

Exploring the Philosophy of Data

Brian Ballsun-Stanton

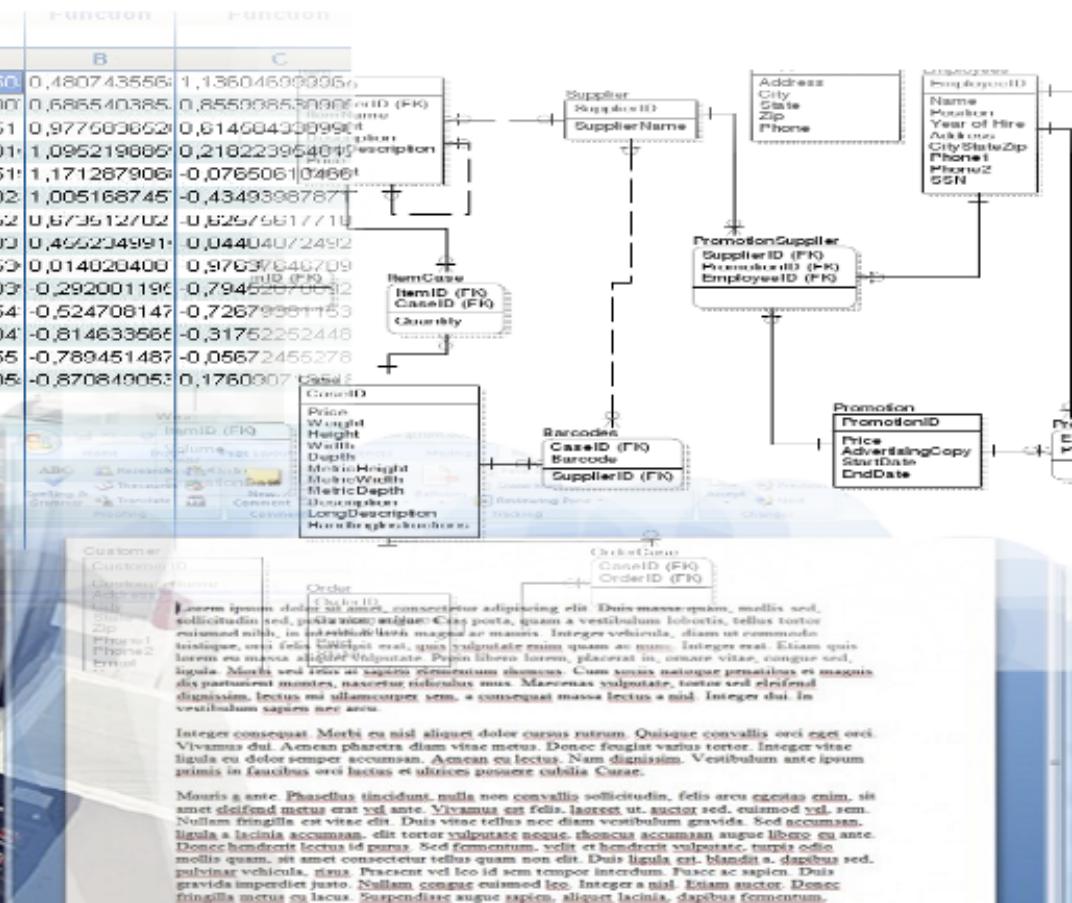
