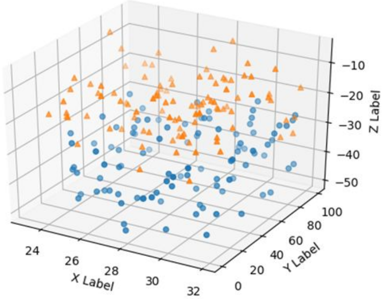
MIDWEST ARCHITECTURE COMMUNITY COLLABORATION 2020

NOVEMBER 5, 2020

Augmenting Analytics with Causal Reasoning AI

WHO NEEDS CAUSAL REASONING AI

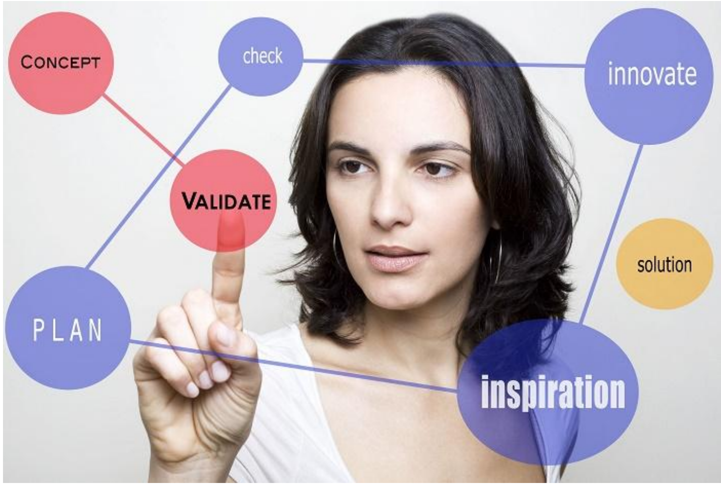
- Most systems can’t tell you why…
  - Human brains are needed to hypothesize, experiment and conclude
- Answering why and how questions is often critical to better decisions
- Why did sales slump in region R at time T?
- Why did disease spread jump in region R at time T?
- Given factors X and Y, what are likely outcomes?

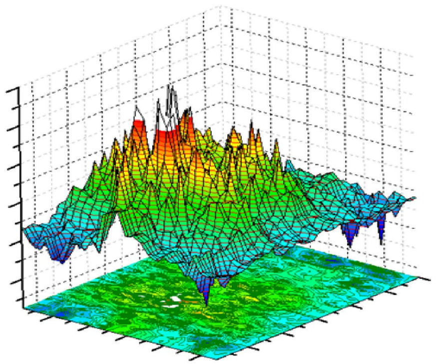


midwest
architecture community
collaboration

LIMITATIONS OF DESCRIPTIVE BI

- "That which is measured improves. That which is measured and reported improves exponentially." ~ Karl Pearson
- "Correlation does not imply causation." ~ Karl Pearson
- "We must be careful not to confuse data with the abstractions we use to analyze them." ~ William James
- "…with no antecedent knowledge of the causation or absence of causation … the calculation of correlation coefficients, total or partial, will not advance us a step toward evaluating the importance of the causes at work." ~ R. A. Fisher
- "Data can tell you that the people who took a medicine recovered faster than those who did not take it, but they can't tell you why." ~ Judea Pearl





