

Cognitive Ecology: A Framework for Cognitively Heterogenous Agents.

Johannes Castner

Columbia University

September 30, 2013

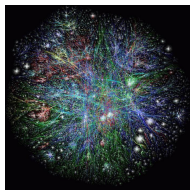


Table of contents

- 1 Motivation
- 2 Debt
- 3 A Modular Research Program
- 4 The Pieces (Modules)
- 5 Mental Models
- 6 Bayesian Beliefs
- 7 Cognition and Social Science
- 8 Measuring Cognitive Distance
- 9 Cognitive Diversity

Motivations and Questions

My Motivations are twofold:

- Short Run: theorize and show empirically that more Cognitive Diversity can be advantageous for an organization or a group.
- Long Run: help to build a cognitive foundation for the social sciences (theoretically and empirically), wherein individuals are really individuals and not clones of each other.

Cognitive Diversity vs. Diversity in Preferences

Economics and Political Science, are focused on differences in

- goals and preferences
- information asymmetries.

But what if information and goals are the same, but beliefs are heterogenous?

- when people work in a team

Or what if both, beliefs and goals are heterogenous?

- Is climate change policy a matter of goals or a matter of beliefs, or both?

“Cognitive Diversity” in teams.

For the moment, loosely define “Cognitive Diversity” as a quantitative measure of *within group* heterogeneity of belief systems (causal beliefs about the world).

The Questions are:

- Does greater “Cognitive Diversity” lead to better group decisions, when the goals of all members are well-matched; and how?
- How and under what forms of Belief Aggregation is “Cognitive Diversity” best utilized?

Condorcet's jury theorem

My work builds on Marquis de Condorcet's "Essay on the Application of Analysis to the Probability of Majority Decisions" (1785).

Assumptions:

- group wishes to reach a decision by majority vote.
- One of two potential outcomes of the vote is correct.
- each voter has an independent probability, p , of voting for the correct decision.

How many voters should we admit to the group?

Answer:

- if $p > \frac{1}{2}$ admit as many voters as possible (in the limit the probability that the majority decision is correct is 1)
- if $p < \frac{1}{2}$ admit only one voter!

Beyond the Condorcet's Jury Theorem

Condorcet's Jury Theorem is one-dimensional and it doesn't include Diversity (i.e. the only diversity is coming from the independents of each person's probability of being right).

- For n people, the probability of k being correct is:

$$Pr(k) = \frac{n!}{k!(n-k)!} * p^k(1-p)^{n-k}$$

- So we need to calculate

$$Pr(k > \frac{1}{2} * n) = \sum_{k > \frac{1}{2} * n} \frac{n!}{k!(n-k)!} * p^k(1-p)^{n-k}$$

- But also: it is a one-shot decision with no learning.

What if an organization can weigh the votes after learning something about the accuracy of each belief?

Suppose we have a partial ranking $p_1 \geq p_i \geq \dots \geq p_n$ over at least some of the n beliefs, based on evidence?

Model derived multi-demsional joint-probabilities

Main departures from Condorcet's Jury Theorem:

- Probabilities are over vectors of variables, v_1, \dots, v_n .
- Joint-distributions come from causal mental models of the world
- Mental models can be compared with each other
- and evaluated against data (using Bayesian Updating)
- Mental models can evolve over time
- The theory is about *Diversity* and not just *Quantity*

In fact, if two identical models are present, the Diversity is no greater than if only one of the two are present.

Thus, my work is very different from the Condorcet's Jury Theorem after all.

The Giants on whose shoulders I'm trying to stand

- Robert Axelrod and collaborators (Structure of Decision, 1976), Cognitive Maps.
- Cognitive Scientists Tom Griffiths (Berkeley) and Joshua Tenenbaum (MIT) for Bayesian Models of Causal Cognition.
- Scott Page (Michigan University) has first built models to show the importance of cognitive diversity. His work inspired mine.

