Ontology Dynamics

Vadim Ermolayev - a knowledge engineer

Intelligent Systems, Dept of IT,
Zaporizhzhya Natl Univ.
Ukraine

vadim@ermolayev.com
http://ermolayev.com/

SemData

FP7 Marie Curie IRSES project

Plan

- Not a narrow-focused approach/technique
- Instead, a broader view – motives …
  - Like "I have a dream …"
  - Shallow, but can go deeper if interested
- Looking at analogies
- Giving an example
  - Something done in my group
  - By people you've seen here
- Sketching out potential R&D problems
- Hopefully provoking questions

Kinematics and Dynamics

- **Mechanics:**
  - **Kinematics** studies the **motion** of objects without reference to its causes
  - **Dynamics** is concerned with the study of **forces** and **torques** and their **effect on motion**
  - **Motion**: **change** of position

- **Knowledge representation and management**:
  - What is **motion** for knowledge representations?
    - Also the **change** of … be detailed later
  - What are the **forces** and **torques**?
    - Also **effects on motion** …

System Dynamics

- **System Dynamics** studies the behaviour of (complex) systems over time
  - The behaviour of the entire system is affected by internal **feedback (causal) loops** and time delays

- **Knowledge representation and management**:
  - What is "**behavior**" wrt knowledge representations?
    - Also the **change** of …

Population Dynamics

- **Studies**:
  - Short- and long-term changes in the **size and age composition** of populations
    - E.g. ageing or population decline
  - Biological and environmental processes influencing those changes
- **Deals with** the way populations are **affected by**:
  - Birth and death rates
  - Immigration and emigration

- **Knowledge representation and management**:
  - What is **birth** and **death** in ontology populations?
  - How these populations **migrate**?

Kinematics OR Dynamics?

- What we do in ontology change and evolution, is it:
  - Dynamics?
  - or Kinematics?

Kinematics OR Dynamics?

- What we do in ontology change and evolution, is it:
  - Dynamics?
  - or Kinematics?

- A: Kinematics:
  - Study of and respond to changes/differences
  - Do NOT really analyze the causes and triggers of change

Kinematics OR Dynamics?

- What we do in ontology change and evolution, is it:
  - Dynamics?
  - or Kinematics?

- A: Kinematics:
  - Study of and respond to changes/differences
  - Do NOT really analyze the causes and triggers of change

- Dynamics – more powerful … example

E.g.: Academic Worker “Motion”

– A trajectory, also the **change** in … motives:

A PhD student/Junior

A Prof/Senior

A Prof Emeritus

E.g.: Academic Worker “Motion”

– A trajectory, also the **change** in … motives:

A PhD student/Junior

A Prof/Senior

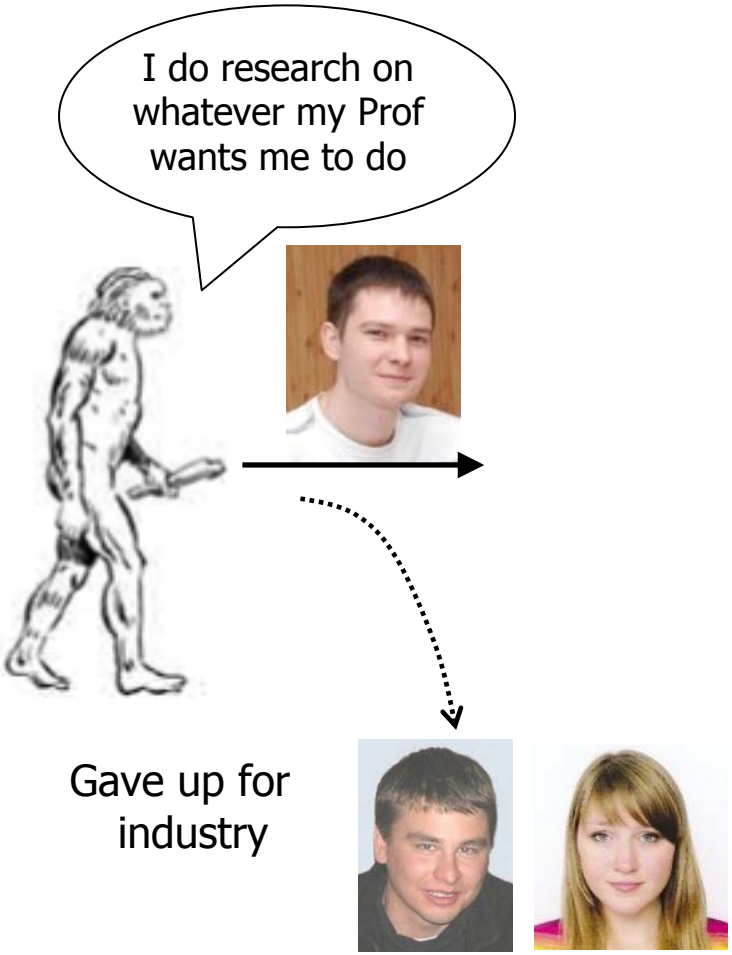
A Prof Emeritus

E.g.: Academic Worker "Motion"

– A trajectory, also the **change** in … motives:

A PhD student/Junior

A Prof Emeritus



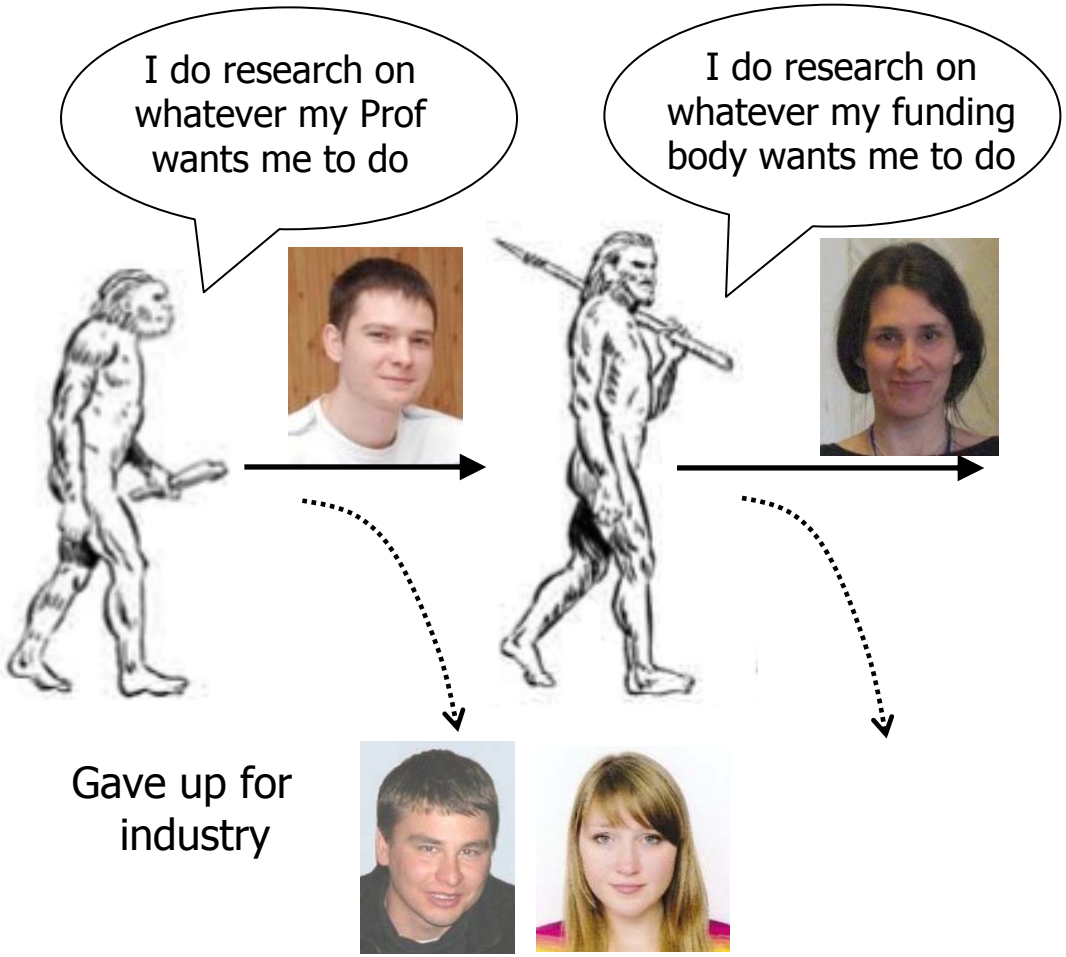
E.g.: Academic Worker "Motion"

– A trajectory, also the **change** in … motives:

A PhD student/Junior

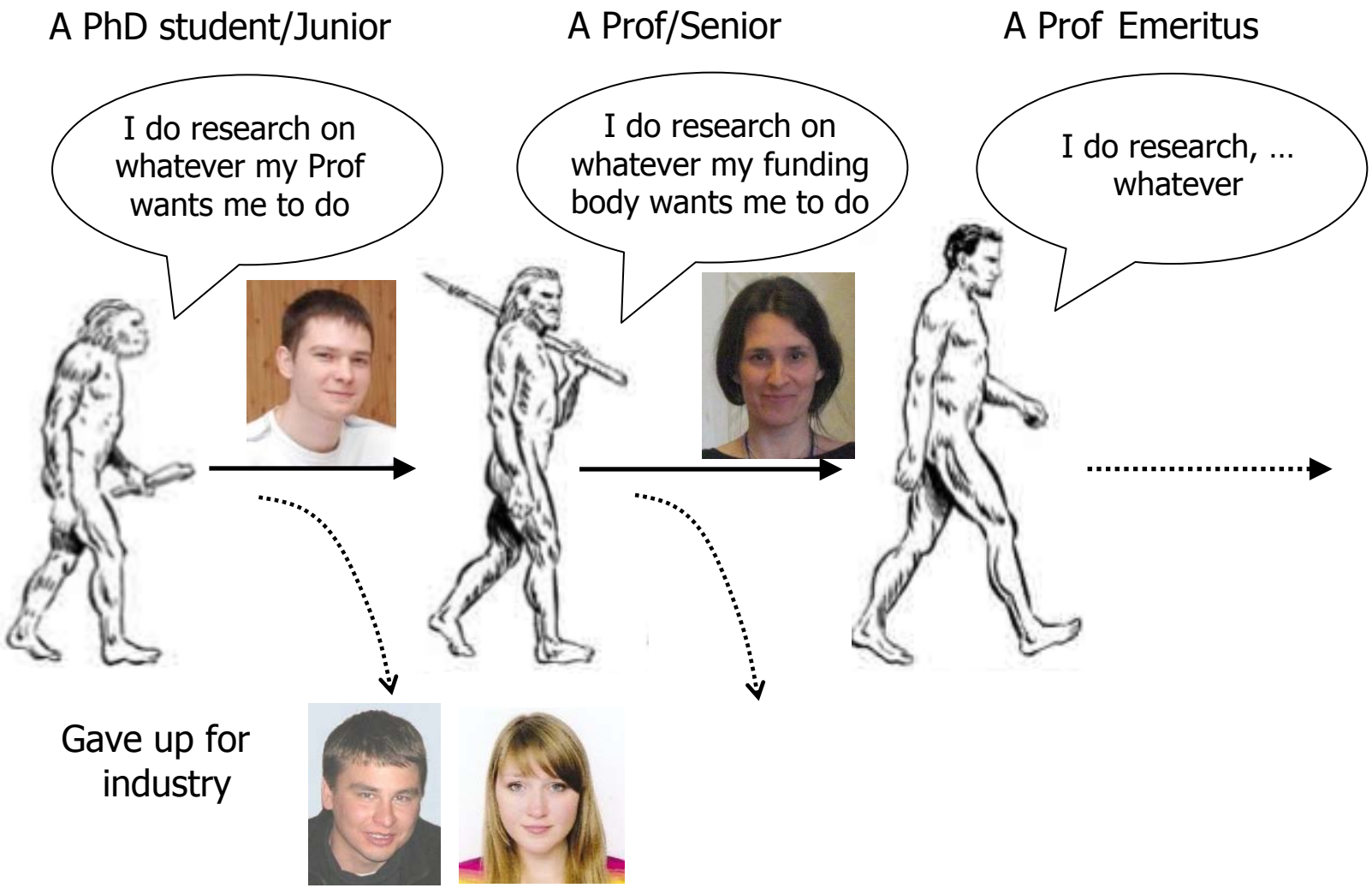
A Prof/Senior

A Prof Emeritus



E.g.: Academic Worker "Motion"

– A trajectory, also the **change** in … motives:



E.g.: Academic Worker "Motion"

– A trajectory, also the **change** in … motives:

A PhD student/Junior

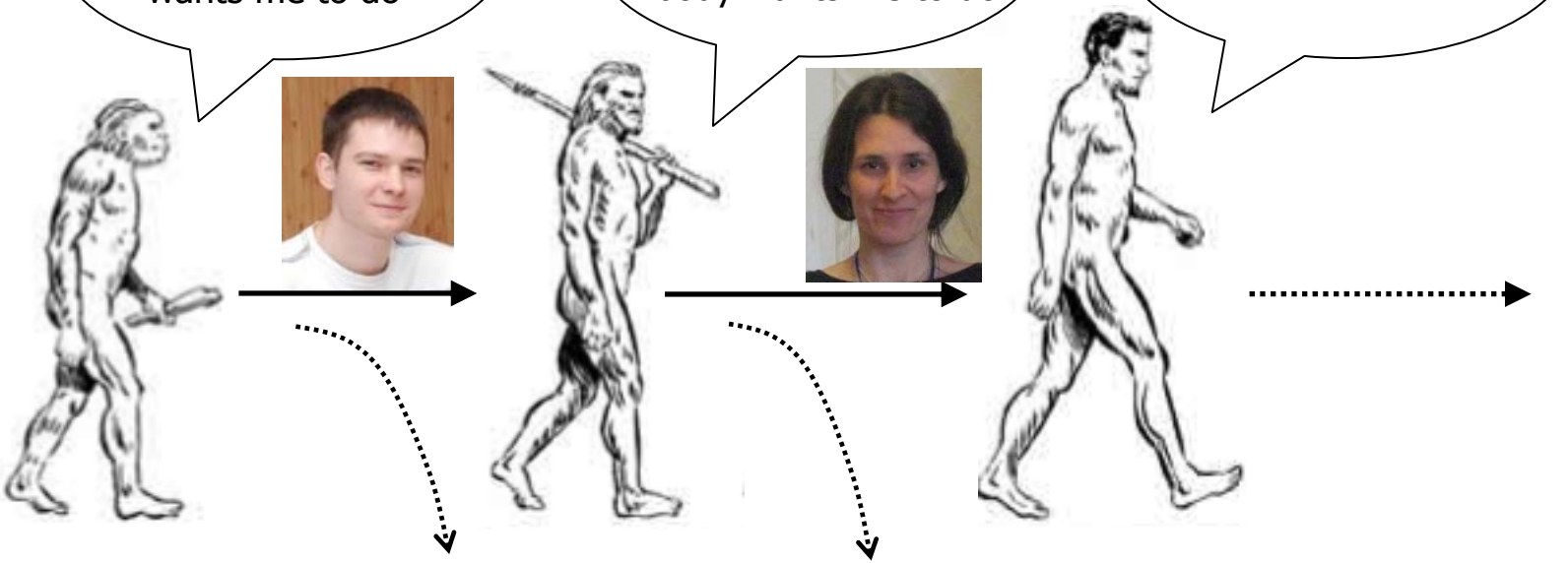
A Prof/Senior

A Prof Emeritus

I do research on whatever my Prof wants me to do

I do research on whatever my funding body wants me to do

I do research, … whatever

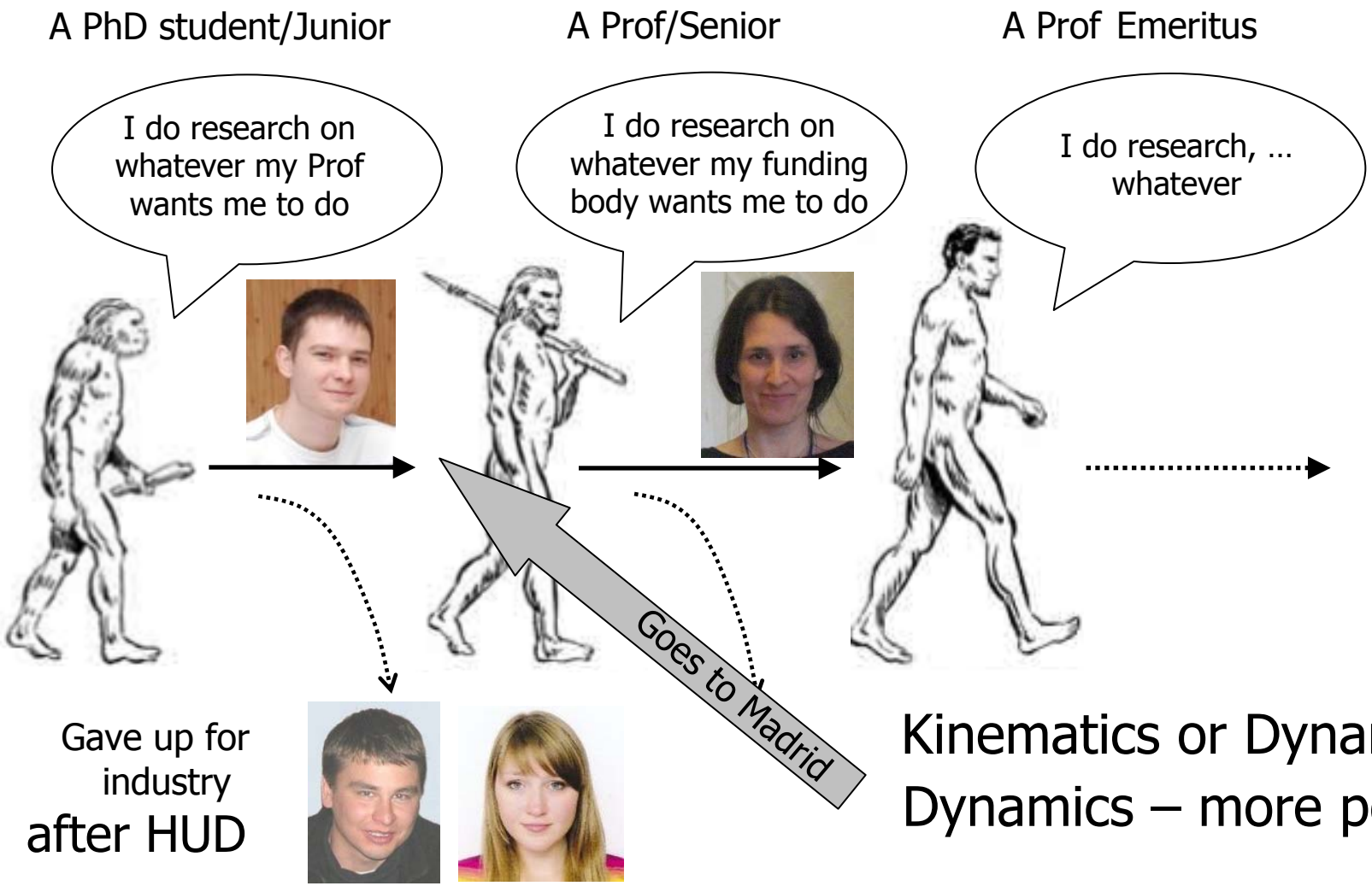


Gave up for industry

Kinematics or Dynamics?

