# The Informative Converse paradox: Windows into the unknown

Harald Martens
Centre for Integrative Genetics CIGENE/IMT-UMB; Nofima, Ås, Norway

**Trapped by our incomplete knowledge?** Correct causal understanding is the goal of natural science. Conventional hypothesis-driven research with analysis in its simplest form - measuring only what is known to be relevant, and then projecting the measurements on prior knowledge – is undoubtedly useful. But incomplete causal knowledge can be a jail for the mind - so we need ways to escape. Extensive empirical observation may hold an escape key – if we can find it in a mass of data.

Disciplined compilation of data from many informative samples, each characterized by many informative variables, creates valuable data tables. Without data-analysis tools, big data tables create mental overflow for the scientist, whether our observation is based on real world sampling, controlled experiment or both. But suitable Do-It-Yourself software now exists for scientists to analyse their own data, searching for causal confirmation and surprise in light of their more or less tacit knowledge, and without unreasonable effort or risk of false discovery [1]. However, these data analysis software tools need to be further improved to quantify the surprises thus discovered.

**The Informative Converse**. In my lecture I shall demonstrate two different windows into the unknown, allowing correct new insight in situations where conventional interpretation is destroyed by unexpected causes that are not properly handled. I show how these windows can be opened in practice. And this reveals a puzzling limitation: Correct new discovery is in principle *only* possible in the *converse* of our prior knowledge – in the domain(s)* in which we have no causal knowledge, or at least have not used such knowledge explicitly in the analysis. In other words: *What we claim to know the least about, we can observe most clearly.* In contrast, in domains where we use our prior knowledge explicitly, the shadow of incomplete knowledge may lock us into a mental jail.

This paradox of the Informative Converse (IC) has recently been described mathematically [2]. It will here be illustrated graphically for simple chemical mixtures and for models with unexpected contributions. The analysis consists in a two-stage multivariate "soft" modelling: First, observed multi-channel data are interpreted in light of our presumed background knowledge by projection, row-wise or column-wise. Then the residuals are analyzed for eigen-structure, in which unexpected phenomena will manifest themselves. On this basis, suggestions will be given for how to extend some common statistical techniques for causal interpretation, e.g. ANOVA, and for how to reveal deficiencies in causal mechanistic models in science via multivariate meta-modelling.

**Is this how we progress?** More speculatively – the IC will be proposed as a possible basis for understanding how culture develops in society. The IC is based on the interplay between two mundane, but important ontological category types, *Samples* and *Variables,* "physical reality" and "pure information", where the former represents *Kinds*, *Objects* or *Instances* while the latter are descriptors of *Attributes* or *Property-universals*, as in Lowe's [3] four-category ontology for natural science. The human mind is a powerful "belief machine", creating systems of theoretical explanation - postulated causal relationships between property-universals, right or wrong. At the same time, our sense perception [4] - including our measuring instruments - allows us to discover patterns in physical timespace – to the extent our pre-judice allows us to look and see. When established explanation collides with observed surprise, individuals searching for truth will experience cognitive dissonance. Over time, society therefore zigzags forward between inductive exploration and deductive orthodoxy. I think science will tack faster forward by measuring more variables than necessary in more samples than necessary, and analyzing the data tables by multistage multivariate soft modelling.

* By *domain* I mean *ways* in a data table, e.g. samples (rows) and variables (columns) in a two-way table. Multichannel spatiotemporal measurements generate data tables with more than two ways and hence offer more opportunities for discovery.

**References***:*


[1] H.Martens and M. Martens (2001) *Multivariate Analysis of Quality. An Introduction*. J. Wiley UK

[2] H.Martens (2011) The Informative Converse paradox: Windows into the unknown. *Chemometrics and Intelligent Laboratory Systems* **107**, pp 124-138.

[3] E. Jonathan Lowe (2006) *The four-category ontology: a metaphysical foundation for natural science*. Oxford University Press, Oxford UK.

[4] M. Martens and H. Martens (2008) The Senses Linking Mind and Matter. *Mind and Matter – an interdisciplinary journal of mind-matter research*, **6**, pp 51-86.