A Modular Research Program

We have now assigned to reason and experience their place within the system of theoretical physics. Reason gives the structure to the system; the data of experience and their mutual relations are to correspond exactly to consequences in the theory. On the possibility alone of such a correspondence rests the value and the justification of the whole system, and especially of its fundamental concepts and basic laws.

-Albert Einstein

My overall research goals are big, but each piece is small. Each theoretical piece must have the possibility of an empirical counterpart.

The Pieces (Modules)

My research approach is modular and here are some of the modules (the first three are foundations, layed by others):

- What are Mental Models and how can they be elicited from people through interviews and texts? (Axelrod et. al 1976)
- Encoding Mental Models as Bayesian Networks (Joint Prior Distributions). (following Tennenbaum and Griffiths, 2003, 2004, 2005)
- Parameterizing Prior Distribution (Pearl, 1988).
- Bayesian updating and a Theory of Social Influence (A sketch).
- Measures of Cognitive Distance and Cognitive Diversity.

Mental Models (Cognitive Maps)

Mental Models of interest here are Causal Models over some restricted domain (Axelrod et. al 1976).

For Macro-Economists, for example, the universe of interest includes variables such as

- Interest Rates,
- Unemployment,
- GDP,
- Inflation etc.

Models could be Neo-classical and Keynesian, for example, but many more models are possible and in fact exist in the minds of policy makers (Neo-classical and Keynesian are stereotypes).

Cognitive Maps as Directed Signed Graphs

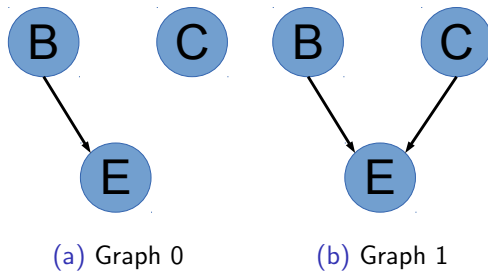


Figure: Two examples of Simple Cognitive Maps

Here “B” stands for Background Cause and “C” stands for Cause (the causal variable believed in by actor 1, but not by actor 0).

Non-Linearities

There are two seeming restrictions:

- A relational restriction: “A has a positive/negative causal effect on B ”,

The Problem:

When relations are non-linear (as in the Laffer Curve).

The Solution:

The statement “Taxes have a negative effect on output” has to be interpreted as: “the instantaneous effect of taxes on output is negative, conditional on the current level of taxes and output.”

Causal Loops

- A structural restriction: Belief Systems can not include causal cycles

The Problem:

When someone expresses beliefs in equilibria (negative causal feedback cycles) as most economists do, or in runaway effects (positive causal cycles), as some climate scientists do.

The Solution:

Including a time dimension (with possibly very short time-intervals).

$$A_{t=0} \rightarrow B_{t=1} \rightarrow A_{t=2} \rightarrow B_{t=3} \rightarrow \dots$$

Eliciting Cognitive Maps: Empirical Data on Beliefs

How beliefs are elicited and Cognitive Maps are constructed from statements is detailed in “Structure of Decision: The Cognitive Maps of Political Elites” (1976), edited by Political Scientist Robert Axelrod.

Basic Idea:

- Either use interview or use transcriptions of speeches/debates
- get statements of the form: CO₂ causes Climate Change
- encode as directed signed arc of the graph representing a particular actor's belief system:

CO₂ →⁺ Climate-Change

Encoding Cognitive Maps as Bayesian Networks

Why Bayesian Networks? Current and planned work:

- For a natural dynamic theory of belief updating and causal learning (Griffiths and Tennenbaum).
- For a theory of social (media) influence, where belief systems are well defined objects.
- For meaningful measures of Cognitive Distance and Cognitive Diversity.
- For theories of Belief Aggregation and Belief Synthesis
- For a theory of Belief System Evolution (including random mutations in the learning process to explore the belief space)

Bayesian Networks (together with Bayesian Updating, or Biased Bayesian Updating) are very useful for modeling dynamic processes involving belief systems.

Some Foundations

The following introduction to the Bayesian Network representation of individual Belief Systems closely follows:

Griffiths, T. L., Kemp, C., and Tenenbaum, J. B. (2008).
“Bayesian models of cognition.” In Ron Sun (ed.), The Cambridge handbook of computational cognitive modeling. Cambridge University Press.

