Norges miljø- og
biovitenskapelige
universitet

Materialteori og -informatikk

Digitalisering på Ås

Subtraction and simulation

Martin Thomas Horsch

11th June 2023
GCAS Conference 2023 // 8th Annual GCAS Summer Institute

Fakultet for realfag og teknologi

Forskergruppe materialteori og -informatikk

Noregs miljø- og biovitskaplege universitet

Materialteori og -informatikk

Digitalisering på Ås

1. Logical subtraction
2. Molecular simulation
3. Epistemic metadata
4. Reproducibility claims

Example from Yablo: Someone who rejects ontological commitment to the **existence of numbers** is asked **how many prime numbers there are** greater than ten. "Infinitely many, of course, **except that** numbers don't exist."

Norwegian University of Life Sciences

Logical subtraction and subject matter

Logical subtraction is a concept from analytic philosophy.[1-3]

Its formalization is closely connected to the theory of **subject matter**.[2, 3]

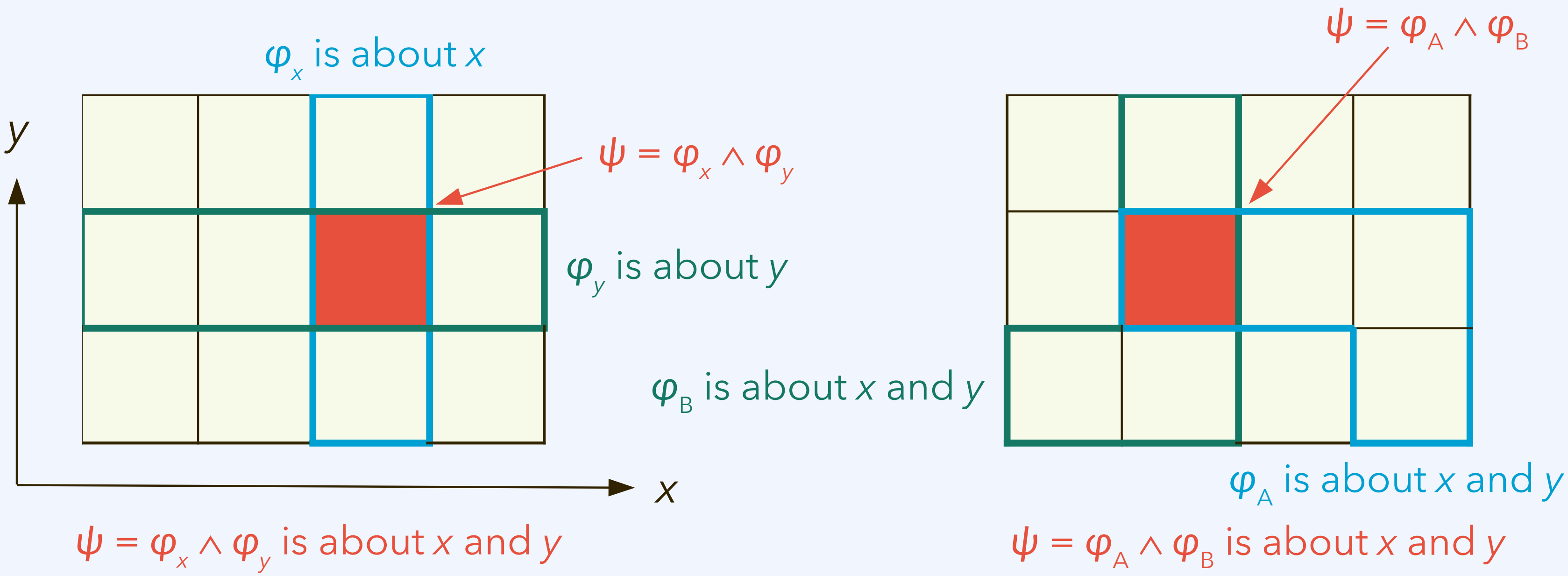
Could you try to **replicate my old simulation result**? Just do the same as I did.
Except that you of course log in with your user account, not mine.
Your result was off by 0,5%? **Don't worry**, that is totally normal.