midwest architecture community collaboration

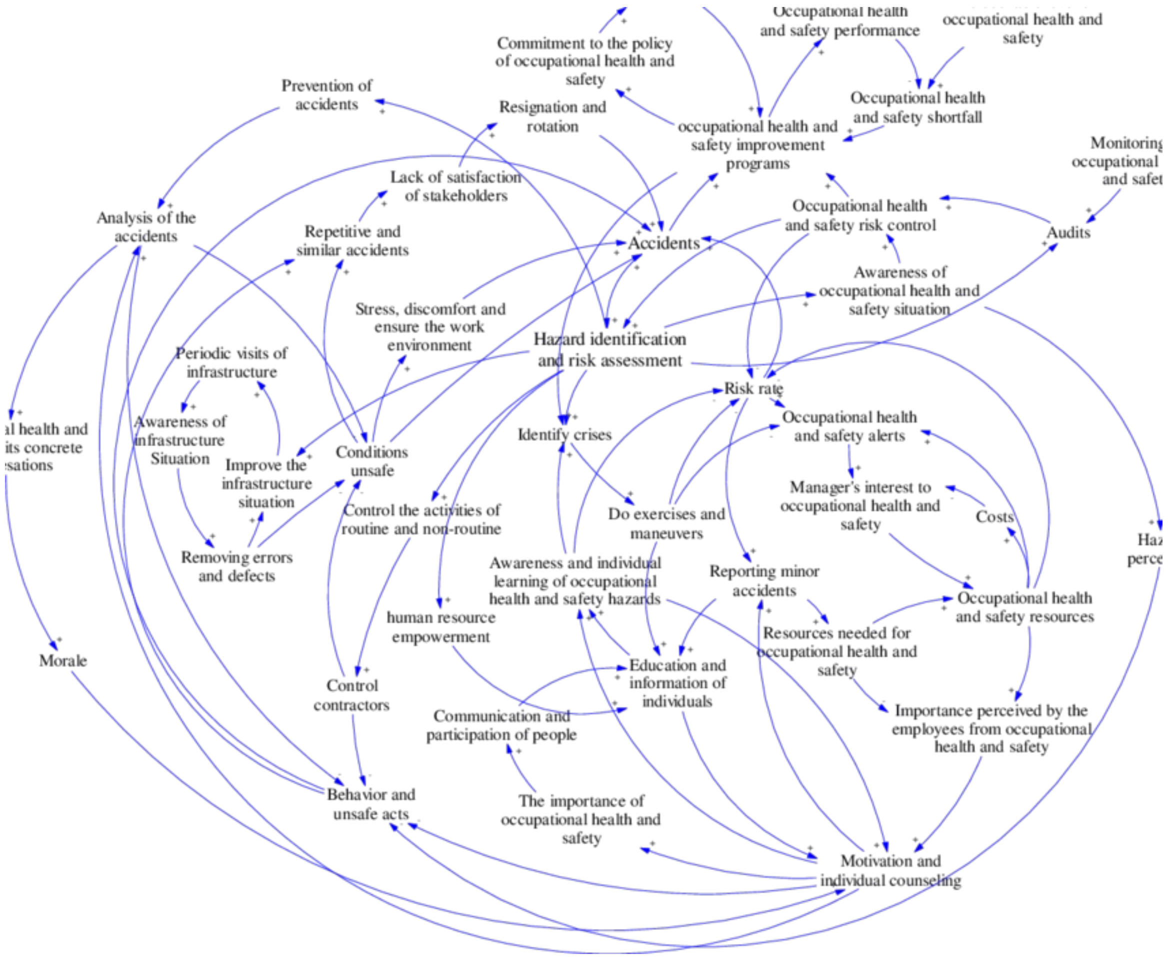
AUGMENTING BI TOOLS IS THE KEY

- BI Tools are great at correlation and filtering, sorting and visualizing
  - But they are best with structured data
- Content management tools are great at managing unstructured information
  - But they do not support BI/Analytics
- We need tools to sit in the middle

midwest architecture community collaboration

ONTOLOGIES, ARISTOTLE AND THE SEMANTIC WEB

- Doug Lenat and Tim Berners-Lee championed the Semantic Web (W3C)
- OWL/RDF is great for tagging web content, but not enough for:
  - Complex causal paths
  - Deep language understanding
  - Effectively marrying structured and unstructured content
- The seeds of the solution were already sprouting in Aristotle's time



ResearhGate

midwest architecture community collaboration

INFORMATION IS NOT ENOUGH TO DETERMINE CAUSE

- Data is flat, and not enough to discover correlations
  - **Information is needed**
- Relational models have more dimensions than flat models…
- but not enough dimensions to efficiently discover causes
  - **Knowledge is needed**
- Graph structured knowledge can support causal reasoning