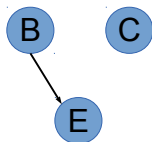
Encoding Cognitive Maps as Bayesian Networks

The problem of learning the “correct” causal graphical model has two aspects:

- Model Selection, and
- Parameter Estimation.

Where, if we have someone’s Cognitive Map at a given time, we have his “class of models,” that is, we know the class of models that the actor has *selected* (based on previous deliberation and experience) up to the particular “parameterization”, which we do *not* know.

Encoding Cognitive Maps as Bayesian Networks

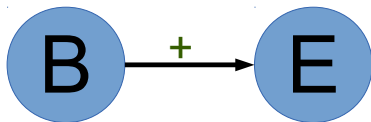


Graph₀ is encoded as follows as a Bayesian Network (or simply joint distribution). The joint distribution of any k variables ($k = 3$ in this case) can be written as the product of all of its conditionals and its marginals. In the case of Graph₀:

$$P_0(B, C, E) = P(E|B) * P(B) * P(C)$$

Note that since B is believed to cause E , the value of E is believed to depend on the value of B and thus the term $P(E|B)$ is included. However, E is independent from C and thus $P(E|B, C)$ collapses, while in Graph₁ the term would have to be $P(E|B, C)$.

Encoding a Causal Arc



For a positive causal relation, where B is the cause of E, we have

$$P(E = 1|B = 1) > P(E = 1|B = 0)$$

for binary random variables B and E. For this discussion, I assume that all variables are binary, but similar encoding can be done for numerical variables of all sorts (discrete or continuous).

Contingency Tables

Contingency Table Representation used in Elemental Causal Induction:

	Effect Present (e^+)	Effect Absent (e^-)
Cause Present (c^+)	$N(e^+, c^+)$	$N(e^-, c^+)$
Cause Absent (c^-)	$N(e^+, c^-)$	$N(e^-, c^-)$

Jenkins and Ward (1965) suggest that the degree of causation is best measured by the quantity (estimated from data using contingency counts, $N(\cdot)$):

$$\Delta P = \frac{N(e^+, c^+)}{N(e^+, c^+) + N(e^-, c^+)} - \frac{N(e^+, c^-)}{N(e^+, c^-) + N(e^-, c^-)} =$$

$$P(e^+, c^+) - P(e^+, c^-)$$

ΔP and Causal Power (Updating Rules)

Remember (previous slide) that

$$\Delta P = P(e^+, c^+) - P(e^+, c^-)$$

Cheng (1997) suggested that peoples judgments are better captured by “causal power”:

$$Power = \frac{\Delta P}{1 - P(e^+|c^-)},$$

which predicts that ΔP will have a greater effect when $P(e^+|c^-)$ is large.

Parameterization of Beliefs, the trivial case

We need to define the form of the conditional probability distribution, $P(E|B, C)$, which is called the “parameterization of the Graph.”

For Graph_0 this can simply be done with one parameter:

$P_0(E = e^+|B = b, \omega_{0,0}) = \omega_{0,0}b$, where $b \in \{0, 1\}$.

$$\Delta_b P_0 = P_0(e^+, b^+) - P_0(e^+, b^-) = \omega_{0,0}$$

Parameterization of Beliefs, the less trivial case

For Graph_1 we could also write

$$\Delta_b P_1 = P_1(e^+, b^+) - P_1(e^+, b^-) = \omega_{1,0}$$

and

$$\Delta_c P_1 = P_1(e^+, c^+) - P_1(e^+, c^-) = \omega_{1,1}.$$

But can we write:

$$P(E|B, C) = \omega_{1,0} * B + \omega_{1,1} * C,$$

as in a simple, linear regression?

Parameterization of Beliefs, the less trivial case

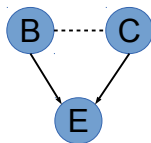
Can we write $P(E|B, C) = \omega_{1,0} * B + \omega_{1,1} * C$?

