead and alive
  - Deborah to be a computer scientist and related to Bootleg Blondie

Annotation and Retrieval – Google's Knowledge Graph Mess

Brazilian computer scientists…

- Fernanda Fannielli, Tiago Prince Sales, and Mara Abel have no info box, nor did most other organisers, nor the session chairs, as is the case with many other scientists not based in the Anglosphere or close to it…

… For sure we won't all be easily findable with a simple query … nor do we have the authority to fix this mess … nor do we know all the terms to determine whether there is a bias in labeling.
AIMS

• Contribute to systematising the sort of bias that may be present in ontologies and similar artefacts
• Provide a preliminary answer to what bias means for ontologies, what their sources are
• How that manifests itself in ontologies
• Assess it across a set of ontologies in the same domain
• Pointers to avenues for more research

PRINCIPAL SOURCES OF BIAS IN ONTOLOGIES

THIS TALK

• Identify and discuss sources of bias
• Cherry-picked examples and more substantive comparison with three COVID-19 ontologies
• Bias in the light of automated reasoning and applications

PRELIMINARIES

• Defining cognitive bias... and differentiate from cognitive styles, alternate perspectives, image schemas, simple mistakes
• Inclusive definition for bias is adopted:
  • "a consequence of interference with honest attempts" [Oreg, 2009]
• Variants: narrow scope of norm deviation and error
• Implicit vs explicit
• Bias wrt effect, source, output vs process, antecedents vs effects of (cognitive) bias
• Types of bias
• (would benefit from a proper ontological investigation...)


Sanfilippo, E.M. et al. (Eds.). CEUR-WS vol. 2969. 17p


Image by planetschwa is licensed under CC BY-NC-ND 2.0
PRELIMINARIES

• For IT and computing, grouped by dimension, e.g.,
  • by type of task for information visualisation [Dimara20]
  • by software engineering “knowledge area” [Mohanani20]

<table>
<thead>
<tr>
<th>Bias</th>
<th>Antecedents</th>
<th>Knowledge area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anchoring and Adjustment</td>
<td>Reusing previously written generic, difficult to identify influential points (anchor) [27]</td>
<td>Construction, Design, Construction</td>
</tr>
<tr>
<td></td>
<td>Knowing, decision, and information from past knowledge [19]</td>
<td>Requirements, Construction, Management</td>
</tr>
<tr>
<td></td>
<td>Not being able to adjust to the new environment [19]</td>
<td>Design</td>
</tr>
<tr>
<td></td>
<td>Uncertainty of future actions, lack of business / functional knowledge and inflexible clients [19]</td>
<td>Design</td>
</tr>
<tr>
<td></td>
<td>Confirmation and availability bias during design [19]</td>
<td>Design</td>
</tr>
</tbody>
</table>

SAMPLING OF COGNITIVE BIASES FROM DIMARA ET AL.’S LIST

• 17 of the 154 seem potentially applicable to ontologies; among others:
  • Mere exposure/familiarity: choice is influenced by exposure to it and thus familiarity with it.
  • Naïve 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.
  • Distinction bias: Choices affected by how many are the alternatives (?)
  • Barnum effect: High accuracy ratings for vague and general statements. [9]

POSSIBLE BIASES, BY SOURCE

<table>
<thead>
<tr>
<th>Type</th>
<th>Subtype</th>
<th>[im/ex]plicit bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Philosophical</td>
<td>-</td>
<td>explicit</td>
</tr>
<tr>
<td>Purpose</td>
<td>-</td>
<td>explicit</td>
</tr>
<tr>
<td>Subject domain</td>
<td>Science, Granularity, Linguistic, Socio-cultural, Political or religious, Economics</td>
<td>explicit, either, either, either, explicit</td>
</tr>
</tbody>
</table>

FOUNDATIONAL ONTOLOGY DIFFERENCES (PHILOSOPHICAL OR OTHERWISE)

• Realism vs idealism, concepts, universals etc.
• Some differences don’t matter in praxis; some do
• Ways to find and resolve the (explicit/implicit) conflict(s)
• Example: BFO’s realism does not accept abstract entities
**PURPOSE: ENCODING BIAS, CF. A COGNITIVE BIAS**

- (a) a class transformation or (b) a relationship transforms to?
- (a) Enzyme is a Protein or (b) Enzyme has bearer Protein?
- (a) Mouse hasColour Colour or (b) Mouse hasQuality Colour that hasQuality Physical Region?