E.g.: Academic Worker "Motion"

– A trajectory, also the **change** in … motives:



Kinematics or Dynamics?

Dynamics – more powerful!

A Law of Gravity?

- For the "motion" of knowledge representations, is there a(n analogue of a):
  - Newton's law?
  - Law of gravity?
  - Notion of entropy?
  - Feedback and causal loop model?
  - Mendel law?
  - Exponential growth model?
  - A system law?
  - Etc. of the kind …
- A positive answer will help a lot:
  - Reasons for the change in knowledge
    - More intelligent and efficient workflows
  - Limits for the scalability of KBS
    - Big Data/Knowledge settings

Aspects / Kinds of "Motion"

- Not complete – just to outline the foci of interest …
  - E.g. moving between network locations - not very much interested

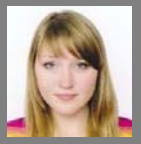
- Temporalized representations
  - Also a change of **state**. E.g. in versioning (Natalya)
- Change in **shape** (Ontology Schema)
  - Also the change in representation **language**
  - OntoElect, ontology learning (Olga, Eugene)
- Change in **population** (Individuals)
  - Ontology instance migration (Maxim)
- Change in pragmatic **context** (e.g. Domain)
  - E.g. would my process ontology for MIC Design fit also for Automotive? Would there be a change?
- Alltogether, inspired by Evolutionary Biology
  - Evolving Knowledge Ecosystems (myself in coop with JU.fi, vestforsk.no)





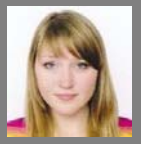
Details: **Ermolayev, V., Akerkar, R., Terziyan, V., Cochez, M.:** Towards Evolving Knowledge Ecosystems for Big Data Understanding. In: Akerkar, R. (ed.) Big Data Computing, pp. 1-55, Taylor & Francis, 2013, ISBN 978-1-46-657837-1

OntoElect: Change to Fit



- Ontology **refinement** methodology
- Ensures better ontology **fitness** through iterations
  - WRT stakeholder requirements
- Responds to the changes in stakeholder requirements
  - Tacit – not revealed explicitly
  - Evidence(s) learnt from the documents written by the stakeholders
    - Treated as stakeholder "**votes**"
  - Representativeness/Completeness assessed by measuring "**saturation**"

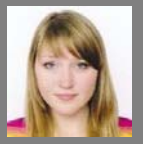
OntoElect: Change to Fit



- Ontology **refinement** methodology
- Ensures better ontology **fitness** through iterations
  - WRT stakeholder requirements
- Responds to the changes in stakeholder requirements
  - Tacit – not revealed explicitly
  - Evidence(s) learnt from the documents written by the stakeholders
    - Treated as stakeholder "**votes**"
  - Representativeness/Completeness assessed by measuring "**saturation**"
- Kinematics:
  - Ontology schema change, between the versions
  - Measured and shaped out using the **fitness** metric

Details: **Tatarintseva, S., Ermolayev, V., Keller, B., Matzke, W.-E.:** Quantifying Ontology Fitness in OntoElect Using Saturation- and Vote-Based Metrics. In: Ermolayev, V., et al. (Eds.) ICTERI 2013, CCIS Vol. 412, pp. 136–162, Springer, 2013, http://link.springer.com/chapter/10.1007/978-3-319-03998-5_8

OntoElect: Change to Fit



- Ontology **refinement** methodology
- Ensures better ontology **fitness** through iterations
  - WRT stakeholder requirements
- Responds to the changes in stakeholder requirements
  - Tacit – not revealed explicitly
  - Evidence(s) learnt from the documents written by the stakeholders
    - Treated as stakeholder "**votes**"
  - Representativeness/Completeness assessed by measuring "**saturation**"
- Kinematics:
  - Ontology schema change, between the versions
  - Measured and shaped out using the **fitness** metric
- Dynamics:
  - What is the decisive "critical mass" of new evidence that causes a change?
    - Majority vote (50%+1) OR a **minority vote** that matters?

Details: **Tatarintseva, S., Ermolayev, V., Keller, B., Matzke, W.-E.:** Quantifying Ontology Fitness in OntoElect Using Saturation- and Vote-Based Metrics. In: Ermolayev, V., et al. (Eds.) ICTERI 2013, CCIS Vol. 412, pp. 136–162, Springer, 2013, http://link.springer.com/chapter/10.1007/978-3-319-03998-5_8

OntoElect: Time Ontos Case

- The review of Time ontologies on the Semantic Web
  - ZNU and HUD, SemData project
- Questions:
  - Are the existing ontologies of time fit for the community requirements?
  - Do we have enough evidence from the community?
  - What is the minimal decisive set of evidence(s) that pictures the gaps?
- Community:
  - Temporal Representation and Reasoning
- Evidence(s):
  - Ranked sets of multi-word Terms
  - Extracted from their (representative?) document corpus
    - Full texts of the Proc. of TIME Symposia series, 1994 – 2013, ~440 papers, http://time.di.unimi.it/TIME_Home.html

Details: **Ermolayev, V., Batsakis, S., Keberle, N., Tatarintseva, O., Antoniou, G.:** Ontologies of Time: Review and Trends. *Int. J. of Computer Science & Applications*. 11(3), 57–115, 2014, http://www.tmrfindia.org/ijcsa/v11i34.pdf

Do we have Enough?

