Towards Ontologies in Variation

Torsten Hahmann$\triangle$
Sheila McIlraith$^+$

$\triangle$ National Center for Geographic Information & Analysis, School of Computing and Information Science, University of Maine, USA

$^+$ Department of Computer Science, University of Toronto

KRR 2015
Real Humans

Textbook Human vs. Real Humans
Variability in Human Anatomy

Existing atlases, textbooks, and digital models of human anatomy:

- Excellent job at standardization of nomenclature
- Projects such as Virtual Human/Soldier (virtualsoldier.us) created a digital image data set of a complete human

- All represent a canonical male or female human
- No sense of what is “normal”, what deviates from “normal”, and the general variability in a population or in subpopulations (beyond male/female distinction)
  - No explicit definitions of canonicity and variability
- Much progress at statistically describing individual bones or organs, but not tied to standard medical nomenclature used by human experts
  - Statistical Shape Analysis (SSA)
Parametric Human Project: parametrichuman.org

Consortium led by UBC and Autodesk Research

“Our goal is to create an **advanced, data-driven, biomechanical, statistical, anatomical model** as a digital ergonomics and design tool to help users augment or accelerate the study of human abilities.”

- Better understand human anatomy: identify what is canonical and quantify the modes of variations in musculoskeletal human
Parametric Human Project:  
A Data-Driven Attempt to Capture Human Variability

Musculoskeletal Anatomy Lab, University of Toronto (Anne Agur) &  
Environment & Ergonomics Research Group, Autodesk Research (Azam Khan)
Ontologies in Variation

A KR scheme that can

- Represent natural kinds, objects, concepts – *Things* – in all their variation

- Support **explicit definitions** of canonicity and variability of *Things* in terms of other *Things*, properties and relations of meaning to humans

- Derive variability from data

- Express propositions about statistical properties of individuals, of groups of individuals, and of one or multiple datasets

- Support both data-driven (derive class characteristics from data) and concept-driven (classify data using a-priori definitions) representations

- Use different reference classes (e.g. female, American Indian, female between 30 and 40, ...) as needed and compute definitions for them

- Support scientific inquiry into data using a given terminology
Starting Points
which is derived from the Rectus femoris and Vastus intermedius. The medial and lateral borders are thinner and converge below; they give attachment to those portions of the Quadriceps femoris which are derived from the Vastus lateralis and medialis.

**Apex.**—The apex is pointed, and gives attachment to the ligamentum patellae.

**Structure.**—The patella consists of a nearly uniform dense cancellous tissue covered by a thin compact lamina. The cancelli immediately beneath the anterior surface are arranged parallel with it. In the rest of the bone they radiate from the articular surface toward the other parts of the bone.

**Ossification.**—The patella is ossified from a single center, which usually makes its appearance in the second or third year, but may be delayed until the sixth year. More rarely, the bone is developed by two centers, placed side by side. Ossification is completed about the age of puberty.

**Articulation.**—The patella articulates with the femur.

### The Tibia (Shin Bone)

The tibia (Figs. 258, 259) is situated at the medial side of the leg, and, excepting the femur, is the longest bone of the skeleton. It is prismatic in form, expanded above, where it passes into the knee-joint, contracted in the lower third, and again enlarged but to a lesser extent below. In the male, its direction is vertical, and parallel with the bone of the opposite side; but in the female it has a slightly oblique direction downward and laterally, to compensate for the greater obliquity of the femur. It has a body and two extremities.

#### The Upper Extremity (proximal extremity)

The upper extremity is large, and expanded into two eminences, the medial and lateral condyles. The superior articular surface presents two smooth articular facets (Fig. 257). The medial facet, oval in shape, is slightly concave from side to side, and from before backward. The lateral, nearly circular, is concave from side to side, but slightly convex from before backward, especially at its posterior part, where it is prolonged on to the posterior surface for a short distance. The central portions of these facets articulate with the condyles of the femur, while their peripheral portions support the menisci of the knee-joint, which here intervene between the two bones. Between the articular facets, but nearer the posterior than the anterior aspect of the bone, is the intercondylar eminence (spine of tibia), surmounted on either side by a prominent tubercle, on to the sides of which the articular facets are prolonged; in front of and behind the intercondylar eminence are rough depressions for the attachment of the anterior and posterior cruciate ligaments and the menisci.