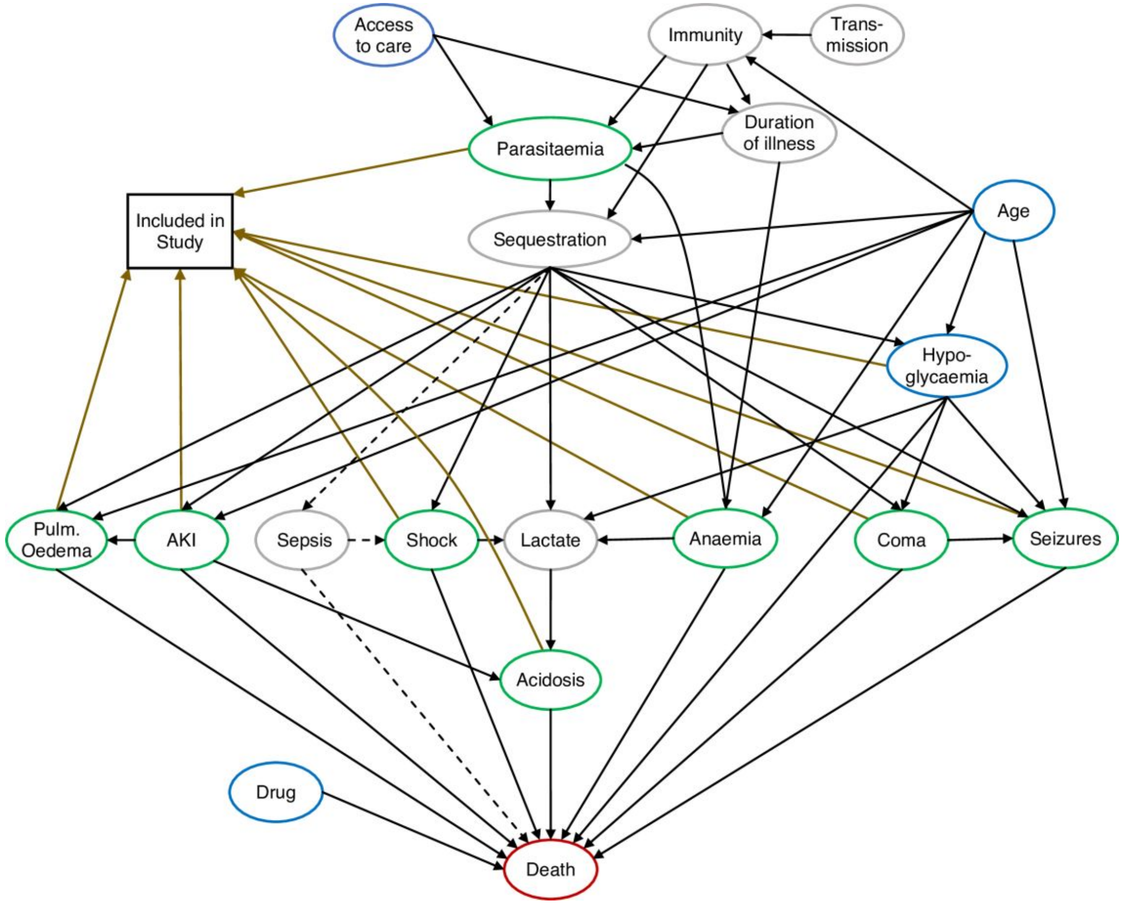
- Causal Paths are core to causal reasoning
  - Necessary but not sufficient

midwest architecture community collaboration

UNDERSTANDING CAUSAL PATHS

The diagram at right shows a causal network with multiple paths from tropical medicine studies on malaria

- The nodes are abbreviated, but understandable
- This shows "Confounders" in Age, Sequestration and Hypoglycemia
- It shows "Colliders" leading to several preliminary conditions and ultimately to Death
- It only tells the negative, omitting positive impacts of therapies or other conditions that lead to healing
- It doesn't show external influences such as congenital or acquired Hypoglycemia or other comorbidities
- Even incomplete, it is extremely useful



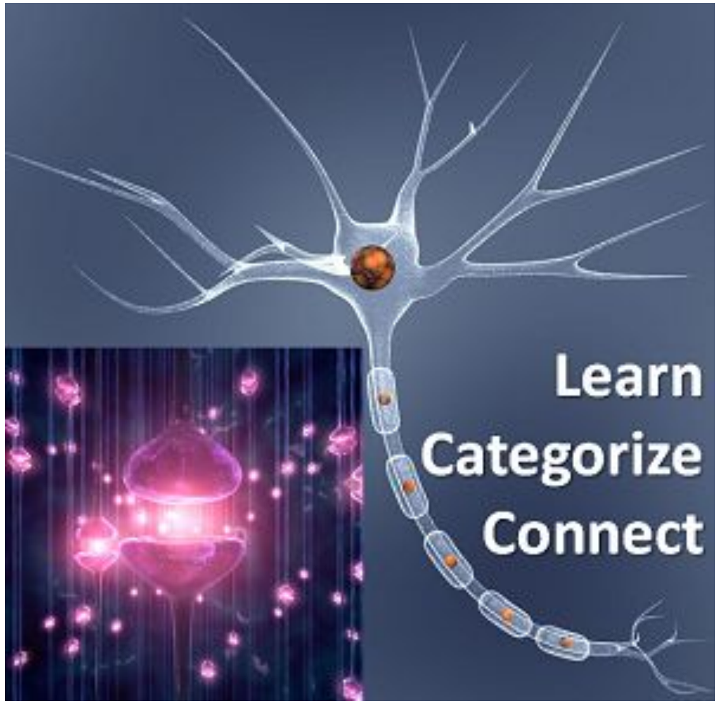
https://www.tropmedres.ac/team/james-watson/