- Full texts:
  - Sorted in their chronological order,
  - Transformed to plain texts
  - Grouped in incremental slices

20 incremental slices:
1994
1994+1995
1994+1995+1996
….
1994+ ……. + 2013

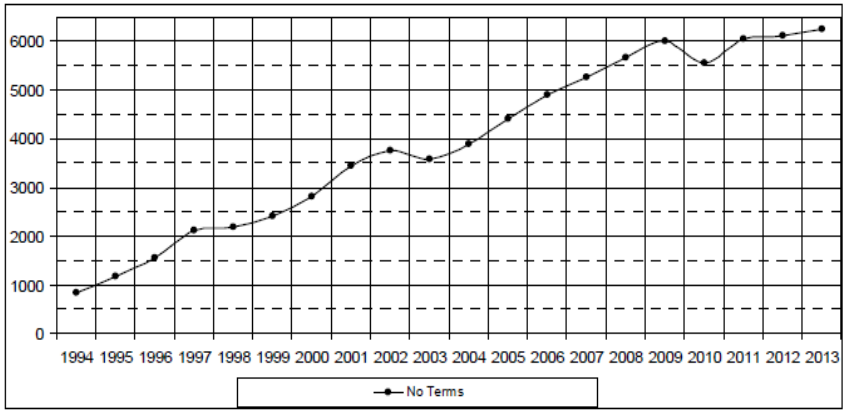
- For each slice S_i :
  - Extracted* the bag of Terms, ranked by score (*sc*)
  - Computed normalized scores (*nsc*)
  - Produced Termhood T_i by filtering out insignificant Terms (*nsc*<*eps*, *eps* computed to retain 50%+1 Term)
  - Computed absolute and relative termhood difference values: thd(T_i , T_{i-1}); $thdr = thd / \sum nsc$

* TerMine service by the UK National Centre for Text Mining (NaCTeM, http://www.nactem.ac.uk/). Scores computed using NaCTeM's multi-word term recognition technique (Frantzi *et al.,* 2000).

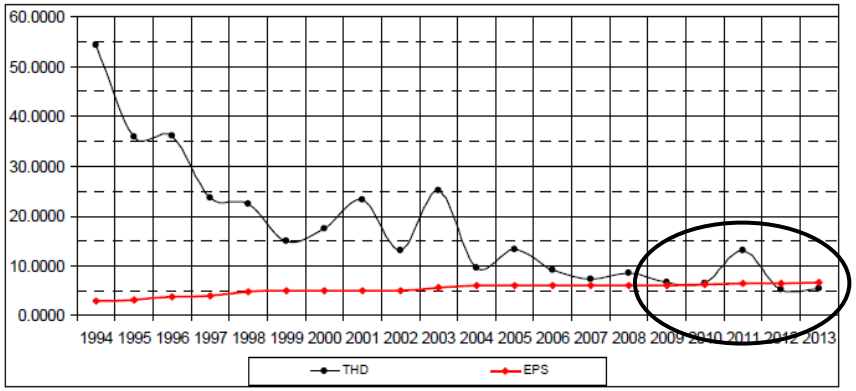
Do we have Enough?

Collection Slice	Terms in the		*eps*	*thd*, value	thdr, %
	Bag of Terms	Termhood			
1994	8546	838	3.0000	54.4448	100.0000
1994-1995	14597	1179	3.1699	35.9807	62.3806
1994-1996	23992	1548	3.7549	36.0855	59.6366
1994-1997	31427	2104	4.0000	23.7044	35.4153
1994-1998	38122	2183	4.7549	22.4341	30.7901
1994-1999	42788	2400	5.0000	14.9911	18.7218
1994-2000	49986	2821	5.0000	17.4853	20.7287
1994-2001	59294	3430	5.0000	23.1877	26.9035
1994-2002	65627	3767	5.0000	13.1819	15.3747
1994-2003	75171	3584	5.6147	25.0810	36.7663
1994-2004	81617	3893	6.0000	9.6005	13.8278
1994-2005	91692	4410	6.0000	13.3894	19.7595
1994-2006	101190	4903	6.0000	9.0502	12.6376
1994-2007	108203	5255	6.0000	7.3260	9.8946
1994-2008	115493	5658	6.0000	8.5976	11.7790
1994-2009	121832	6007	6.0000	6.6174	9.0302
1994-2010	128171	5564	6.3043	6.3422	9.0829
1994-2011	137918	6043	6.3399	13.0734	20.2061
1994-2012	145173	6109	6.3549	5.1033	8.0395
1994-2013	151075	6259	6.6667	5.4895	8.7677

- Termhoods became **saturated**
  - Termhood Difference goes below Individual Term Significance
- Terminological shift in time
  - Still not 0. Indicates domain changes over time
- The (representative) **majority vote**, but still <u>too many</u> terms retained



No of retained terms

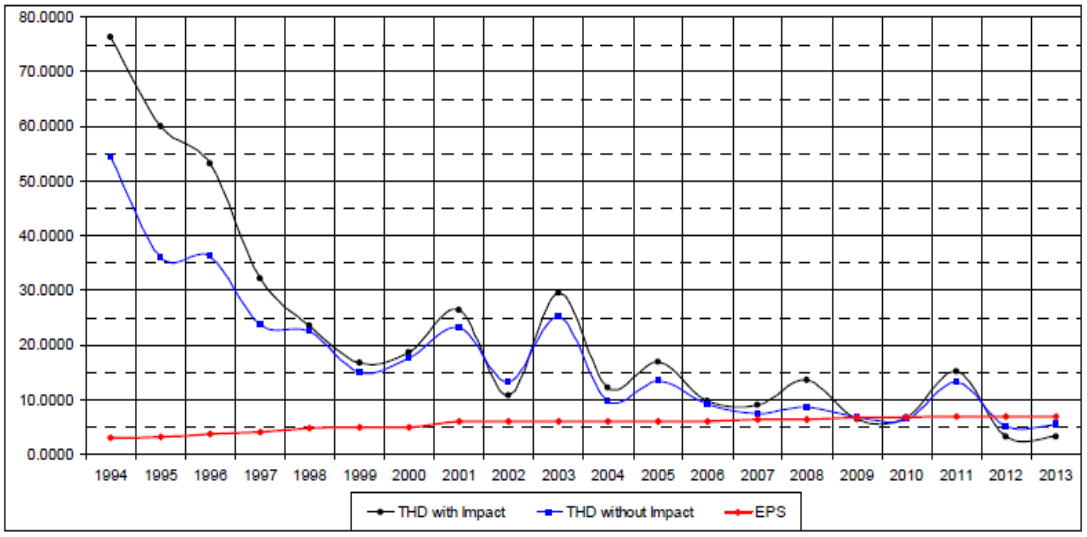


Absolute Termhood Difference (*thd*) and Individual Term Significance threshold (*eps*)

A Decisive Minority Vote?