The anterior surfaces of the condyles are continuous with one another, forming a large somewhat flattened area; this area is triangular, broad above, and perforated by large vascular foramina; narrow below where it ends in a large oblong foramen for the nutrient artery of the tibia, which gives attachment to the ligamentum patellae; a bursa intervenes between the deep surface of the ligament and the part of the bone immediately above the tuberosity. Posteriorly, the condyles are separated from each other by a shallow depression, the posterior intercondylar fossa, which gives attachment to part of the posterior cruciate ligament of the knee-joint. The medial condyle presents a small pit on its posterior surface, for the insertion of the thickened part of the ligamentum patellae. The lateral condyle presents near its posterior surface a sharp ridge, the posterior margin of which serves as the upper margin of the intercondylar fossa.
The Body or Shaft (*corpus tibiae*).—The body has three borders and three surfaces.

Borders.—The **anterior crest or border**, the most prominent of the three, commences above at the tuberosity, and ends below at the anterior margin of the medial malleolus. It is sinuous and prominent in the upper two-thirds of its extent, but smooth and rounded below; it gives attachment to the deep fascia of the leg.

The **medial border** is smooth and rounded above and below, but more prominent in the center; it begins at the back part of the medial condyle, and ends at the posterior border of the medial malleolus; its upper part gives attachment to the tibial collateral ligament of the knee-joint to the extent of about 5 cm., and insertion to some fibers of the Popliteus; from its middle third some fibers of the Soleus and Flexor digitorum longus take origin.
Surfaces.—The medial surface is smooth, convex, and broader above than below; its upper third, directed forward and medialward, is covered by the aponeurosis derived from the tendon of the Sartorius, and by the tendons of the Gracilis and Semitendinosus, all of which are inserted nearly as far forward as the anterior crest; in the rest of its extent it is subcutaneous.

The lateral surface is narrower than the medial; its upper two-thirds present a shallow groove for the origin of the Tibialis anterior; its lower third is smooth, convex, curves gradually forward to the anterior aspect of the bone, and is covered by the tendons of the Tibialis anterior, Extensor hallucis longus, and Extensor digitorum longus, arranged in this order from the medial side.
<table>
<thead>
<tr>
<th>Code</th>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A02.5.006.001</td>
<td>Tibia</td>
<td>Superior articular surface</td>
</tr>
<tr>
<td>A02.5.006.002</td>
<td>Facies articularis superior</td>
<td>Medial condyle</td>
</tr>
<tr>
<td>A02.5.006.003</td>
<td>Condylus medialis</td>
<td>Lateral condyle</td>
</tr>
<tr>
<td>A02.5.006.004</td>
<td>Condylus lateralis</td>
<td>Fibular articular facet</td>
</tr>
<tr>
<td>A02.5.006.005</td>
<td>Facies articularis fibularis</td>
<td>Anterior intercondylar area</td>
</tr>
<tr>
<td>A02.5.006.006</td>
<td>Area intercondylaris anterior</td>
<td>Posterior intercondylar area</td>
</tr>
<tr>
<td>A02.5.006.007</td>
<td>Area intercondylaris posterior</td>
<td>Intercondylar eminence</td>
</tr>
<tr>
<td>A02.5.006.008</td>
<td>Eminencia intercondylaris</td>
<td>Medial intercondylar tubercle</td>
</tr>
<tr>
<td>A02.5.006.009</td>
<td>Tuberculum intercondylare mediale</td>
<td>Lateral intercondylar tubercle</td>
</tr>
<tr>
<td>A02.5.006.010</td>
<td>Tuberculum intercondylare laterale</td>
<td>Shaft: Body</td>
</tr>
<tr>
<td>A02.5.006.011</td>
<td>Corpus tibiae</td>
<td>Tibial tuberosity</td>
</tr>
<tr>
<td>A02.5.006.012</td>
<td>Tuberositas tibiae</td>
<td>Medial surface</td>
</tr>
<tr>
<td>A02.5.006.013</td>
<td>Facies medialis</td>
<td>Posterior surface</td>
</tr>
<tr>
<td>A02.5.006.014</td>
<td>Facies posterior</td>
<td>Soleal line</td>
</tr>
<tr>
<td>A02.5.006.015</td>
<td>Linea musculi solei</td>
<td>Lateral surface</td>
</tr>
<tr>
<td>A02.5.006.016</td>
<td>Facies lateralis</td>
<td>Anterior border</td>
</tr>
<tr>
<td>A02.5.006.017</td>
<td>Margo anterior</td>
<td>Medial border</td>
</tr>
<tr>
<td>A02.5.006.018</td>
<td>Margo medialis</td>
<td>Interosseus border</td>
</tr>
<tr>
<td>A02.5.006.019</td>
<td>Margo interosseus</td>
<td>Medial malleolus</td>
</tr>
<tr>
<td>A02.5.006.020</td>
<td>Malleolus medialis</td>
<td>Malleolar groove</td>
</tr>
<tr>
<td>A02.5.006.021</td>
<td>Sulcus malleolaris</td>
<td>Articular facet</td>
</tr>
<tr>
<td>A02.5.006.022</td>
<td>Facies articularis malleoli medialis</td>
<td>Fibular notch</td>
</tr>
<tr>
<td>A02.5.006.023</td>
<td>Incisura fibularis</td>
<td>Inferior articular surface</td>
</tr>
<tr>
<td>A02.5.006.024</td>
<td>Facies articularis inferior</td>
<td>Fibula</td>
</tr>
<tr>
<td>A02.5.007.001</td>
<td>Tibia</td>
<td></td>
</tr>
</tbody>
</table>
Existing Ontological Representations of Anatomy