midwest architecture community collaboration

KNOWLEDGE DIMENSIONS

Concept data Concept graph

- Causal data is needed to associate causes with possible outcomes: Presuppositions
- Taxonomical data is needed for inheritance
- Language data is needed for:
  - Concept Learning
  - Causal Path Construction
  - Solution Validation
- Time, location and other context for NLU



midwest architecture community collaboration

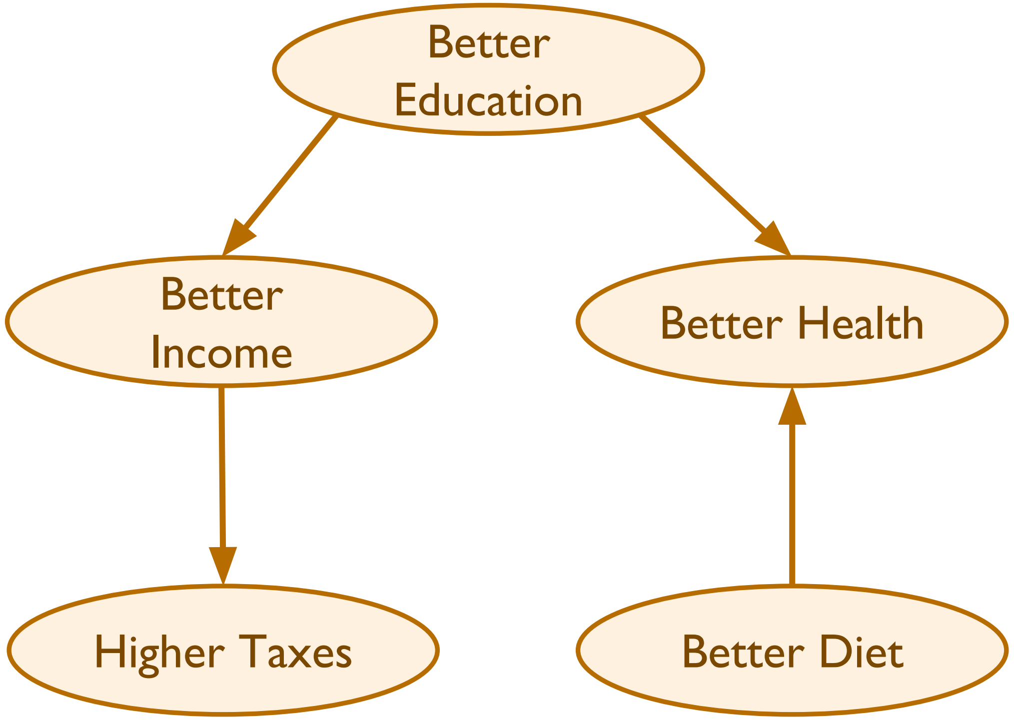
CHALLENGES TO CAUSAL INFERENCE

- Confounders and Colliders make causal inference more challenging
- Causal knowledge alone is often not enough
- Taxonomical, meronomical and space/time knowledge often serve critical roles
- The amount of concept knowledge needed is huge except for narrow domains
- Causal inference paths may be chaotic