We know that $0 < P(E|B, C) < 1$, because it is a probability. But then,

$$0 < \omega_{1,0} + \omega_{1,1} < 1,$$

which is a dependence between the parameters $\omega_{1,0}$, $\omega_{1,1}$ that didn't come explicitly out of the beliefs.

The “Noisy-OR” Parameterization (Pearl, 1988)



To break this artificial dependency, the “Noisy-OR” parameterization (Pearl, 1988) is used:

$$P_1(e^+|b, c, \omega_{1,0}, \omega_{1,1}) = 1 - (1 - \omega_{1,0})^b * (1 - \omega_{1,1})^c$$

which breaks the dependency between $\omega_{1,0}$ and $\omega_{1,1}$.

The “Noisy-OR” Parameterization

$$P_1(e^+|b, c, \omega_{1,0}, \omega_{1,1}) = 1 - (1 - \omega_{1,0})^b * (1 - \omega_{1,1})^c,$$

gives that

$$P_1(e^+|B = 1, C = 0, \omega_{1,0}, \omega_{1,1}) = \omega_{1,0},$$

$$P_1(e^+|B = 0, C = 1, \omega_{1,0}, \omega_{1,1}) = \omega_{1,1}$$

and

$$P_1(e^+|B = 1, C = 1, \omega_{1,0}, \omega_{1,1}) = \omega_{1,0} + \omega_{1,1} - \omega_{1,0} * \omega_{1,1}.$$

If $\omega_{1,0} = \omega_{1,1} = 1$, the “Noisy-OR” becomes the *or*-function:
 $E = 1$ whenever $B = 1$, or $C = 1$ or both, $E = 0$ otherwise.

Psychology and parameterizing beliefs

Two important theories as to how people update their causal beliefs:

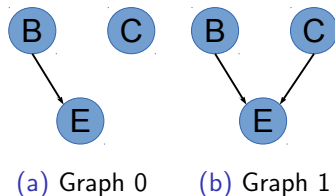
- $\Delta P = P(e^+|c^+) - P(e^+|c^-)$
- Causal Power ($\frac{\Delta P}{1 - P(e^+|c^-)}$).

Tenenbaum and Griffiths (2001, 2005) showed that both psychological theories correspond to maximum-likelihood estimates of the causal strength parameter $\omega_{1,1}$:

- ΔP results from assuming the linear parameterization
- and Causal Power results from the “Noisy-OR” parameterization.

As the “Noisy-OR” parameterization is logically more sound, I will use a Causal Power Likelihood, along with a “Noisy-OR” parameterization.

Two Cognitive Maps as two Hypotheses



If we have some data of the world, D , we can write the *posterior odds* as follows:

$$\frac{P(\text{Graph}_1|D)}{P(\text{Graph}_0|D)} = \frac{P(D|\text{Graph}_1)}{P(D|\text{Graph}_0)} \frac{P(\text{Graph}_1)}{P(\text{Graph}_0)}$$

Likelihood ratio and Prior Odds Ratio

$$\frac{\overbrace{P(\text{Graph}_1|D)}^{\text{Posterior Odds}}}{\overbrace{P(\text{Graph}_0|D)}^{\text{Posterior Odds}}} = \frac{\overbrace{P(D|\text{Graph}_1)}^{\text{Likelihood Ratio}}}{\overbrace{P(D|\text{Graph}_0)}^{\text{Likelihood Ratio}}} \frac{\overbrace{P(\text{Graph}_1)}^{\text{Prior Odds}}}{\overbrace{P(\text{Graph}_0)}^{\text{Prior Odds}}}$$

The Likelihood Ratio:

$$\frac{P(D|\text{Graph}_1)}{P(D|\text{Graph}_0)} = \frac{P_1(E = e|B = b, C = c) * P_1(B = b) * P_1(C = c)}{P_0(E = e|B = b) * P_0(B = b) * P_0(C = c)},$$

where I use the “Noisy-OR” parameterization of P_1 , P_0 (Pearl, 1988).