- Terminology contribution peaks: 2001, 2003, 2005, 2008, and 2011
- Citation info collected (Google Scholar)
- Paper impact computed based on citation frequency (*cfr*)
- Papers with *imp*=*n* replicated n times – changing the incremental slices
- *thd*/ *thdr*/ *eps* re-computed
- Strong correlation
- Termhood based on high-impact (24) papers only

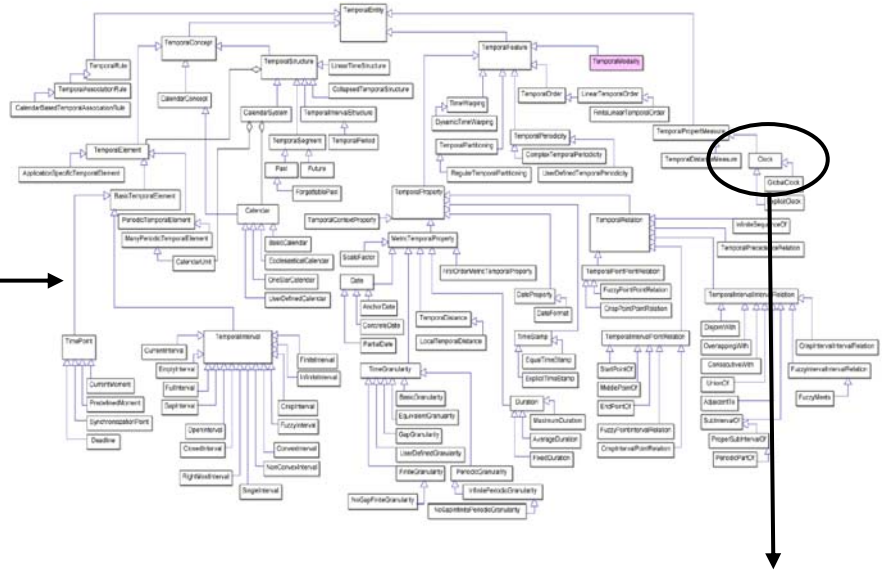
$$imp = \begin{cases} [0.2 \times cfr] + 1, cfr > 0 \\ 0, cfr = 0 \end{cases}$$

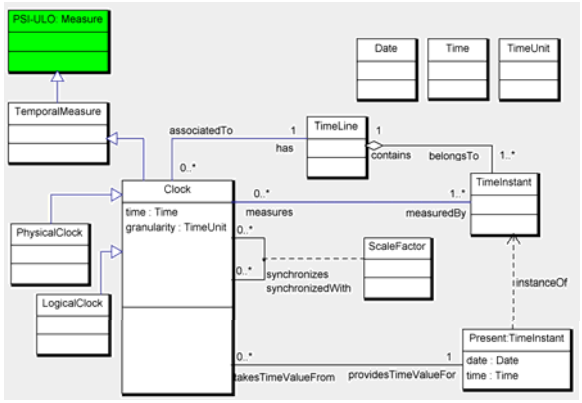


- 686 Terms vs 6,109
  - The "**influence**" that triggers change

Influence: Bag of Onto Tokens

Score	Term	Logic	Problem	Formula	Formalism	Operator	Method	Model	Reasoner	Domain	Language	Feature	Constraint	Instance	Pattern	Application	Project	Author
	Total No of terms: **686**	44	27	6	36	8	22	24	1	4	8	175	28	1	13	110	1	178
147.11	temporal logic	✓																
100.11	calendar pattern														✓			
86.54	temporal constraint												✓					
68.73	temporal operator					✓												
59.58	fuzzy match											✓						
52.25	temporal structure											✓						
49.83	calendar schema											✓						
46.25	temporal representation				✓													
41.00	temporal reasoning						✓											
40.00	freeze quantifier				✓													
37.73	fuzzy interval											✓						
36.36	xml document															✓		
36.00	crisp interval											✓						
34.00	satisfiability problem		✓															