midwest architecture community collaboration

COLLIDERS

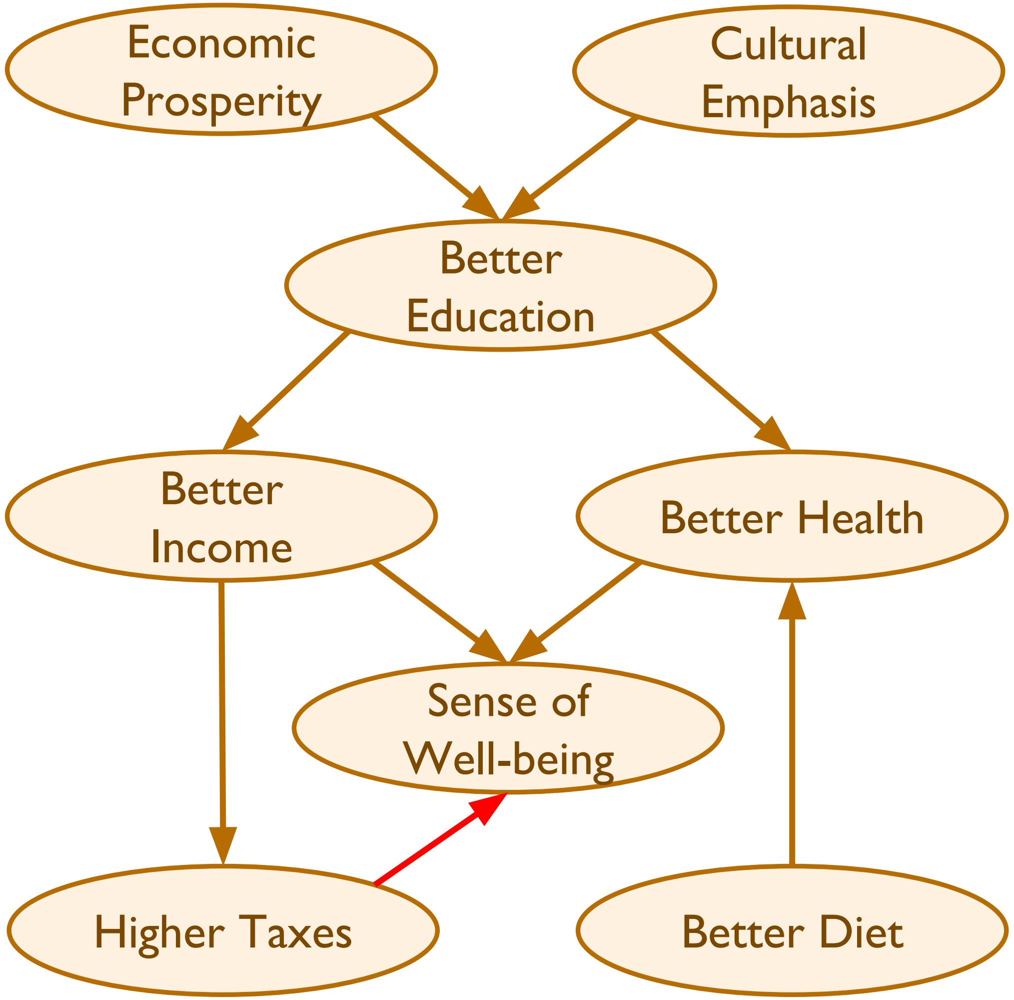
- Some outcomes have many contributing factors
- Both better education and diet improve health
- It may be that better education affects diet making it both a direct and intermediate cause
- Root causes may exist a-priori
  - Education may be impacted by location, prosperity or cultural factors



