# Optimal religion: optimality theory accounts for ritual dynamics

Biró, T.S.

*Published in:*
Groningen studies in cultural change

# CHANGING MINDS

## *Religion and Cognition Through the Ages*

EDITED BY

István Czachesz and Tamás Biró

PEETERS
LEUVEN - PARIS - WALPOLE, MA
2011

CONTENTS

# OPTIMAL RELIGION

## OPTIMALITY THEORY ACCOUNTS FOR RITUAL DYNAMICS

Tamás Biró

### 1. *Introduction: possible cognitive approaches to religion*

#### 1.1 *How do linguists do cognitive science?*

Researchers in the cognitive sciences have in mind two very different realisations of a cognitive system:[1] the human brain and the intelligent computer. While the first dominates psychology, neurology or anthropology, the second plays a central role in industrially oriented artificial intelligence, such as robotics.

Linguistics exhibits an interesting trichotomy. Psycholinguists and neurolinguists focus on the brain's linguistic skills, while language technology develops industrial products. However, the third branch of linguistics does not aim at either of those realisations: mainstream theoretical linguistic research follows its own historically determined methodology with only very specific, often quite indirect connections to developments in other cognitive fields.[2] Rather than learning from other cognitive sciences, it developed into a quite peculiar cognitive discipline. This task-sharing allows for scholars with very diverse educational backgrounds and different institutional affiliations to work together efficiently. Notwithstanding frequent complaints about the lack of communication between different subfields, linguistics – despite its upbringing in the humanities – has become a cutting-edge discipline among the cognitive sciences.

Nevertheless, the 'cognitive' nature of modern theoretical linguistics is a complex issue. The path launched by Noam Chomsky is cognitive in the sense that it considers language as a biological phenomenon that is best described by some mathematical formalism and not so much as an arbitrary symbol system based on social conventions, as used to be seen by the structuralists in the first half of the twentieth century. Thanks to Chomsky's hypothesis about the independence of the language faculty, linguists have been

---

1 Cf. the notion of 'multiple realisation' in Ilkka Pyysiäinen's article in the present volume.

2 For more details, see John Nerbonne's article in the present volume.

and still are pursuing their own methodologies, including those adopted from pre-Chomskyan linguistics,[3] those borrowed from other cognitive sciences (and then significantly transformed),[4] and those newly developed in a cognitive scientific style.[5] Most contemporary linguists, even if not taking a theoretical stance on the question of the independence of the linguistic faculty, follow this methodology in practice. It was only in the 1970s and 1980s that an 'anti-Chomskyan' functionalist approach – featuring figures such as Charles Fillmore and George Lakoff – called *cognitive linguistics* emerged, which aimed at deriving linguistic phenomena directly from general cognitive capacities and from language use, without stipulating an abstract grammar in the brain.[6] Note that even these linguists are 'Chomskyan' in a broad sense: they view language as a cognitive function and employ cognitive (though, different) methodologies, as is reflected in the label of this school.

The cognitive science of religion (CSR), another field growing out of the traditional sphere of humanities, I conjecture, will also develop a similar trichotomy. This trichotomy may consist of a neurological-psychological line, a social engineering line (for example, CSR models supporting decisions on policies about fundamentalism), and an autonomous research line based on a combination of traditional methodologies used in religious studies (for example, anthropology, sociology of religion, textual criticism) with cognitive approaches. As the example of linguistics shows, this trichotomy best fits the educational background and the institutional embedding of the scholars involved. Each scholar has to find the methodology most suited to his or her personality, and therefore the wider the methodological scope, the more successful the cognitive science of religion will be in the near future.

In this paper, I am advocating a Chomskyan approach – even if the model I will introduce is based on a linguistic architecture that is independ-

---

[3] For example, the hierarchical constituents (e.g., phrases in a sentence) or the distinctive phonological features (such as [± voiced], [± nasal], [± rounded]) originate in the structuralist schools of the early twentieth century.