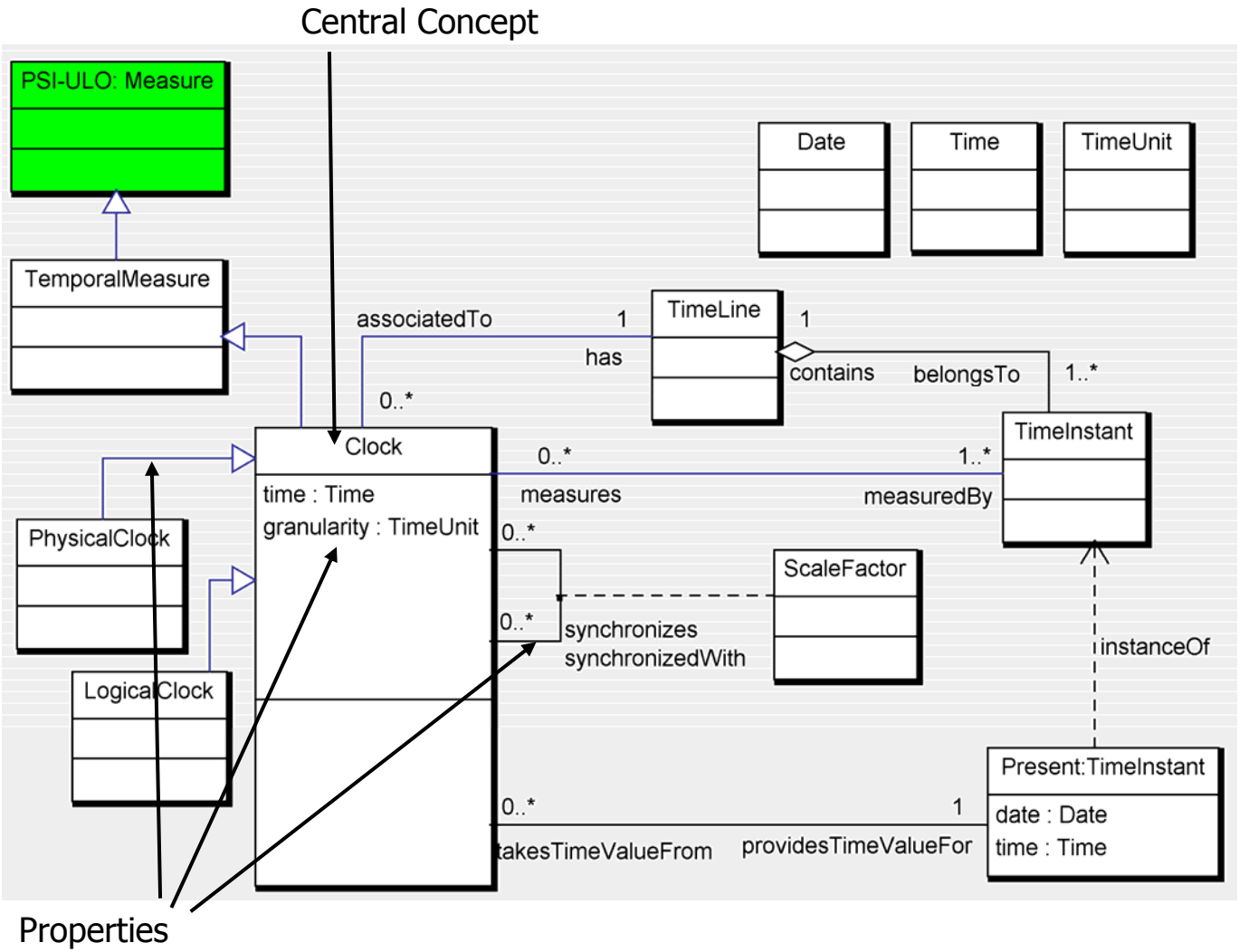


A **Clock** is a specific **TemporalMeasure** which measures **TimeInstant**s on a **TimeLine** by taking the value of current time from the **TimeInstant** instance of *Present.* A **Clock** is always associated with a particular single **TimeLine** (though there could be **TimeLine**s with no **Clock**). The **Clock**s which are associated with different **TimeLine**s may "run" quicker or slower compared to each other – thus reflecting the velocities of the time flow characteristic to their **TimeLine**s. These **Clock**s may be syncronized based on the use of the appropriate **ScaleFactor** (which is a **Rule** for comparing the time values of different **Clock**s). The granularity of the time value, provided by a **Clock**, is specified by the used **TimeUnit**.

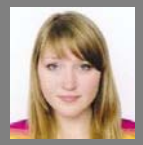
+ OWL

... ongoing work ...

Ontology Token



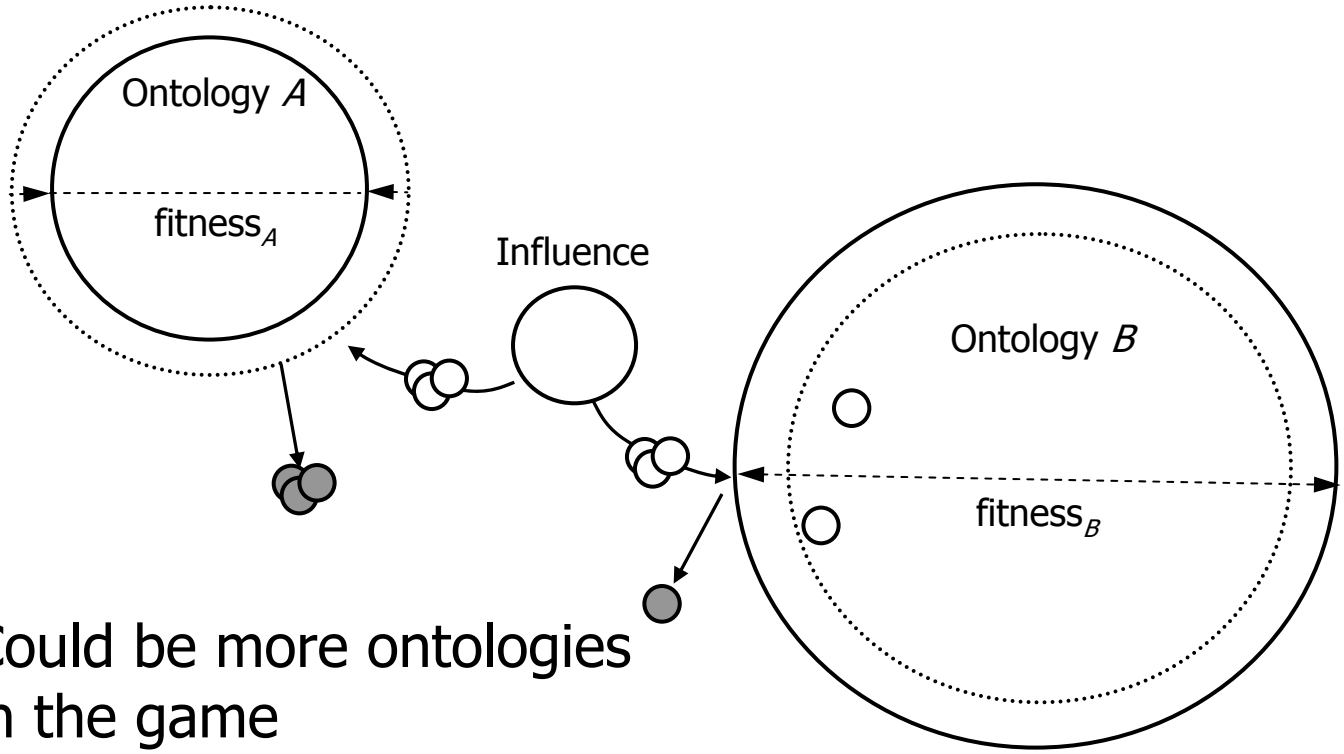
Ontology Fitness




- Token mappings: $\mu = (t, r, o, cf)$
  - *t* – ontology token (central concept, properties)
  - *r* – relationship {equivalence, membership, subsumption, meronymy, association}
  - *o* – ontology element
  - *cf* – confidence factor
- Positive votes: $v_o = ns \times w(r) \times cf$
  - *ns* – normalized score of the corresponding term (central concept)
  - *w*(*r*) – mapping relationship type weight
  - *cf* – mapping *cf*
- Propagated votes: $v_o^p = att \times v_{o^{sub}}$
  - Reflect the contribution of *o* to the semantics of the ontology element o^{sub} subsumed by o
  - *att* – attenuation coeff, chosen empirically
- Negative votes: $v_{t_i}^- = -ns_i$
  - No mapping – missing in *O* or contradicts to some *o*
- **Fitness**

$$\Phi_O = \Phi_O + \sum_{o \in O} v_o + \sum_{o \in O} v_o^p + \sum_{t \in T^{miss}} v_t^-$$