midwest architecture community collaboration

CONFOUNDERS

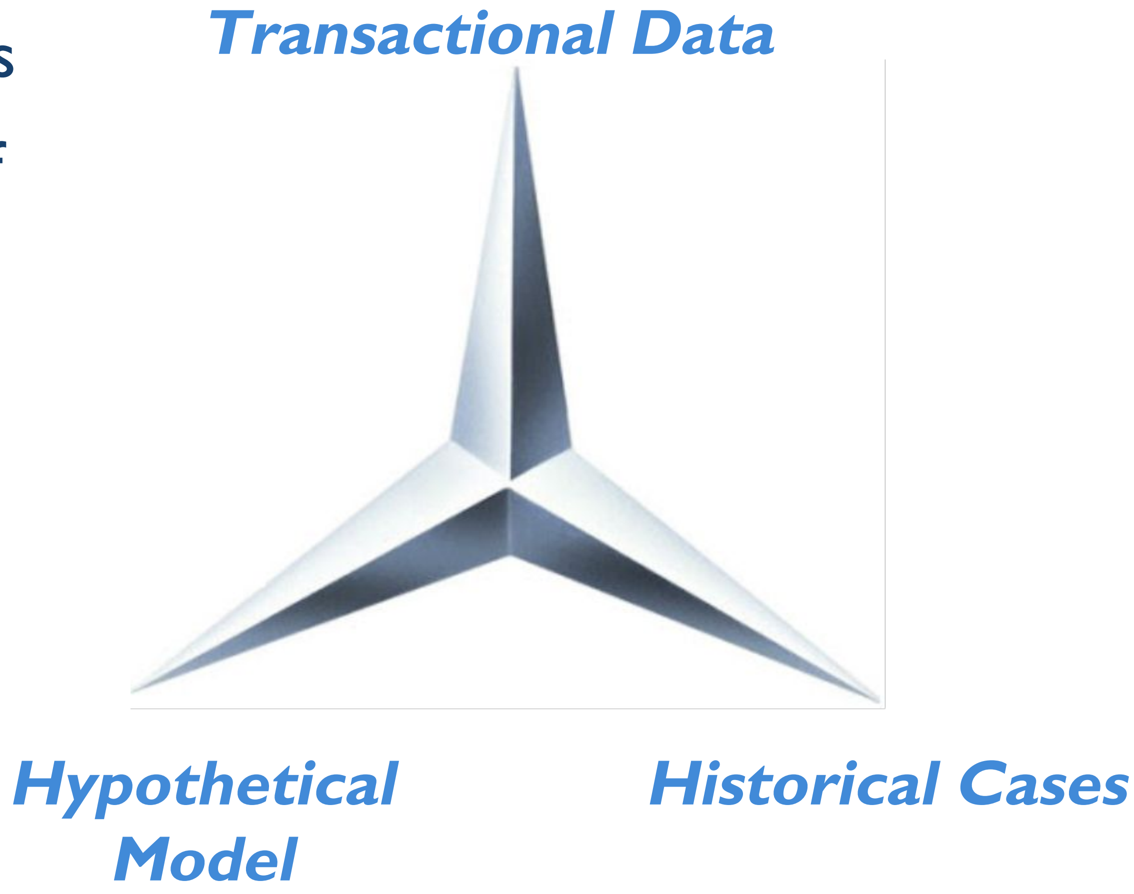
- One cause or causal factor may have many possible outcomes
  - Better education may improve several outcomes
- Causality also involves detractors
  - Higher taxes could impair a person's sense of well-being
- Computational models must support multiple modes of inference



midwest architecture community collaboration

THREE SOURCES

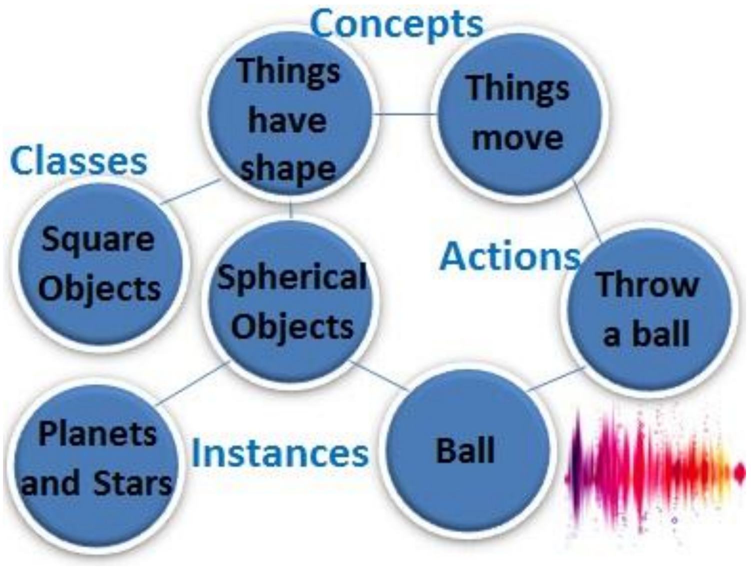
- Transactional data can provide input cases
- Hypothetical models support definition of presuppositions
- Historical Cases support both a-priori learning and a-posteriori validation
- Causal inference traverse paths so all three sources are best structured to support path navigation