The Likelihood Function (updating)

The question that the Likelihood Function answers is the following:

Given our model of the world, how plausible is the data that we are experiencing?

For example:

$$P_1(B = 1, C = 1, E = 1) = \pi_{1,b} * \pi_{1,c} * (\omega_{1,0} + \omega_{1,1} - \omega_{1,0} * \omega_{1,1}),$$

where $\pi_{1,b}$ and $\pi_{1,c}$ are the parameters associated with the marginal probabilities of our “un-caused” variables, $P_1(B = 1)$ and $P_1(C = 1)$.

The Likelihood Ratio

The Likelihood Ratio:

suppose $D = (B = 1, C = 1, E = 1)$ then the **Likelihood Ratio** is:

$$\frac{P_1(B = 1, C = 1, E = 1)}{P_0(B = 1, C = 1, E = 1)} = \frac{\pi_{1,b} * \pi_{1,c} * (\omega_{1,0} + \omega_{1,1} - \omega_{1,0} * \omega_{1,1})}{\pi_{0,b} * \pi_{0,c} * \omega_{0,0}}.$$

But note that $P_1(B = 0, C = 1, E = 1) = \pi_{1,c} * \omega_{1,1} > 0$, while $P_0(B = 0, C = 1, E = 1) = \pi_{0,c} * 0 = 0$, so that the Likelihood Ratio is undefined when $D = (B = 0, C = 1, E = 1)$!

In that case, we just say that the ratio is infinite and Model_1 , explains that data infinitely better than Model_0 , no matter what the prior was.

The Prior and Posterior Odds as Basis for a Theory of Trust and Social Influence

Suppose that you have two friends, call them 1, and 0, endowed with beliefs specified by Graph_1 and Graph_0 respectively.

If you trust 1's opinion more than 0's opinion, then

$P(\text{Graph}_1) > P(\text{Graph}_0)$ and the prior odds ratio of the two graphs is greater than 1. Otherwise, if you trust 0 more than 1, the prior odds ratio of their two models (if 1 is in the numerator) is a positive fraction.

Persuasion Processes

The above suggests how the modeling of belief updating (learning) can be made to include social influence in a straight forward and natural way.

Updating uses data to evaluate ones own and others opinions, particularly those peoples' opinions with whom the actor has contact.

And of course, media influence can be studied using this framework!

$$\frac{\overbrace{P(\text{Graph}_1|D)}^{\text{Relative Trust}_{1,0,t=1}}}{P(\text{Graph}_0|D)} = \frac{\overbrace{P(D|\text{Graph}_1)}^{\text{New Evidence}_{1,0}}}{P(D|\text{Graph}_0)} \frac{\overbrace{P(\text{Graph}_1)}^{\text{Relative Trust}_{1,0,t=0}}}{P(\text{Graph}_0)}$$

By comparing ones own model and ones friends' models, in light of data, one might be swayed over time to find a more likely synthesis. This swaying process is made precise with Bayesian model comparison, where initial biases are expressed as prior odds.

Biases and Evidence

Again, we have:

$$\overbrace{\frac{P(\text{Graph}_1|D)}{P(\text{Graph}_0|D)}}^{\text{Posterior Odds}} = \overbrace{\frac{P(D|\text{Graph}_1)}{P(D|\text{Graph}_0)}}^{\text{Likelihood Ratio}} \overbrace{\frac{P(\text{Graph}_1)}{P(\text{Graph}_0)}}^{\text{Prior Odds}}.$$

Taking the log gives:

$$\overbrace{\log\left(\frac{P(\text{Graph}_1|D)}{P(\text{Graph}_0|D)}\right)}^{\text{Log Posterior Odds}} = \overbrace{\log\left(\frac{P(D|\text{Graph}_1)}{P(D|\text{Graph}_0)}\right)}^{\text{Log Likelihood Ratio}} + \overbrace{\log\left(\frac{P(\text{Graph}_1)}{P(\text{Graph}_0)}\right)}^{\text{Log Prior Odds}}.$$

While the Posterior Ratio ranges from 0 to ∞ , the Log Posterior Ratio ranges from $-\infty$ to ∞ and is 0 when both models are believed to be equally plausible (or equally implausible).