Our **simulation of object *o*** confirms theory *s*.
Except that theory *s* deals with physical reality, and *o* is so simplified that **we know it cannot exist** or be built exactly in physical reality.

Example from Yablo:[2] Someone who rejects ontological commitment to the **existence of numbers** is asked **how many prime numbers there are** greater than ten.
"Infinitely many, of course, **except that** numbers don't exist."

[1]R. A. Jaeger, *Philos. Rev.* **82**(3): 320–329, doi:10.2307/2183898, **1973**.
[2]S. Yablo, *Aboutness*, Princeton Univ. Press (ISBN 978-0-691-14495-5), **2014**.
[3]K. Fine, *J. Philos. Log.* **46**: 675–702, doi:10.1007/s10992-016-9419-5, **2017**.

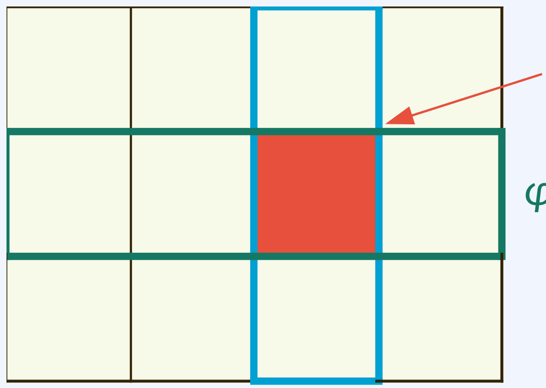
Conjunction as addition of information

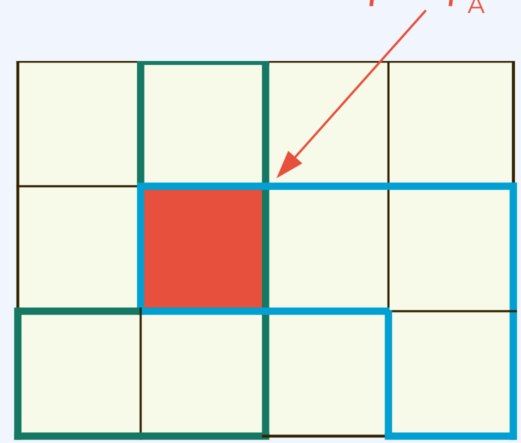


System with two separate elementary **topics** / **subject matters**: x and y.

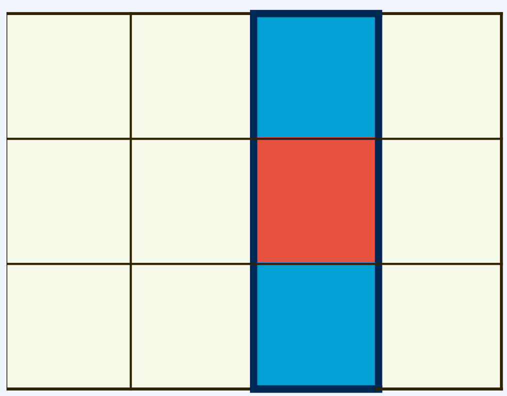
There are four possible values for x and three possible values for y.

Subtraction of information



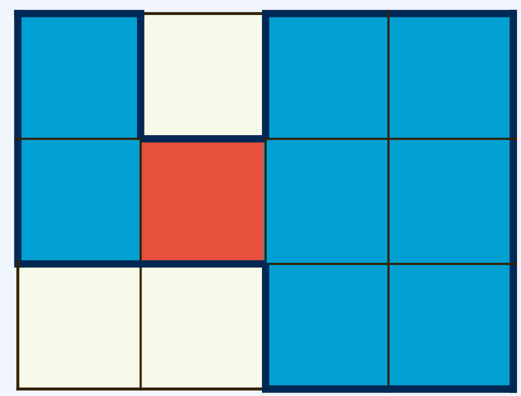


$$\psi - \varphi_y = (\varphi_x \wedge \varphi_y) - \varphi_y \equiv \varphi_x$$



φ_x is recovered by subtracting φ_y

$$\psi - \varphi_B = (\varphi_A \wedge \varphi_B) - \varphi_B \not\equiv \varphi_A$$