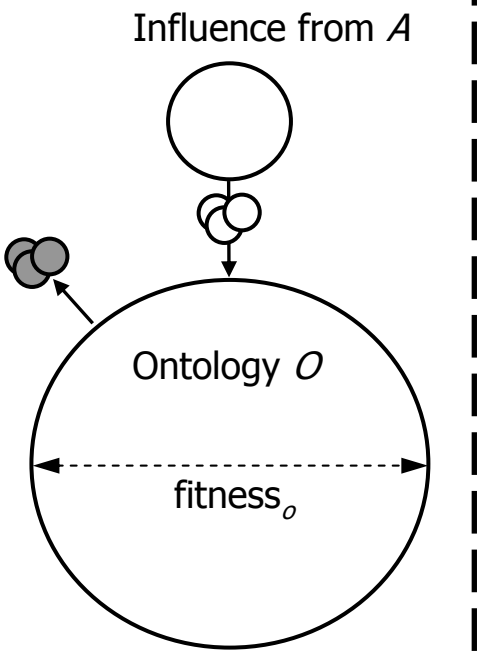
Competing Ontologies



- Could be more ontologies in the game
- Seams to be an equilibrium system
- A Law of "Preservation of Fitness" OR "Ontology Entropy"
  - To be further researched
- Target – 50%+1 (OntoElect)

Competing Domains

Domain *A*

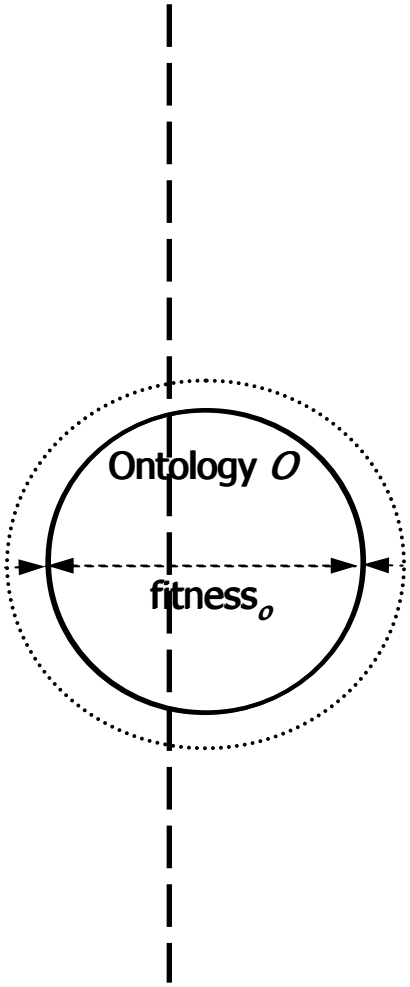


Domain *B*

Competing Domains

Domain *A*

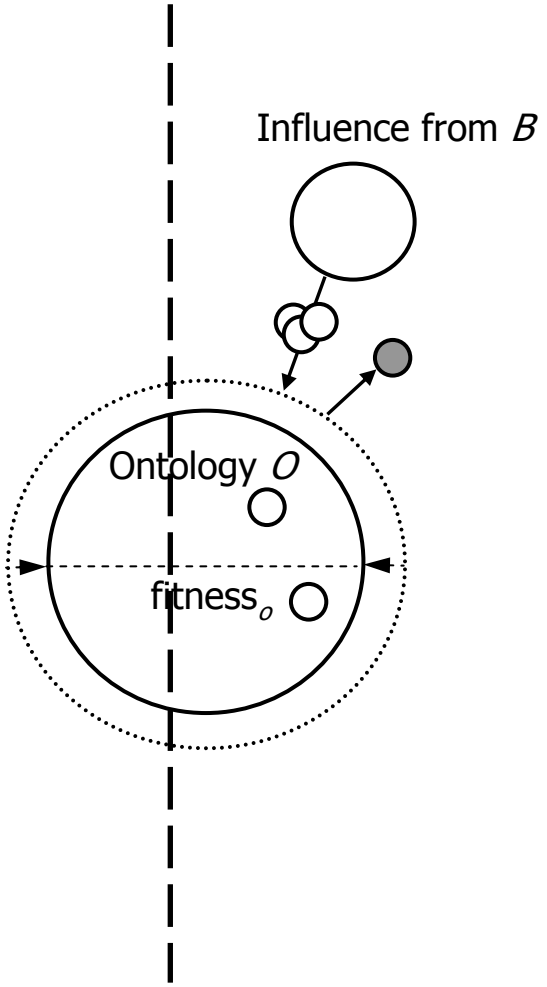
Domain *B*



Competing Domains

Domain *A*

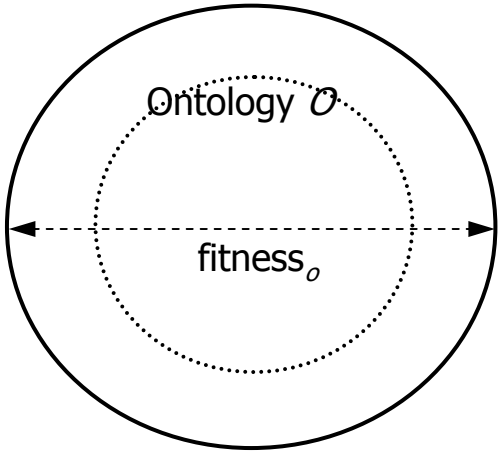
Domain *B*



Competing Domains

Domain *A*

Domain *B*



Some Conclusions

- In Ontology change and evolution studies:
  - Lack of focus on Dynamics vs Kinematics
- The "Laws of Dynamics" may be sought:
  - Looking at the analogies in:
    - System Dynamics
    - Population Dynamics
    - Statistical Mechanics
- Ontology Fitness:
  - May perhaps be used as an adequate feature for the "Laws of Dynamics"
  - Seems to be useful in:
    - Ontology refinement
    - Ontology reuse across domains
    - Choosing the best ontology among alternatives

Will be happy to answer your questions …

Will be also happy to continue discussions …

Room CW1/13,
vadim@ermolayev.com