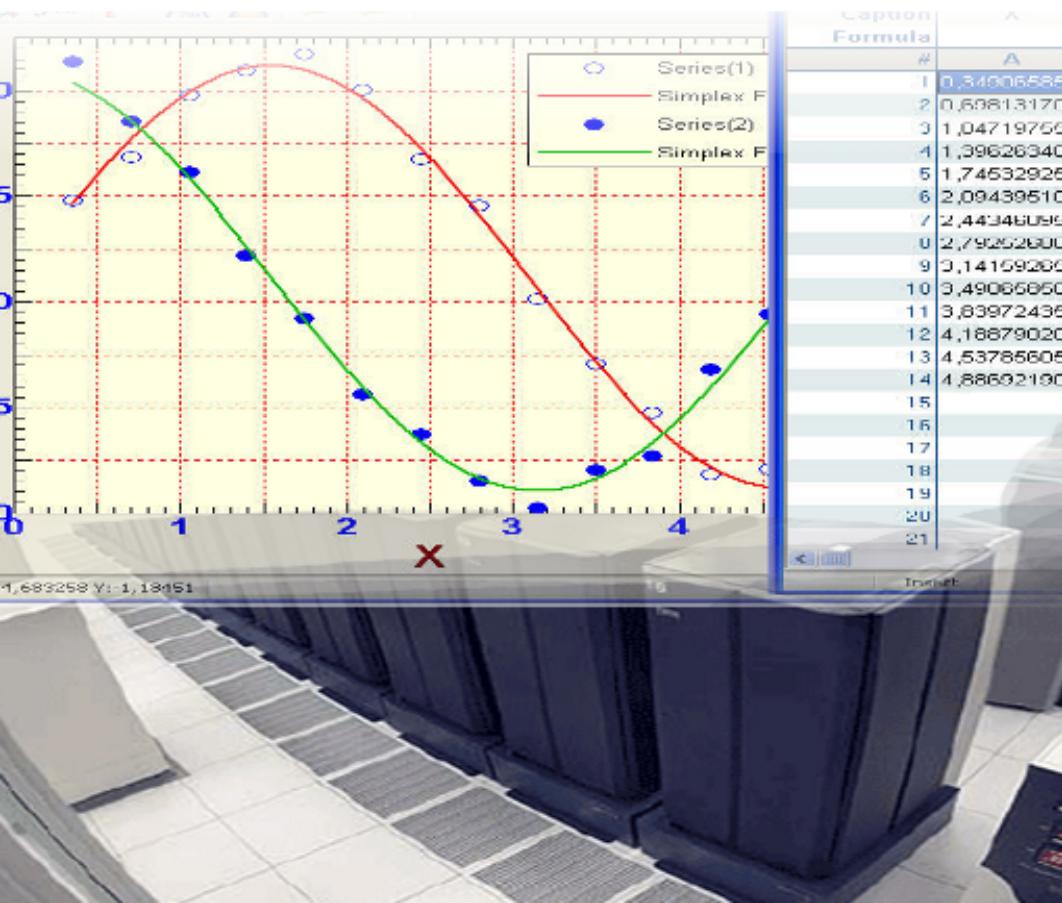
Ph.D in the Philosophy of Science at UNSW