φ_A cannot be perfectly recovered

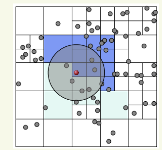
Noregs miljø- og biovitskaplege universitet

Materialteori og -informatikk

Digitalisering på Ås

1. Logical subtraction
2. Molecular simulation
3. Epistemic metadata
4. Reproducibility claims

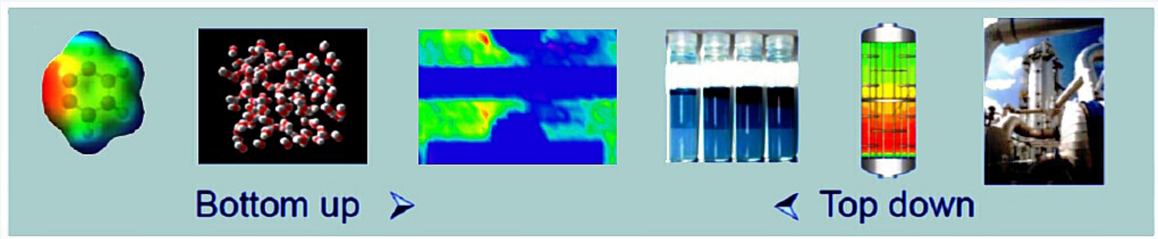
Molecular simulation:



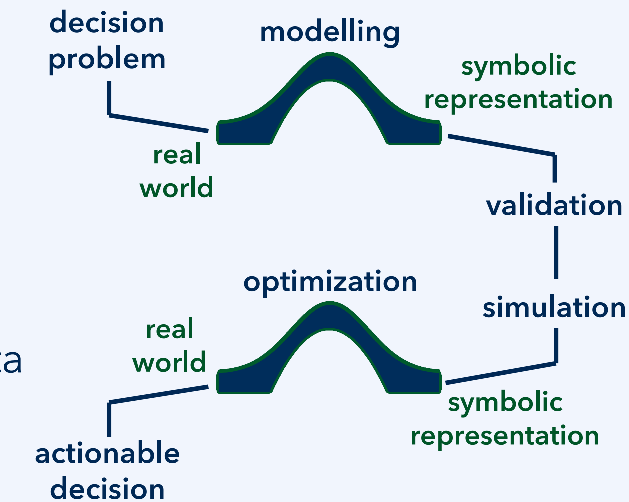
Computational Molecular Engineering

Simulation is a kind of **fiction**.
We must **suspend our disbelief** to accept the simulated scenario.
Can this suspension be understood as **subtraction**?

Molecular simulation in engineering



Modelling and simulation based decision support



physics-driven aspects (qualitative validity) ⟷ **data-driven aspects (quantitative reliability)**

- Realistic representation of underlying physical features
- Parts/aspects of the model represent aspects of the actual physical world

- Models with parameters that can be adjusted to data
- Reliable interpolation, extrapolation, prediction

(compared *e.g.* to purely machline learning based models)

Computational Molecular Engineering

- **Process Engineering**
- **Data Management and Technology**
- **Scientific and High-Performance Computing**

Fictional objects in computational engineering

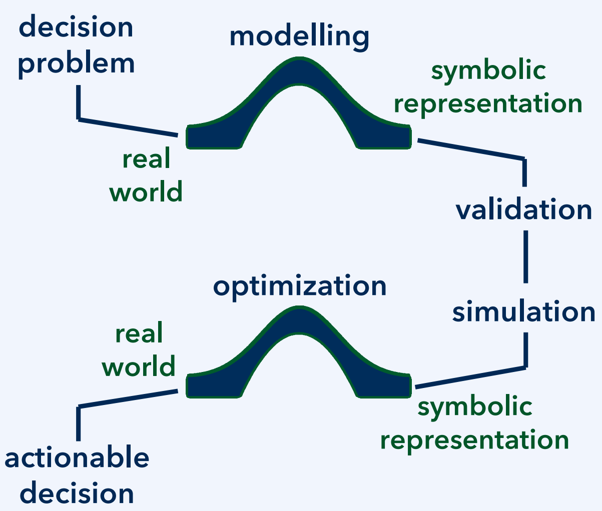
Naive view: The simulation **represents** a real physical process, the model represents a real physical system.

