

# Religious Mental Structures Counterintuitiveness represented in Optimality Theory (OT)

TAMÁS BIRÓ

Amsterdam Center for Language and Communication, University of Amsterdam, The Netherlands

t.s.biro@uva.nl, <http://www.biro.t.hu/events/OTGCA/>, <http://www.csr-arc.com/>

CSS Workshops on "OT as a General Cognitive Architecture" and "New Developments in the Cognitive Science of Religion"

## Cognitive Science

**GOAL:** Understand data structures and algorithms used by mind/brain.

**Data structures:** numbers, strings, graphs, feature matrices, distributed activation patterns, etc.  
We use: **AVM = attribute-value matrices** (originally employed in linguistics, viz. HPSG and LFG).

**Algorithms:** rewrite rules, symbol manipulating programs, differential equations, activation spreading, etc.  
We use: **OT = Optimality Theory** (Prince and Smolensky 1993/2004):  
Ranked *constraints* pick the best one among the *candidates*.

An over-simplified example: *Different languages have different stress patterns:*

/amerika/	NOFINALSTRESS	LATESTRESS	EARLYSTRESS
[ámérika]	good	worst	best
[ámérika]	good	bad	good
[ámérika]	good	good	bad
[ámériká]	bad	best	worst

Constraints are ranked into hierarchies. The highest ranked one is applied first; if draw, then second is applied; etc. In turn, different hierarchies yield different language types:

Word initial stress languages [ámérika]: EARLYSTRESS ≫ LATESTRESS, NOFINALSTRESS  
 NOFINALSTRESS ≫ EARLYSTRESS ≫ LATESTRESS  
 Word final stress languages [ámériká]: LATESTRESS ≫ EARLYSTRESS, NOFINALSTRESS  
 LATESTRESS ≫ EARLYSTRESS, NOFINALSTRESS  
 Penultimate stress languages [ámérika]: NOFINALSTRESS ≫ LATESTRESS ≫ EARLYSTRESS  
 Second syllable stress languages [ámériká]: No such hierarchy.  
 Correct prediction: (almost) no such language among languages of the world.

## PROPOSAL:

Apply these data structures and algorithms to models of religions, too.

## Cognitive Science of Religion

**GOAL:** Understand the religious concepts in the human mind/brain.

Religious concepts are **counterintuitive**: they violate *ontological expectations* (P. Boyer 1994, 2001):

- **Ontological categories**, such as *human, animal, plant, object, artifact* (Keil 1989).
- **Folk-theories**, introducing ontological expectations (constraints), such as
  - *folk-physics* (applies to all categories):
    - VISIBLE: Objects are visible.
    - LOCALITY: Objects at one place at a time.
  - *folk-biology* (applies to living categories):
    - NEEDFOOD: Living objects die, unless fed.
    - BEGETSAME: Progeny belongs to same species.
  - *folk-psychology* (applies to *human* agents):
    - FINITEPOWER: Agents have restricted power.
    - NOOMNISCIENCE: Knowledge is restricted.
- Gods, spirits, ancestors: **counterintuitive agents**.
  - ‘Fast and frugal heuristics’ (Gigerenzer et al. 1999), supporting fast computation (hence, adaptive).
  - Natural objects expected to satisfy constraints.
  - “Supernatural”, a.k.a. *counterintuitive* concepts are those that violate these constraints.
  - Concepts with a low-level of counterintuitiveness (“minimally counterintuitive”) are most memorable. Hence, cultural transmission: idea survives, and becomes cross-culturally recurrent.
  - Concepts with a high-level of counterintuitiveness (“maximally counterintuitive”) are prone to change.

## Open issues:

- What makes something “maximally counterintuitive”? Where is the borderline between “minimal counter-intuitiveness” and “maximal counterintuitiveness”? How to measure counterintuitiveness?
- Do factors dependent on specific cultures influence memorability/the level on counterintuitiveness?

## PROPOSAL:

View counterintuitiveness as the violation of OT-like constraints.

## A counterintuitive AVM representation and Optimality Theoretical constraints

**Candidates** are AVM structures:

- Either objects, agents, etc. belonging to one of the five ontological categories, with attribute-value pairs encoding their properties;
- Or events and actions, with attribute-value pairs corresponding to their semantic arguments: agent, patient, instrument, location, time, etc.