[4] Optimality Theory, the topic of the present contribution, originates in the field of neural networks, though linguists use it in a way that completely conceals its connectionist roots. It is important to note that neural networks left most of mainstream linguistics (and much of computational linguistics) absolutely untouched, despite important episodes such as the *past tense debate*.

[5] Most of the linguistic models belong to this category: Chomsky's Government and Binding, his Minimalist Program, or alternatives to Chomsky, such as LFG, HPSG, Lexical Phonology, Autosegmental Phonology, Government Phonology, and many, many others.

[6] For an introduction to this approach, which is not by coincidence very popular among scholars of the cognitive science of religion, see Croft and Cruse, *Cognitive Linguistics*.

ent of Noam Chomsky's *oeuvre* – in the sense that I consider religious phenomena in themselves in order to develop a formal model to describe them. The primary question is whether the model is able to account for the observed phenomena and – as with mainstream theoretical linguistics – the cognitive underpinning is only secondary.[7] This approach is different from much of the contemporary work in the cognitive science of religion which follows the anti-Chomskyan cognitive linguists in searching for direct connections between the domain of research and general cognitive capacities.[8] However, the fact that I consider religious phenomena to be autonomous and make broader cognitive connections only secondarily is only a methodological or epistemological issue. Unlike Chomsky himself,[9] and like many other linguists, I do not postulate ontologically the existence of an independent brain faculty for the domain I am describing. For religion, E. Thomas Lawson and Robert N. McCauley demonstrate how one can develop formalisms specifically for rituals while still arguing that this proposal is embedded in the general cognitive capacities of human beings.[10] Similarly, the specific formalism I shall adapt to religion has been developed for language but is asserted to be a general cognitive architecture.[11]

To summarise, the approach proposed in the present article follows the methodology of contemporary mainstream theoretical linguistics. The goal is to build a formal model that is able to describe empirical observations, namely, the dynamics of religious rituals. No connection to general cognitive capacities will be made; suffice it to say that we adopt an architecture that has proven to be very efficient in linguistics, another cognitive domain, and whose connectionist underpinning – creating a possible bridge to brain structures – has been developed by Paul Smolensky and his colleagues.[12]

---

7 Actually, in much mainstream linguistic research, cognitive adequacy is even less important and remains only relevant at the rhetorical level. Scholars base their theories exclusively on grammaticality judgements of well-designed sentences. Only few theoretical linguists (not including neurolinguists and psycholinguists) allow themselves to be influenced by non-linguistic considerations, or even by neurolinguistic and psycholinguistic experiments.

8 For a quick overview, see, for example, the papers and squibs submitted to the online *Archive for Religion & Cognition* (http://www.csr-arc.com).

9 Even Noam Chomsky has recently partially withdrawn his strong claim on the autonomy of the language faculty; see Hauser, Chomsky and Fitch, 'The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?'

10 Lawson and McCauley, *Rethinking Religion*.

11 Smolensky and Legendre, eds., *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*.

12 Smolensky and Legendre, *The Harmonic Mind.*

## 1.2 *Formal models should really be formal*

Formal models – at times making more use of mathematics than at others, and frequently leading to computer simulations – usually become the link between the different aspects of cognitive sciences. These models form the bases of practical applications in robotics or language engineering, even if concrete applications often simplify certain aspects of the theories while having to solve practical issues. At the same time, these models can also guide psychological and neurological research. Many research questions are formulated in terms of these models, and the goal of an experiment is often to supply evidence for them. Without these models, one would be lost in the jungle of neural structures.

In linguistics, these models are often formulated as the result of work using traditional methodologies. Thus, for example, a structuralist analysis of a high number of languages and the subsequent setting up of language typologies – a pre-Chomskyan methodology – is the starting point for the creation of a linguistic model. The Chomskyan or generative turn in its broadest sense, as I view it, was nothing but the introduction of more formal, mathematically more elaborate models, as well as the introduction of a rhetoric supporting these models that sees language not so much as an arbitrary social convention but as a biological phenomenon (hence issues such as innateness, universal grammar, and so forth). Once such a model is proposed, its adequacy is challenged from all directions. The major questions being: Does it really describe the relevant phenomena in all languages, or are there counter-examples? Is it a convincing cognitive model (whatever 'convincing' means), or is it *ad hoc*? Does it also match results in psychological and neurological research? Can it also describe language acquisition data (for example, child language phenomena)? Can the grammar be learned with an efficient algorithm? Can it be used in language technology? Note that even though all these questions can be posed theoretically, not all of them affect the fate of a model.