Causal Support

We say that Model_1 is better supported by D than Model_0 whenever:

$$\log \left(\frac{P(D|\text{Graph}_1)}{P(D|\text{Graph}_0)} \right) > 0$$

and this is why, in the context of Causal Models, Tenenbaum and Griffiths renamed the Log Likelihood Ratio the “Causal Support” of Model_1 against Model_0 .

The **magnitude** of the **difference in explanatory power** matters and is meaningful.

Calculating “Causal Support”

We have that the “Causal Support” is

$$\text{Causal Support} = \log \left(\frac{P(D|\text{Graph}_1)}{P(D|\text{Graph}_0)} \right).$$

Not wanting to introduce artificial differences (that we don't know from data), we assume:

$$\pi_{1,b} = \pi_{0,b} \text{ and } \pi_{1,c} = \pi_{0,c},$$

so that these parameters cancel out.

Integrating over the Parameters

And then we integrate out the causal parameters (assuming that they come from a uniform distribution on the interval $[0, 1]$):

$$P(D|\text{Graph}_1) = \int_0^1 \int_0^1 P_1(D|\text{Graph}_1, \omega_{1,0}, \omega_{1,1}) P(\omega_{1,0}, \omega_{1,1}|\text{Graph}_1) d\omega_{1,0} d\omega_{1,1}$$

and

$$P(D|\text{Graph}_0) = \int_0^1 P_0(D|\text{Graph}_0, \omega_{0,0}) P(\omega_{0,0}|\text{Graph}_0) d\omega_{0,0}$$

to obtain

$$\text{Causal Support} = \log \left(\frac{\int_0^1 \int_0^1 P_1(D|\text{Graph}_1, \omega_{1,0}, \omega_{1,1}) P(\omega_{1,0}, \omega_{1,1}|\text{Graph}_1) d\omega_{1,0} d\omega_{1,1}}{\int_0^1 P_0(D|\text{Graph}_0, \omega_{0,0}) P(\omega_{0,0}|\text{Graph}_0) d\omega_{0,0}} \right)$$

Expected Causal Support when Model₁ is true

If we repeatedly generate data from $P_1(B, C, E)$ and calculate the Causal Support each time, we will get the Kullback-Leibler divergence (also Information Divergence, Information Gain, Relative Entropy, or KLIC):

$$D_{KL}(P_1||P_0) = E_1 \left(\frac{P_1(D)}{P_0(D)} \right), \text{ with } D \sim P_1,$$

where $E_1(\cdot)$ is the expectation operator under Graph₁ (not the effect variable).

But we know that $D_{KL}(P_1||P_0) = \infty$, because Graph₀ puts zero probability on $D = (B = 0, C = 1, E = 1)$, which will in expectation be drawn $\pi_{1,1} * \omega_{1,c} * N$ times for every N draws.

Cognitive Distance

For a measure of “Cognitive Distance” between beliefs P_1 and P_0 , we would naively want something in the form of

$$\lambda D_{KL}(P_1||P_0) + (1 - \lambda)D_{KL}(P_0||P_1)$$

with which we would hope to measure the average distance between the models when the prior probabilities of Model₁ and Model₀ are λ and $(1 - \lambda)$ respectively.

NOT useful as a general difference measure because it is infinite for many pairs of models.

λ divergence and Jensen-Shannon divergence

λ Divergence:

$$D_{\lambda} = \lambda * D_{KL}(P_1 || \lambda * P_1 + (1 - \lambda) * P_0) + (1 - \lambda) * D_{KL}(P_0 || \lambda * P_1 + (1 - \lambda) * P_0),$$

where the prior probabilities of Model₁ and Model₀ are λ and $(1 - \lambda)$ respectively.

$$\text{with } 0 < \lambda < 1 \Rightarrow 0 < D_{\lambda} < \infty.$$

For $\lambda = \frac{1}{2}$ this gives the Jensen-Shannon Divergence:

$$D_{JS} = \frac{1}{2} D_{KL}(P_1 || M) + \frac{1}{2} D_{KL}(P_0 || M),$$

where $M = \frac{1}{2} P_1 + \frac{1}{2} P_0$.