**Foundational Model of Anatomy (FMA)**
- developed by Rosse & Mejino, Univ. of Washington since 1995
- provides **taxonomy of anatomical structures** similar to TA
- provides terminology for the **classes of Things** we want to characterize, but no definitions yet
- **other anatomical ontologies could also serve as starting point**
Surface Segmentation of a Human Tibia

Based on named anatomical Things (primarily surfaces), relationships between them, and their properties as captured in the Adjacency Graph (Mogk et al. 2014)
Based on manual inspection of 2,000 adult tibia

“The most striking peculiarity of the normal tibia is its variability in shape. The bone is hardly ever exactly alike in any two skeletons, and it will occasionally differ markedly in the same body.”

“. . . the shapes may be reduced to six principal groups or types.”

“The shape of the shaft most frequent in both the white male and the female is that of a prism. About 3/5 of all tibiae are of this variety, [. . . ]”

“The next modification of the shaft is characterized by a pronounced concavity involving the upper 2/3 of the external surface of the tibia. . . . This character is frequent in the tibiae of the American Indians.”

“The average male tibia is distinctively longer than that of the female. The average length of the male tibia on the right was 36.45 cm; on the left 36.48 cm. In females the right tibia measured on an average 34.5 cm; the left 34.6 cm.” The scale of length in the male varies from 31.0 cm to 45.5 cm, and in the female from 28.0 to 39.0 cm. In the male the right tibia was found longer in 42.8% of cases; the left longer in 25%; equal length in 28.6% . . . ”
Tool: Symbolic-Probabilistic Definitions

Describe canonicity: What do humans share anatomically?
- Start with describing **classes of bones**
- Use standard anatomical terminology, as reflected by the terms hierarchically organized in the TA
  - **Features**: parts, surface areas, borders, points

⇒ Symbolic KR?

Describe variability: How do humans vary anatomically?
- What component Things may be missing or duplicated?
- How frequent are features in certain **relations**
  - **Relations**: adjacent, bounds, overlaps, anterior of
- How are **property values** of features statistically distributed?
  - **Properties**: convexity, smoothness, length, width

⇒ Probabilistic KR?

More general question: Can we support such a scenario by combining symbolic and probabilistic KR such that we play to the strengths of both?
A More General Problem
Ontologies in Variation

A KR scheme that can

- Represent natural kinds, objects, concepts – *Things* – in all their variation
- Support **explicit definitions** of canonicity and variability of *Things* in terms of other *Things*, properties and relations of meaning to humans
- Derive variability from data
- Express propositions about statistical properties of individuals, of groups of individuals, and of one or multiple datasets
- Support both data-driven (derive class characteristics from data) and concept-driven (classify data using a-priori definitions) representations
- Use different reference classes (e.g. female, American Indian, female between 30 and 40, . . . ) as needed and compute definitions for them
- Support scientific inquiry into data using a given terminology
General Problem Description

- **Input:**
  - An **ontology** of structural knowledge about **Things**
  - One or multiple **datasets** consisting of **datapoints (facts)** about observed/measured **Things** (ground facts)

- **Desired Output:** A concise description of the classes of Things in terms of other Things, relationships, and properties
  - **logical assertions:** every adult tibia has exactly one medial surface, which is always convex and adjacent to its lateral surface
  - **statistical assertions about a specific dataset:** the medial surface of an adult tibia has an average size of 28 $cm^2$ (in the range of 19 $cm^2$ and 45 $cm^2$) and has an average principal curvature of (0.5, 0.1)

= Ontologies in Variation

- What approach to use? **We argue that it requires marrying logical and statistical knowledge representation**
Why Not Just Logical AI?