An omnipotent, omniscient, omnipresent and invisible counterintuitive agent (deity) that requires feeding (sacrifices) can be represented as the following attribute-value matrix (AVM):

<i>agent</i>	
KNOWLEDGE	<i>all</i>
COLOR	<i>invisible</i>
LOCATION	<i>all</i>
POWER	<i>infinite</i>
...	

**Constraints** include:

- ‘Markedness constraints’ punishing violations of the universal ontological expectations.
- ‘Input-output faithfulness constraints’ punishing divergence from input.
- ‘Output-output faithfulness constraints’ punishing divergence from further, culturally acquired pieces of information (avoid discrepancy).

and it violates the constraints as:

LOCALITY	NEEDFOOD	NOOMNISCIENCE	VISIBLE	FINITEPOWER	...
<i>violates</i>	<i>satisfies</i>	<i>violates</i>	<i>violates</i>	<i>violates</i>	...

Since it violates (at least) some of the constraints, it is a counterintuitive concept.

## Mental algorithms processing an event description: theological incorrectness emerges in re-told stories

**Input = the story, as told by the experimenter:**

- G. saved a man’s life, *and at the same time* he helped a woman find her lost purse.
- Though G. was answering another prayer in another part of the world, *before long* G. saved his life.

**Output = the story, as re-told by atheist and theologically educated believer subjects:**

- G. saved a man’s life, *and only then* he helped a woman find her lost purse.
- “This story suggests that G. cannot listen to more than one prayer at a time.”

NB: In Pascal Boyer’s (2001) paraphrase of Justin Barrett (1996, unpublished thesis), the input contained simultaneity. In Barrett and Keil (1996), temporal relations in the input seemed to be rather vague. In either case, the subjects explicitly avoided simultaneity. This experiment proves the ranking LOCALITY ≫ FAITHFULNESS\_TO\_ORIGINAL\_STORY. To see why, observe the following OT tableau. The  $\mathfrak{E}$  symbol points to the optimal candidate.

/input story, with simultaneity, or vague temporal relations/	LOCALITY	FAITHFULNESS TO_ORIGINAL_STORY	FINITEPOWER	VISIBLE
$\left[ \begin{array}{l} \textit{saving action} \\ \text{AGENT} \left[ \begin{array}{l} \textit{agent} \\ \text{POWER} \textit{infinite} \\ \text{COLOR} \textit{invisible} \end{array} \right] \\ \text{PATIENT} \textit{A1} \\ \text{LOCATION} \textit{L1} \\ \text{TIME} \textit{T1} \end{array} \right] \ \& \ \left[ \begin{array}{l} \textit{helping action} \\ \text{AGENT} \left[ \begin{array}{l} \textit{agent} \\ \text{POWER} \textit{infinite} \\ \text{COLOR} \textit{invisible} \end{array} \right] \\ \text{PATIENT} \textit{A2} \\ \text{LOCATION} \textit{L2} \\ \text{TIME} \textit{T1} \end{array} \right]$	<i>violates !</i>	<i>satisfies</i>	<i>violates</i>	<i>violates</i>
$\mathfrak{E} \left[ \begin{array}{l} \textit{saving action} \\ \text{AGENT} \left[ \begin{array}{l} \textit{agent} \\ \text{POWER} \textit{infinite} \\ \text{COLOR} \textit{invisible} \end{array} \right] \\ \text{PATIENT} \textit{A1} \\ \text{LOCATION} \textit{L1} \\ \text{TIME} \textit{T1} \end{array} \right] \ \& \ \left[ \begin{array}{l} \textit{helping action} \\ \text{AGENT} \left[ \begin{array}{l} \textit{agent} \\ \text{POWER} \textit{infinite} \\ \text{COLOR} \textit{invisible} \end{array} \right] \\ \text{PATIENT} \textit{A2} \\ \text{LOCATION} \textit{L2} \\ \text{TIME} \textit{T2} \end{array} \right]$	<i>satisfies</i>	<i>violates</i>	<i>violates</i>	<i>violates</i>
... (candidate set must be worked out in a precise way)				

Note that FAITHFULNESS is ranked higher than the rest of the constraints. Were it not the case, we would expect subjects change the story to also satisfy these constraints.