M.S. in Information Technology at
Rochester Institute of Technology

Data can be different things...

however, each thing is *data* to someone.

Bits Hard Numbers Observations



Some Different Philosophies

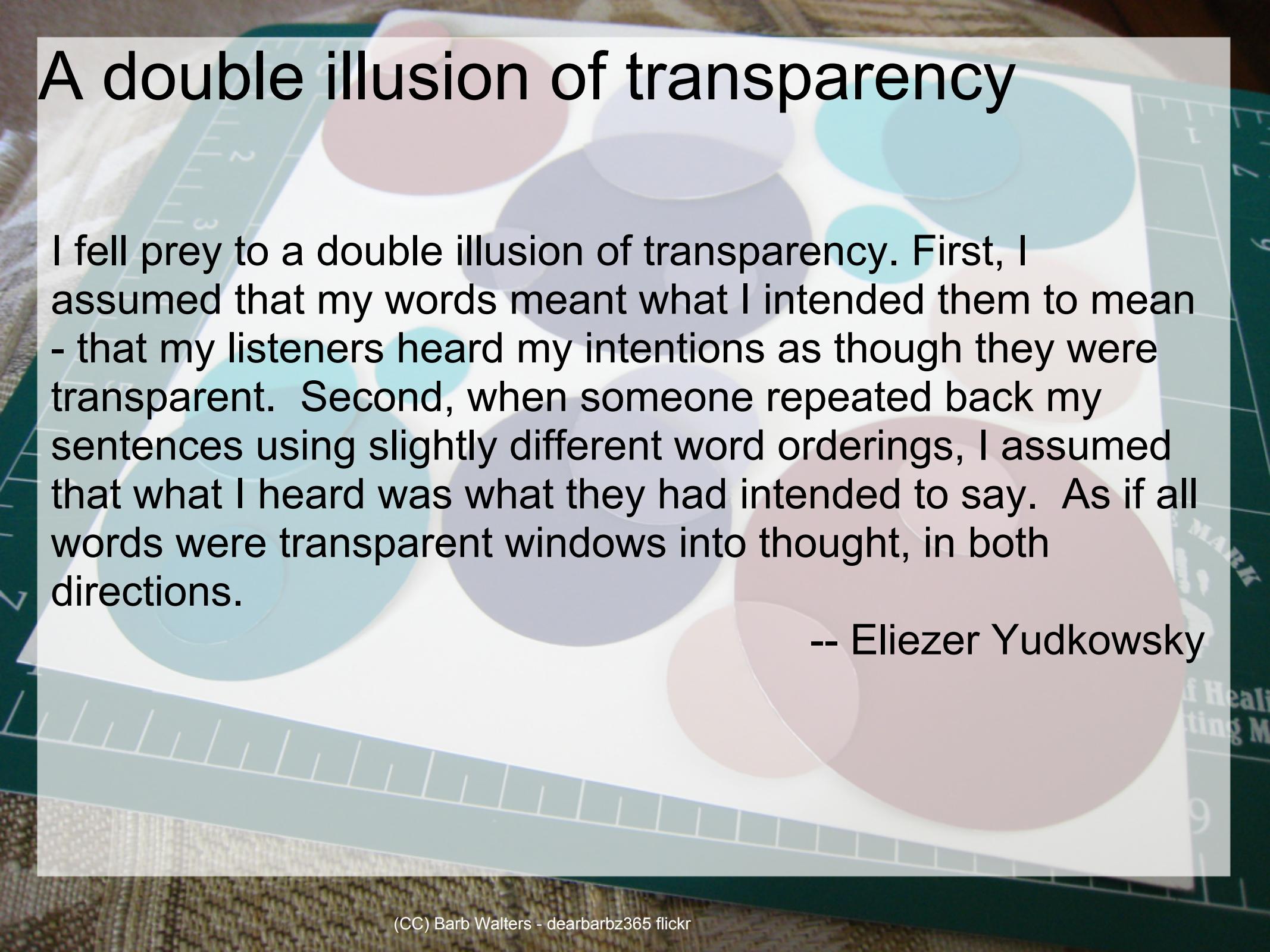
Data is something inherently technological, electronic storage for human information and knowledge.

Data are hard numbers, objective, quantifiable, reproducible. Only factual numbers need apply.

Data are recorded observations about anything. Qualitative or quantitative. They can be embodied in things by their design and automatically produced by machines.

Ontologies of information and knowledge above this are so mixed to be not quickly reportable.

A double illusion of transparency



I fell prey to a double illusion of transparency. First, I assumed that my words meant what I intended them to mean - that my listeners heard my intentions as though they were transparent. Second, when someone repeated back my sentences using slightly different word orderings, I assumed that what I heard was what they had intended to say. As if all words were transparent windows into thought, in both directions.

-- Eliezer Yudkowsky

OCHRE and ArchaeoML

- "Each XML document in the database corresponds to a real-world entity such as an archaeological site, context, or artifact; a written text ... a person ... a map or drawing... and so on."
- "...[I]t is difficult to capture the variability and complexity of the entities encountered ... there are few common standards for how entities should be described. **Observations** are forced into a rigid tabular mold, relegating information about idiosyncrasies ... to unstructured ... free-form notes."

= Data as Subjective Observation

Heurist

"Heurist is designed as a collaborative database containing - in a single database - a wide variety of items, from bibliographic records to internet bookmarks or photographs; from geological samples and museum artefacts to historical events or university courses."

= Data as Encoded Human Communications

Conclusions

To figure out your philosophy, start trying to classify things as "Data", "Information", or "Knowledge." There are more philosophies out there than what I've found.

Local languages have evolved: "raw data and derived data" that create a local understanding of the nature of data.

Be aware that different philosophies of data exist. If there seem to be misunderstandings and confusion, sanity check basic philosophical principles like the philosophy of data.



Discussion?