Actual practice:

The simulated process is almost always **fictitious**; often, it is **impossible** – it cannot technically occur.

While models legitimately represent real systems, they simplify them. Often, simulations really aim at **characterizing the model** as such, not a real system.

Modelling and simulation based decision support



Searle, *The logical status of fictional discourse:*[1]

- "to explore the **difference between fictional and serious** utterances [...] is not to explore the **difference between figurative and literal** utterances, which is another distinction quite independent of the first"
- "work[s] of fiction are made possible by [...] a set of conventions which **suspend** the normal operation of the **rules relating illocutionary acts and the world**"

[1] In J. R. Searle, *Expression and Meaning:* Chapter 3, Cambridge Univ. Press, **1979**.

Suspension as subtraction

Modelling and simulation has a **figurative/metaphorical aspect**: In the virtual reality of a simulation, there are *p*, *T*, *etc.*, and in physical reality, there are also *p*, *T*, *etc.*, but despite the same symbols, these are very different quantities.

But to be productive, this mechanism also requires an **aspect of fiction**. The model represents a fictitious entity *o*, but it is "**not about**" whether *o* can exist."

Our **simulation of object *o*** confirms theory *s*.
Except that theory *s* deals with physical reality, and *o* is so simplified that **we know it cannot exist** or be built exactly in physical reality.

Searle, *The logical status of fictional discourse:*[1]

- "to explore the **difference between fictional and serious** utterances [...] is not to explore the **difference between figurative and literal** utterances, which is another distinction quite independent of the first"
- "work[s] of fiction are made possible by [...] a set of conventions which **suspend** the normal operation of the **rules relating illocutionary acts and the world**"

[1]In J. R. Searle, *Expression and Meaning:* Chapter 3, Cambridge Univ. Press, **1979**.

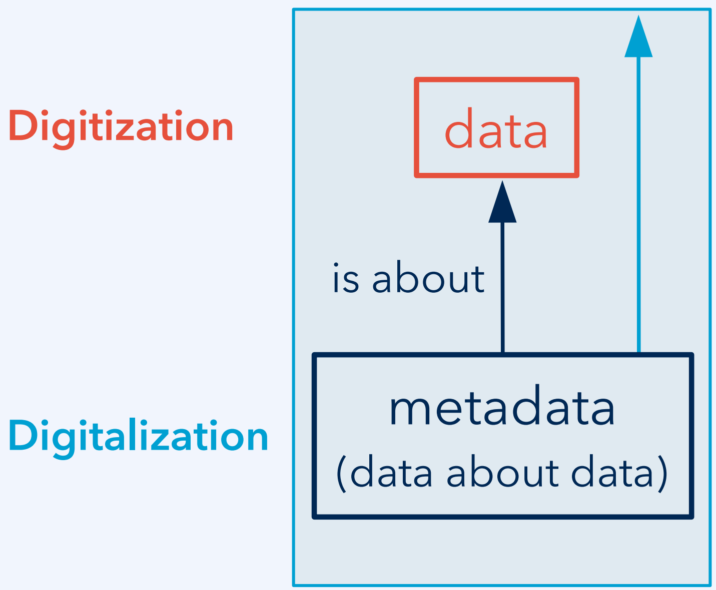
1. Logical subtraction
2. Molecular simulation
3. <u>Epistemic metadata</u>
4. Reproducibility claims

Epistemic opacity (Humphreys, 2011): A cognitive "process is **epistemically opaque** relative to a cognitive agent *X* at time *t* just in case *X* does not know at *t* all of the **epistemically relevant elements** of the process."

European AI Act proposal: "To address the **opacity** that may make certain AI systems **incomprehensible to or too complex for natural persons**, a certain degree of transparency should be required for high-risk AI systems. [...] High-risk AI systems should therefore be accompanied by **relevant documentation**".