Hypothesis: there is an universal ranking of the folk-theory constraints such that folk-physics ≫ folk-biology ≫ folk-psychology. The more fundamental experience a constraint is based on, the higher it is ranked.

## Conclusion: a novel view on counterintuitiveness thanks to Optimality Theory

We have presented a computable model with representations (data structures) and mechanisms (algorithms) applicable on these representations, in order to formalize a standard concept in the Cognitive Science of Religion:

- The advantage of using Optimality Theory is that constraints are soft, they can be violated, and hence, counterintuitive representations are allowed. And yet, they must be violated as little as possible.
- “Minimally counterintuitive”: optimal in an OT sense, with respect to a hierarchy. Open questions: (1) Is this hierarchy universal? (2) Place for culture-specific constraints? (3) Exact formulation of faithfulness constraints?
- “Maximally counterintuitive”: replaced by the human mind with an alternative, more harmonic candidate. Hence, such concepts will not survive. For instance, as observed in Justin Barrett’s experiments.
- Optimality Theory has been used for a number of cognitive domains (especially in linguistics). OT can be implemented in a connectionist network, and thus, argued to be cognitively plausible (Smolensky and Legendre 2006).

**Tamás Biró**

ACLCLC, University of Amsterdam

t.s.biro@uva.nl, <http://www.biro.t.hu/events/OTGCA/>, <http://www.csr-arc.com/>

## **Religious mental structures:**

### **Counterintuitiveness represented in Optimality Theory**

Two central topics all along the history of the cognitive sciences have been the “data structures” encoding information in the mind (graphs, strings, feature matrices, distributed activation patterns...), as well as the (discrete or continuous) algorithms manipulating these data structures. Despite the seminal work of Lawson and McCauley (1990), scholars in the cognitive science of religion have, thus far, lagged behind in developing formal models that apply the computer metaphor of the brain to religious phenomena.

The goal of my presentation is to lay down a formal model of counterintuitiveness, a central and heavily debated concept in the cognitive science of religion. The model will make use of *Optimality Theory* (OT), a linguistic model developed originally by Alan Prince and Paul Smolensky (1993), underpinned cognitively, philosophically and computationally by Smolensky and Legendre (2006). The discussion among CSR scholars on what counts as minimally counterintuitive – what is a cognitively optimal representation and what is doomed to oblivion – will be naturally reformulated in terms of OT. Counterintuitiveness will simply correspond to violating certain OT constraints. In particular, I will argue that a counterintuitive concept is cognitively optimal if it is a *locally optimal* representation with respect to the constraints.

The term “constraint” refers here to specific, formally defined OT constraints, which replace the general, loosely used notion of cognitive constraints. In a way analogous to Optimality Theory in linguistics, we introduce two kinds of constraints. *Markedness constraints* prefer certain (surface) structures over others. For instance, such constraints will be introduced by folk-theories, punishing representations that violate ontological expectations. Furthermore, there are also *faithfulness constraints*, struggling against changes in a representation. For instance, even if the representation of a counterintuitive concept violates some of the markedness constraints, it must fit into a general narrative: into a story, legend, myth, into a religious explanation, into the interpretation of a certain rite, and so on.

The advantage of using Optimality Theory is that constraints are soft, they can be violated, and hence, counterintuitive representations are allowed. And yet, they must be violated as little as possible, and the exact meaning of “as little as possible” is worked out in an exact way: constraints are ranked in a hierarchy, which determines the relative goodness (well-formedness) of the structures. For instance, the constraints introduced by folk-physics are ranked higher than those introduced by folk-psychology; specifically, constraint NOOMNIPRESENCE is stronger than constraints NOINVISIBILITY and NOOMNISCIENCE.

This model explains Justin Barrett’s experiments by developing the mental algorithm that replaces representations violating the constraints more than minimally by better representations. Additionally, Optimality Theory naturally invites us to formulate hypotheses regarding the acquisition of counterintuitive concepts and the cross-cultural typology of these concepts.