midwest architecture community collaboration

RESOLUTION OF DEIXIS

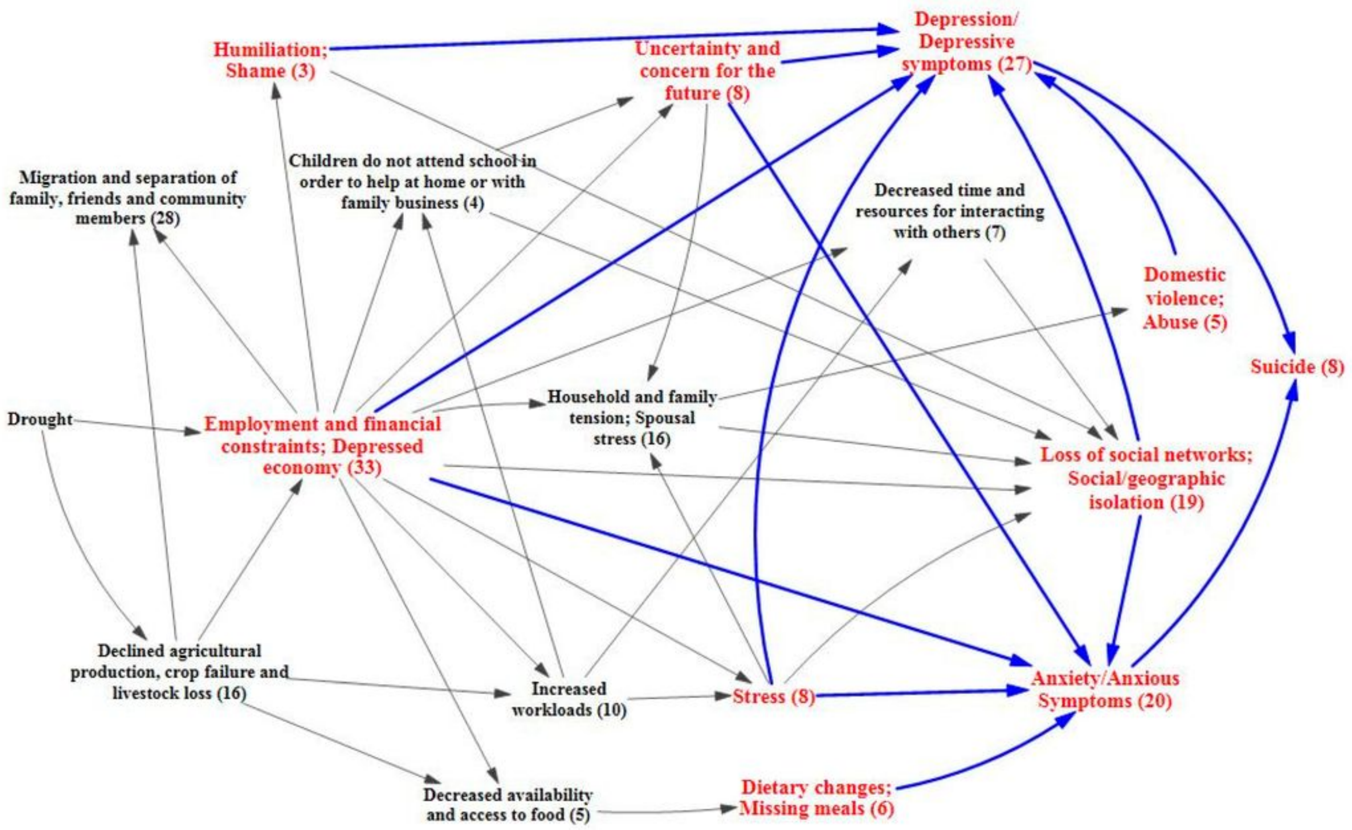
- Deixis, context and pragmatics are deep layers of natural language (NL)
- Presuppositions about what "can" happen can show what did happen or will happen
- Exformation is often needed for resolution
  - Characteristics of the subject like shape
  - Actions that can apply to the subject
- Logical syllogisms are key to inference



midwest architecture community collaboration

HYPOTHETICAL MODELS AND PRESUPPOSITION

- Simple hypothetical models are usually naïve and unreliable
- The more model data you can have, the better the inference
- Data mining requires some NL
- The more powerful the NLU the more accurate model data you can gather



midwest architecture community collaboration

RESOURCES

- The Book of Why: Judea Pearl, Basic 2018
- Wolfram Physics
- Pragmatics: Yan Huang, Oxford 2007
- The Semantic Web: Tim Berners-Lee, 2001
- The Grammar of Science: Karl Pearson, Black 1892
- Metaphysics: Aristotle, Self-Published 340 BC

midwest architecture community collaboration

midwest
architecture community
collaboration