In the cognitive science of religion, the model of religious rituals by McCauley and Lawson is probably one of the earliest and best examples of a model whose development followed the same path.[13] It was developed on the basis of traditional anthropological methodologies, namely, ethnographic data such as that collected by Harvey Whitehouse in Dadul.[14] How-

---

[13] McCauley and Lawson, *Bringing Ritual to Mind*.

[14] H. Whitehouse, *Inside the Cult*. Harvey Whitehouse also provides a cognitive account of his observations, but the structure of his argumentation is closer to explanations found in traditional anthropological literature than to models in linguistics.

ever, the other eye of the developers of the model remained continuously focused on cognitive science, for example, memory research. These two fields have been combined into a novel abstract formalism referring to mathematical concepts. The McCauley-Lawson model has been tested against an increasing amount of empirical data, both anthropological[15] and psychological, which will certainly lead to substantial refinements in the model in the coming years. The neurological foundations of the theory should also be constantly revised in the light of the most recent research on memory. In turn, the model will hopefully prove to be *adequate* both on a descriptive level (correctly describing empirical data) and on a cognitive level (consistent with what we know about the human brain/mind). However, to the best of my knowledge, the formal details of the theory have not yet been worked out.

This situation is a problem if the cognitive science of religion aims to meet the highest standards of the cognitive sciences. The model represents religious rituals in a three-dimensional space: ritual form, ritual frequency, and arousal associated with the performance of the ritual. In this space a certain dynamics applies force to the rituals, due to which only some positions are stable. McCauley and Lawson speak of *attractor positions*, that is, positions in the space towards which rituals converge in time. Now, the problem is that unless the intention is to use these heavy mathematical concepts only as metaphors, a real scientific model employing notions of dynamical systems is expected to define the dynamics precisely and to demonstrate that the suggested positions are indeed attractors.

It would probably be easy enough to suggest some simple, illustrative mathematical models that yield the expected positions as attractors. It would suffice to choose one of the paradigmatic examples in complex system theory, and somehow to interpret its parameters as the parameters of religion. However, I doubt that the result would go beyond a very superficial parallelism between the behaviour of certain complex systems and religious phenomena, and that this model could quantitatively explain the real dynamics of the parameters involved in rituals. The proponent of such a model should justify *why* the specific equations of the dynamical system apply to religion. Nonetheless, I invite fellow scholars to refute my intuition and to come up with such models, even if they are initially too simplistic, as they may have the potential to develop into convincing theories over the longer term.

Consequently, we move back one step. Instead of directly tackling the dynamics, we will first tackle the mental model, and turn to a cognitive architecture borrowed from linguistics. It will be argued that this ‘grammar’ can describe the way a congregant’s mind works, and the dynamics of the

[15] Cf. Dimitris Xygalatas in the present volume.

ritual system will subsequently be derived from this architecture. By applying Optimality Theory[16] to rituals, I do not want to claim to have found *the* solution, but rather to show what I mean by *a* formal model in CSR. Again, in this way, I would like to call fellow researchers to pursue something better, more convincing, and more adequate on a descriptive and a cognitive level.

### 1.3 *Do not be afraid of formal models!*

Many readers might find the following sections more difficult to read than some other contributions to this volume. Indeed, one may find it useful to stop reading here and there, and just reproduce the argumentation using paper and pen in order to understand (and 'digest') the formalism. Yet, I argue, this is unavoidable in hard-core cognitive sciences.

Many readers will probably ask what the advantage is of introducing such a complex formalism. It may often be useful to translate the formalism into simple terms, but there is a risk that the entire formal enterprise will thus appear to be unnecessarily complicated, an abracadabra just to say something very simple. In fact, while the formalisation actually says something very simple, it also has the potential to reveal much more, and it is only this potential that motivates the enterprise.