Weaknesses:

- Ill-suited to capture variability (explicit enumeration of cases or non-monotonic reasoning necessary)
- Underspecified: capture necessary, but rarely sufficient conditions
- Require labour-intensive hand-crafting
- Brittle: one wrong assertion breaks the ontology

Strengths:

- Highly expressible
- Can build on and use existing human-comprehensible terminology
- Results will have direct meaning to humans: helps understand classes, not just classify Things
Why Not Just Statistical AI?

**Weaknesses:**
- Cannot easily reuse existing terminology and structural knowledge
- Comprehensibility: not easy to make sense of trained classifiers
- Lack of insight and explanation
  - Why is a certain bone classified as a tibia?
  - Why is a certain bone not classified as a tibia?
- Where do probabilities come from?
- How to manually correct/modify representation?

**Strengths:**
- Robust in the presence of minor errors or deviations
- Adaptable to new data (can be trained and easily evolved)
- Logical reasoning can be treated as a special case
Marrying Logical and Statistical AI: Idea

How Can They Complement Each Other?

- Logic as basic structure ("skeleton"), data-driven statistics refine

- Relies on a frequentist view of probabilities (statistics about the datasets) as opposed to the more common interpretation of probabilities as degrees of belief
  - Does not require us to specify probabilities a-priori
  - Similar in some aspects to LP (Bacchus 1990)
  - But no attempt to add probabilistic semantic to a logic

- Instead, we propose to embed statistical expressions about datasets in classical logical expressions
  - Statistical expressions are evaluated to true/false based on datasets
  - No probabilistic inference necessary or supported
  - Classical logical inferencing outside of statistical expressions
Marrying Logical and Statistical AI: More Related Work

- **Halpern (e.g., Halpern 1990)**
  - Explicitly considers two semantics for probabilities: statistical (“90% of birds fly”) and degrees of belief (“We are 90% certain that Tweety flies”) in a logic formalization of probabilistic reasoning.

- **Markov Logic Networks (e.g, Domingos et al. 2006)**
  - Classical logical formulas are assigned a probability indicating the certainty with which the formula holds true.
  - Probabilistic inferencing necessary.
  - Requires certainties to be given a-priori.
  - Similarly, many probabilistic extensions to description logics.

- **Semantic Science (e.g., Poole et al. 2008)**
  - Extends an ontology with a statistical/probabilistic representation.
  - Based on datasets, but not clear how certainties are derived.
  - Ontology only provides terminology and classes are rigidly defined.
  - Probabilities assigned to relational expressions indicate certainty.
  - Probabilistic inferencing necessary.
Outline of a Solution
Straw Proposal

Essentially a classical first-order language, extended by queries about datasets and statistical aggregations of query results (similar to those supported by SQL):

(1) Dataset and measurement variables and constants
Straw Proposal

Essentially a classical first-order language, extended by queries about datasets and statistical aggregations of query results (similar to those supported by SQL):

1. Dataset and measurement variables and constants

2. Extract collections of tuples from a dataset $d$:
   
   members($d$, tibia)
   
   rel-instances($d$, adjacent)
   
   prop-instances($d$, length)
   
   rel-instances(filter($d$, adjacent, [1], [medial-surface]))
   
   prop-instances(filter($d$, length, [1], [tibia]))
   
   prop-instances(value($d$, length, > 20cm))
Straw Proposal

(3) Aggregate collections of tuples (counting): results in a property:

How many surfaces are adjacent to each medial surface?

\[
\text{members(aggregate(filter}(d, \text{adjacent}, [1], \text{medial-surface})), [1, 2]))
\]

How many sellers did a realtor represent?

\[
\text{aggregate(sale, [realtor])}
\]
(3) **Aggregate collections of tuples (counting):** results in a property:

*How many surfaces are adjacent to each medial surface?*

\[
\text{members(aggregate(filter}(d, adjacent, [1], [medial-surface]), [1, 2]))}
\]

*How many sellers did a realtor represent?*

\[
\text{aggregate(sale, [realtor])}
\]

(4) **Filter collections:**

*Which medial surfaces are adjacent to more than 3 surfaces?*

\[
\text{value(aggregate(filter}(d, adjacent, [1], [medial-surface]), [1, 2]), > 3)}
\]

*Which realtors represented at least 50 sellers?*

\[
\text{members(value(aggregate(sale, [realtor]), > 50))}
\]
Straw Proposal

(5) **Statistically evaluate collections of property tuples:** using standard statistical functions (mean, min, max, standard deviation), which result in a single constant that can be treated just as any other constant in the embedding logical language.