Data management: Librarianship vs. engineering

Metadata are “descriptive data about an object” (ISO 11179).



FAIR digital object[1, 2]

The librarian:

- Focus on **archival** and **curation**
- **Help humans** use digital artefacts
- Focus on **provenance**, like for artefacts in a museum, so humans understand where they come from

The engineer:

- Computers must **understand** what the digital artefacts mean
- **Focus on knowledge**
- FAIR digital objects[1, 2]
- Aim: Machine-actionability[2]

[1] I. Anders *et al.*, *FAIR Digital Object Technical Specification*, doi:10.5281/zenodo.7824713, **2023**.

[2] C. Weiland, S. Islam, *et al.*, *FDO Machine Actionability*, doi:10.5281/zenodo.7825649, **2023**.

Epistemic metadata

Metadata are "descriptive data about an object" (ISO 11179).

Epistemic metadata are those that help establish the knowledge status of data.[1]

Epistemic metadata in the PIMS-II mid-level ontology:

a) "what **knowledge claim φ** has been formulated?,"
b) "where do the data and the claim come from?" (**provenance**),
c) "what **validity claim** was made about φ?,"
d) "why should we accept any of this?" (**grounding**).

Case study from molecular thermodynamics

- First stage, evaluating ten journal articles, doi:10.5281/zenodo.7516532.
- Second stage, discussing twelve claims, doi:10.5281/zenodo.7608074.

[1] M. T. Horsch, B. Schembera, in *Proc. JOWO 2022*, CEUR *vol.* **3249**: *p.* 2 (CAOS), **2022**.

Reproducibility and falsification[1, 2]

Research data infrastructures must accommodate mutually contradicting claims. They should also assist researchers at validating/falsifying each other's work.

Let us look into a "falsification" or "unsuccessful reproduction" of *a*'s work by *b*:

Knowledge claim (KC), including the provenance
«Researcher a did κ and found φ (and thus claims to know φ).»
→ Therefore, when research process κ is carried out, it **can** lead to the outcome φ.

1) Reseacher *a* did κ and found φ.
2) Reseacher *b* did γ, which is very similar to κ, and found ζ, not very similar to φ.
3) Nobody disputes a's integrity. Nobody disputes that "*a* did κ and found φ."

What allows *b* to claim that this is some sort of falsification?

[1] M. T. Horsch, S. Chiacchiera, G. Guevara, M. Kohns, *et al.*, in *Proc. FOIS 2023*, to appear, **2023**.

[2] H. E. Plesser, *Frontiers Neuroinform* **11**: 76, doi:10.3389/fninf.2017.00076, **2018**.