Three different patterns with a purpose bias:
- Pattern A: biased toward a scientific approach
- Pattern B: conceptual data modelling influence or purpose
- Pattern C: a thesaurus-like approach useful for document annotation

**ENCODING BIAS ISSUES**

Different representation decisions in different ontologies impedes:
- ontology interoperability and alignment
- ontology reuse
- ontology development: in automation and, e.g., CQ translations
- deployment in ontology-driven information systems
**CONFLICTING MODELING STYLES ILLUSTRATION**

- **Requirement**: align/integrate/merge the COVID-19 ontologies
- **CIDO ontology + CODO ontology**
  - CODO: "Laboratory test finding" = {positive, pending, negative}
  - CIDO: "COVID-19 diagnosis" as a class, with three subclasses (negative/positive/presumptive positive) COVID-19 diagnosis
- **Issue**: class vs. instance representations of the same idea
- **Solution options**:
  1. Change CODO
  2. Change CIDO
  3. Outside option [e.g.: approach with attribute+values]


---

**GRANULARITY**

- How detailed should it be? How many levels in the taxonomy? Unbalanced tree ok?
- Less details...
  - Act of omission
  - E.g.: aggregating ex-military persons with non-involved persons (civilians) as one group of "Civilians"
  - Not needed
  - Ran out of time
  - ... but may be very difficult to determine, unless stated
- Not needed and ran out of time may be explicit choices and prioritization (explicit bias), or honestly out of scope for v1 or ever

---

Water purification
Sewerage
Electricity

**ARITY**

- Telephone lines
- Cellphone towers
- FTTH switches
- Authentication
- Malware

---

**Section of the cyber terrorism ontology**

**GRANULARITY**

<table>
<thead>
<tr>
<th>Category</th>
<th>Terminology</th>
<th>ANLROD-CT</th>
<th>RCDD</th>
<th>CPT</th>
<th>MCH</th>
<th>MedIRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthcare Institution</td>
<td>Retention of examination</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>10005840</td>
</tr>
<tr>
<td></td>
<td>Other referral to specialist</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>10001582</td>
</tr>
<tr>
<td></td>
<td>Unrelated chronic condition/valid Examination</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>10001115</td>
</tr>
<tr>
<td></td>
<td>Nutrition counseling</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>10001115</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2170000175105</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>10001115</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6400000175100</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>10001115</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10001115</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>10001115</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10001115</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>10001115</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10001115</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>10001115</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10001115</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>10001115</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10001115</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>10001115</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10001115</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>10001115</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10001115</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>10001115</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10001115</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>10001115</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10001115</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>10001115</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10001115</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>10001115</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10001115</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>10001115</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10001115</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>10001115</td>
</tr>
</tbody>
</table>

Source: [Geller21]
SOCIO-CULTURAL ISSUES IN ONTOLOGIES

• Relates to how society is organised, the assumptions that underlie it and history
  • organisational structures, who lives with whom, demographics, allocation of resources, or social geography
• Example: GoodRelations Ontology’s payment methods and legal status of a Business [Hepp08]

POLITICAL MOTIVATIONS

• The easy one: Aggrieved group vs Terrorist organisation
• Concretely,
  • terrorist and terroristgroup in the terrorism ontology [Jindal20]
  • ActorEntity with various types of Insiders and Protestors in the Cyberterrorism ontology [Veerasamy12]

• Alcohol use and disorder across time and cultures
  • DSM-IV, DSM-V, and ICD-10 coding differences
  • modifications due to a combination of socio-cultural factors and some scientific disagreement [Wakefield15]
DO THESE TYPES OF BIASES APPEAR IN ONTOLOGIES IN THE SAME DOMAIN?