*What is the average length of a tibia?*

\[
\text{avg}(\text{filter}(d, \text{length}, [1], [\text{tibia}]))
\]

*What is the mean number of sellers a realtor represented?*

\[
\text{mean}(\text{aggregate}(\text{sale}, [\text{realtor}]))
\]

*Which realtors represented an above-average number of sellers?*

\[
\text{members}(\text{value}(\text{aggregate}(\text{sale}, [\text{realtor}]), \> \text{avg}(\text{aggregate}(\text{sale}, [\text{realtor}]))))
\]
Embedding these dataset expressions in logical sentences

Is the average length of femurs greater than the average length of tibias?

What percentage of humans have a femur that is longer than their tibia?

Do all people with a tibia that is above average in length also have a femur that is above average in length?

Is every femur longer than any tibia?

Is there a tibia that has more than 5 surfaces adjacent to the medial surface?

The average weight of a male in study 1 (from 2000) is greater than the average weight of a male in study 2 (from 1970).
Symbolic-Statistical Definitions

Can extract a definition of a tibia: A tibia is a long bone that has an average length of 36cm, varying from 29 to 41cm. It consists of a shaft with three surfaces, of which in 90% (of the observed humans) the medial surface is the largest width an average area of $28cm^2$, . . . . The medial surface meets the lateral surface at the anterior border, which is on average 17cm long, and . . . 75% of the observed human tibia’s have a clearly delineated “anterior triangular area”, . . .

- We can extract such definitions for customized reference classes: males over 45, Asian females between 30 and 40, those with a tibia that has certain specific criteria, etc.
- Such definitions can be dynamically computed and/or archived
- Such definitions are explicit: they make sense without having to refer back to the data
- Because such definitions are expressed in a logical language, we can use them for classical first-order inferencing
Symbolic-Statistical Definitions (contd.)

Equally, we can manually define classes using a mix of definite and statistical assertions: Want to identify all “gold-star realtors”, each of which has at least 50 sales during the last year AND has an above average number of repeat customers AND (has a median sales price 10% above average OR sold at least 5 house worth more than $2M) AND has NOT sold a single house below estimate...
Acknowledgements

- Collaboration with Autodesk Research, Toronto (Azam Khan, Jeremy Mogk, Jacky Bibliowicz) and Reza Samavi (McMaster Univ.)

- Discussions with partners of the Parametric Human Project; especially Anne Agur (Musculoskeletal Anatomy Lab, Univ. of Toronto)

- Funding provided by Natural Sciences and Engineering Research Council of Canada (NSERC)
Open Questions

- What do we gain or lose by restricting reasoning to logical interpretations and treating statistical expressions as a “blackbox” that operates strictly only on the datasets (like a statistical database)?

- Are there important queries/applications that are not expressible in this language that we have overseen?

Do you have problems or datasets for which such an approach might be useful? If so, we’d love to hear about them.
F. Bacchus.
LP – a logic for representing and reasoning with statistical knowledge.

P. Domingos, S. Kok, H. Poon, M. Richardson, and P. Singla.
Unifying logical and statistical AI.
In *AAAI-06*, 2006.

Towards generating bone-specific ontologies from an organic shape descriptor (extended abstract).

J. Halpern.
An analysis of first-order logics of probability.

A. Hrdlicka.
Study of the normal tibia.
J. Mogk, J. Bibliowicz, and A. Khan.
Towards the development of a biomechanical ontology to support the initiatives of the parametric human project.

J. Mogk, J. Bibliowicz, and A. Khan.
The organic shape descriptor: A computational technique for characterizing and annotating skeletal surface anatomy (extended abstract).

J. Mogk, J. Lloyd, J. Bibliowicz, A. Agur, S. Fels, and A. Khan.
The parametric human project: building a probabilistic atlas of human anatomy.

D. Poole, C. Smyth, and R. Sharma.
Semantic science: Ontologies, data and probabilistic theories.

C. Rosse and J. L. Mejino Jr.
A reference ontology for bioinformatics: the Foundational Model of Anatomy.