Noregs miljø- og biovitskaplege universitet

Materialteori og -informatikk

Digitalisering på Ås

1. Logical subtraction
2. Molecular simulation
3. Epistemic metadata
4. Reproducibility claims

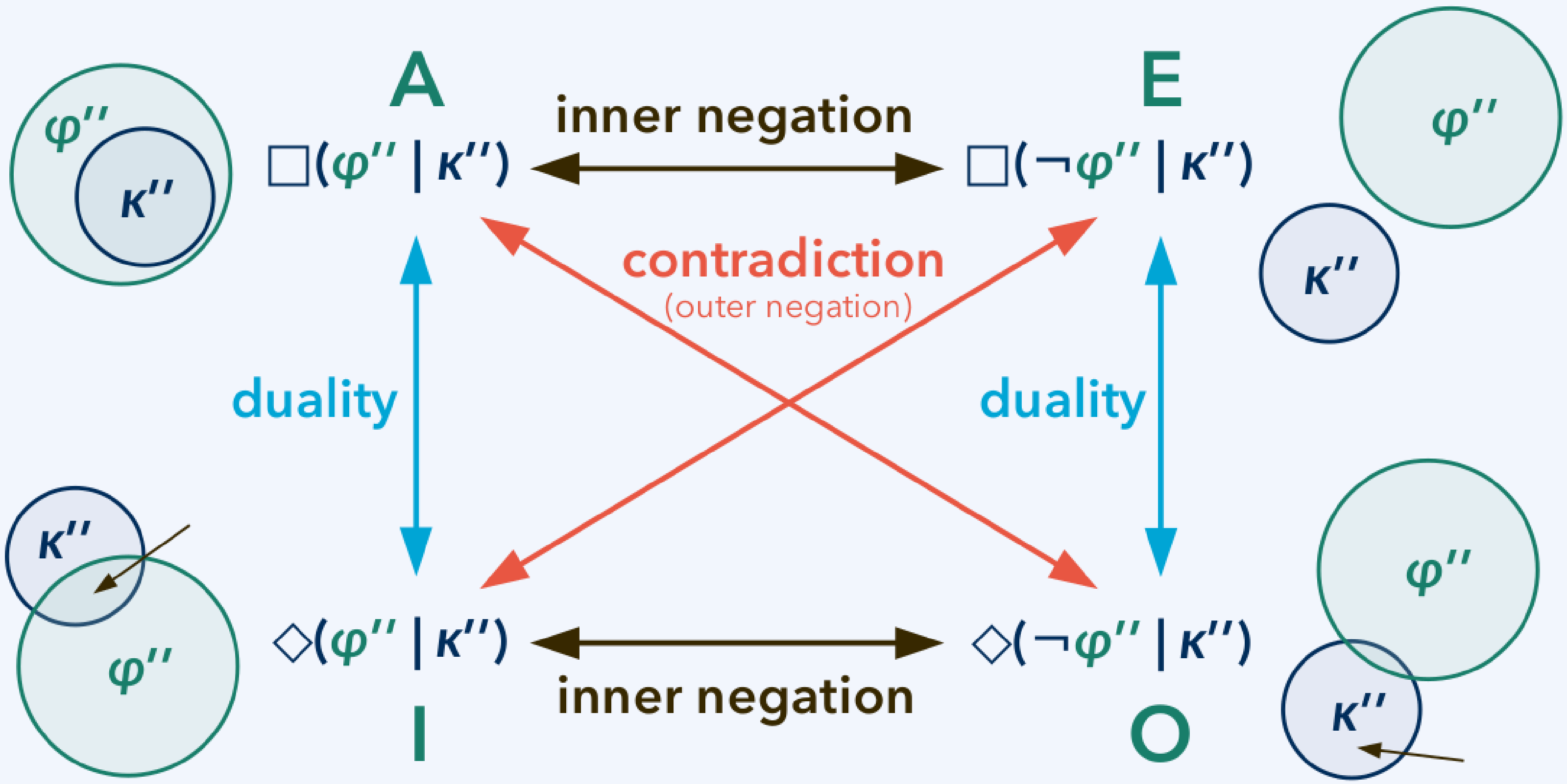
Reproducibility claim (RC)

«Whenever the research process κ'' is carried out, it **must** lead to the outcome φ''.»

Modal square of opposition

If the research process conforms with κ'', the outcome **must conform** with φ''.