BIAS BY COGNITIVE BIAS

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

COVID-19 ONTOLOGIES

- "same" topic on COVID-19, developed at same time by different groups:
  - Coronavirus Infectious Disease Ontology (CIDO) [He20]
  - COVID-19 Ontology (CODO) [Dutta20]
  - Coronavirus Vocabulary (COVoc) [Pendlington20]
- Assess their documentation, characteristics, content
- Iteratively note observations of bias and check subset of cognitive bias list and consider wrt the ontologies

PRESENCE/ABSENCE, BY SOURCE

<table>
<thead>
<tr>
<th>Bias (Source/type)</th>
<th>CIDO</th>
<th>CODO</th>
<th>COVoc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Philosophical</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Purpose</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Science</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Granularity</td>
<td>±</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Linguistic</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Socio-cultural</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Political or religious</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Economics</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Experimental factor</td>
<td>+</td>
<td>±</td>
<td>+</td>
</tr>
<tr>
<td>Close contact</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Gathering</td>
<td>±</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Infected Case- Patient</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Infected Case- Worker</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Infected Family- Member</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Infected Spouse</td>
<td>±</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
AUTOMATED REASONING

ANY EFFECTS ON AUTOMATED REASONING?

- ‘Incoherence’ (one or more unsatisfiable classes), inconsistencies, or undesirable deductions
- For TBox only: incoherence and undesirable deductions will be found at authoring time already, not during deployment
- For the knowledge base (TBox + ABox [in owl or secondary storage]): inconsistencies or undesirable deductions either at authoring time or during deployment
  - Querying data
  - Annotating data (and subsequent retrieval)

DOMAIN ONTOLOGY EXAMPLE: DESIRABLE OR UNDESIRABLE DEDUCTION

DOMAIN ONTOLOGY EXAMPLE: DESIRABLE OR UNDESIRABLE DEDUCTION


EXAMPLE: OBDA DATA INTEGRATION EXAMPLE WITH THE CIDO

Query "retrieves all COVID-19 drugs"
ENCODING BIAS EXAMPLE: QUERING CELL TYPES IN THE CL ONTOLOGY

Comparing alternative encodings:

The "CL way"  

\[ \text{an IC} \subseteq \exists \text{capable of, a process} \]

versus

Kupffer cell \( \subseteq \exists \text{clears, erythrocyte} \)
acid secreting cell \( \subseteq \exists \text{secretes, acid} \)
Be1 cell \( \subseteq \exists \text{produces, interferon-gamma} \)
type I NK T cell \( \subseteq \exists \text{binds, glycosphingolipid} \)

vomeronasal organ \( \subseteq \exists \text{detects, pheromone} \)

....

Example of an "applied" way

ENCODING BIAS EXAMPLE: QUERING CELL TYPES IN THE CL ONTOLOGY

Comparing alternative encodings:

The "applied way"

• Cell Annotation Platform (Celltype.info) only allows search by cell type to find data sets: no effects
• Others that use CL also have mostly only simple term search
• But, if you were to want to search for, e.g.:
  • the function/role of a cell or
  • all cells that perform a certain role
  • all datasets that are from experiments on erythrocyte clearance by macrophages
  • projects on capabilities of some cell type
then the "capable of" version will be much easier cf. over very many possible names for relations: fewer names to remember, more structured, consistency in approach

 cardboard  

ENCODING BIAS EXAMPLE: QUERING CELL TYPES IN THE CL ONTOLOGY

Comparing alternative encodings:

The "applied way"

• Cell Annotation Platform (Celltype.info) only allows search by cell type to find data sets: no effects
• Others that use CL also have mostly only simple term search
• But, if you were to want to search for, e.g.:
  • the function/role of a cell or
  • all cells that perform a certain role
  • all datasets that are from experiments on erythrocyte clearance by macrophages
  • projects on capabilities of some cell type
then the "capable of" version will be much easier cf. over very many possible names for relations: fewer names to remember, more structured, consistency in approach
REFERENCES