Jensen-Shannon divergence as a measure of Cognitive Distance

While for a theory of trust $\lambda \neq \frac{1}{2}$ has an obvious meaning, for cognitive distance the Jensen-Shannon divergence can be interpreted as the total divergence to the average, or the Information Radius (IRad). It is symmetric (because $\lambda = \frac{1}{2}$) and always finite. In fact, using the base 2 logarithm:

$$0 \leq D_{JS}(P_1||P_0) \leq 1.$$

What this means for you (a theorist). A possible scenario in the near future:

- A colleague asks you “before you start, can you tell me what difference adding your variable will likely make?”
- You answer, “hold on, let me just calculate the Cognitive Distance between our two models (you don’t need data for that)!”

Properties of $\sqrt{(D_{JS}(P||Q))}$

$\sqrt{(D_{JS}(\cdot))}$ is a metric $\Rightarrow \forall$ Joint Distributions P, Q, R

- $\sqrt{(D_{JS}(P||Q))} \geq 0$ (non-negativity).
- $\sqrt{(D_{JS}(P||Q))} = 0$ iff $P = Q$. (Coincidence Axiom)
- $\sqrt{(D_{JS}(P||Q))} = \sqrt{(D_{JS}(Q||P))}$ (Symmetry)
- $\sqrt{(D_{JS}(P||R))} \leq \sqrt{(D_{JS}(P||Q))} + \sqrt{(D_{JS}(Q||R))}$ (Triangle Inequality)

The Jensen-Shannon Divergence measures how reliably one can decide which one of two joint distributions generated a piece of data, given that we know that it must have been one of them. Thus, this measure ($\sqrt{(D_{JS}(\cdot))}$) is theoretically meaningful and it satisfies all that we need from a measure of Cognitive Distance.

Cognitive Diversity

Martin Weitzman proposes a recursion formula to calculate the Diversity of a set of objects, using the pairwise distances between the elements of the set.

- As long as we have a set of Belief Systems, Ω with finite cardinality $n + 1$
- and a metric we call “Cognitive Distance”

$$(CD(P||Q) = \sqrt{(D_{JS}(P||Q))})$$

we can calculate the “Cognitive Diversity” of *the Organization*, Ω .

The Weitzman Algorithm for “Cognitive Diversity”

The Algorithm:

- ① let $W = 0$ be the initial Diversity.
- ② Take a random belief system, P_0 , from the organization Ω .
- ③ Remove P_0 from Ω and add it to a previously empty set, S
- ④ Find the belief system P_j , such that
 $CD(P_0||P_j)) = \min(CD(P_0||P_1), \dots, CD(P_0||P_n)),$ for
 $P_1, \dots P_n \in \Omega$
- ⑤ Remove P_j from Ω , add P_j to S and increment W by
 $CD(P_0||P_j)$
- ⑥ Find $P_i \in S, P_j \in \Omega$ such that $CD(P_i||P_j) = \min(CD(P_i||P_j)),$
 $\forall P_i \in S, \forall P_j \in \Omega$ and repeat (5) until Ω is empty.

The algorithm could be altered to compare each new graph with the mixture of those that are already in S .

Properties of the Weitzman Measure

Intuitively, the addition of belief system P_j , to the collection Ω should increase W by its distance to the collection.

$$CD(P_j||\Omega) = \min_{P_i \in \Omega} (CD(P_j||P_i))$$

- Monotonicity in Belief Systems:

$$W(\Omega) \geq W(\Omega \setminus P_i) + CD(P_i||\Omega), \forall P_i \in \Omega$$

- The Twin Property: The addition of an element that is identical to an element in Ω does *not* increase W .
- Continuity in Distances: small changes in the pairwise distances result in small changes in W .
- Monotonicity in Distances: if every pairwise distance is increased, W increases.