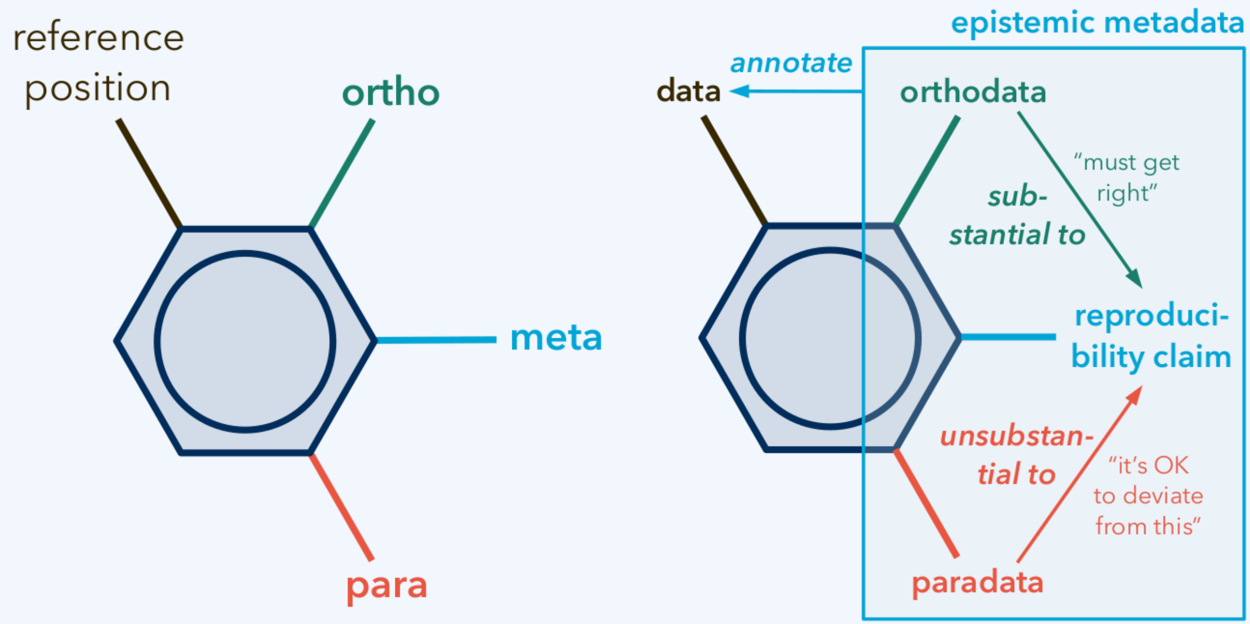
If the research process conforms with κ'', the outcome **must not conform** with φ''.



If the research process conforms with κ'', the outcome **can conform** with φ'' (and it is possible to conform with κ'').

If the research process conforms with κ'', the outcome **can disagree** with φ'' (and it is possible to conform with κ'').

Orthodata and paradata



provenance metadata κ
provenance paradata κ'

---

provenance orthodata κ'' $= \kappa - \kappa'$

«repeat κ, but no need to retain κ'»

knowledge claim metadata φ
knowledge claim paradata φ'

---

knowledge claim orthodata φ'' $= \varphi - \varphi'$

«obtain φ again, except for φ' maybe»

Reproducibility claims[1]

Norwegian University of Life Sciences

Common formulation and schema for reproducibility claims (RCs):

«Whenever research process κ'' is carried out, it must lead to the outcome φ''.»

we argue that there is a mechanism from pragmatics at work here[1]

1) Reseacher *a* did κ and found φ.

   Here, *a* also made the **positive reproducibility claim** $\boldsymbol{\psi} = \Box(\varphi'' \mid \kappa'')$.

2) Reseacher *b* did γ, **consistent with** $\boldsymbol{\kappa''}$, and found ζ, **inconsistent with** $\boldsymbol{\varphi''}$.

   Here, *b* made the **negative reproducibility claim** $\Diamond(\neg\varphi'' \mid \kappa'') \equiv \neg\Box(\varphi'' \mid \kappa'') \equiv \boldsymbol{\neg\psi}$.

3) What is relevant there is the **contradiction between** $\boldsymbol{\psi}$ **and** $\boldsymbol{\neg\psi}$.

provenance	knowledge claim
provenance metadata κ	knowledge claim metadata φ
provenance paradata κ'	knowledge claim paradata φ'
provenance orthodata $\boldsymbol{\kappa''} = \kappa - \kappa'$	**knowledge claim orthodata** $\boldsymbol{\varphi''} = \varphi - \varphi'$
«repeat κ, but no need to retain κ'»	«obtain φ again, except for φ' maybe»

[1]M. T. Horsch, S. Chiacchiera, G. Guevara, M. Kohns, *et al.*, in *Proc. FOIS 2023*, to appear, **2023**.

Norges miljø- og biovitenskapelige universitet

Materialteori og -informatikk

Digitalisering på Ås

Subtraction and simulation

Martin Thomas Horsch

11th June 2023

GCAS Conference 2023 // 8th Annual GCAS Summer Institute

Fakultet for realfag og teknologi

Forskergruppe materialteori og -informatikk