An analogy is offered by the history of science. What did physics gain by introducing Newton's laws and the heavy mathematics needed for classical mechanics? Would it not have been simpler to say that the apple fell from the tree? Newton's mechanics had at least three advantages. Firstly, it produced more exact (quantitative), and therefore more verifiable or refutable predictions. Secondly, it created connections between topics that earlier were considered to be unrelated, such as between the falling apple and celestial motion. Finally, it had unexpected consequences: the motion of a space probe can be calculated using mechanics, but it is neither a terrestrial object, nor a star or planet. Newton's mechanics, unlike earlier physics, could correctly predict precisely what would happen to an object dropped by an astronaut on the surface of the Moon. It also contributed to the discovery of the planet Neptune in 1846, whose position could be derived from perturbations in the motion of Uranus.

I believe that the approach presented in this paper has all three of these potentials. Namely, by introducing exact models describing ritual dynamics, which can be tested on computers, the cognitive science of religion becomes a discipline with stronger predictions to be faced with empirical data. Furthermore, the approach is related to cognitive architectures developed in

---

16 Prince and Smolensky, *Optimality Theory*.

linguistics, and thereby helps build bridges between understanding religion and understanding other cognitive phenomena. Finally, a full-fledged version and the computer implementation of the model might exhibit surprising features and explain more phenomena which one would not discover by just speculating with paper and pen about the model.

## 2. *A quick introduction to Optimality Theory*

The reasons for choosing Optimality Theory (OT) are manifold. First, Optimality Theory has proved to be successful in many fields of linguistics for more than a decade, including issues related to thematic roles,[17] the source of Lawson and McCauley's analysis of ritual form.[18]

More importantly, Optimality Theory has its roots in a general cognitive architecture developed by the PDP (Parallel Distributed Processing) Group led by Jay McClelland and David Rumelhart in the mid-1980s. As a member of this group, Paul Smolensky developed his connectionist Harmony Grammar, the precursor to Optimality Theory.[19] Later on, in the early 1990s, collaboration with the fervent anti-connectionist, Alan Prince, resulted in Optimality Theory (OT).[20] Indeed, OT is meant to form the bridge between the low-level connectionist network present in the brain and the high-level symbol-manipulative processes, such as language – and religion, as I suggest.[21] Most linguists use it as a symbol-manipulative architecture for a grammar, while a cognitive scientist and a connectionist can translate it at any time into a neural network. Moreover, it can also be related to ex-

---

[17] See: Legendre, Raymond and Smolensky, *Analytic Typology of Case Marking and Grammatical Voice.* See also: Smolensky and Legendre, *The Harmonic Mind*, vol. II, 161-181. *Thematic roles* – such as agent, patient and instrument – form the bridge between semantics and syntax. For example, the patient of a verb is expressed in an active English sentence by the object, but in a passive one by the subject. The idea was introduced in the 1970s, and became widely used in Chomsky's Government and Binding Theory (Chomsky, *Lectures on Government and Binding*; *idem*, *Some Concepts and Consequences of the Theory of Government and Binding*).

[18] Lawson and McCauley, *Rethinking Religion.*

[19] See Smolensky's 'Information processing in dynamical systems: Foundations of Harmony Theory' in the famous *Parallel Distributed Processing* volume of 1986.

[20] Prince and Smolensky, *Optimality Theory*. The Optimality Theory-Harmony Grammar connection, as well as several theoretical and computational points are elaborated in Smolensky and Legendre, *The Harmonic Mind.*

[21] The authors of *The Harmonic Mind,* vol. I, p. 45, summarise their contribution to what they call the 'cognitive science of language': 'At a more general level than that of any particular results, we hope that, taken as a whole, the book provides some evidence for the value of an approach to cognitive science that is grounded in neural computation, yet centred on formally articulated general cognitive principles'.

isting heuristic approaches to non-linguistic domains.[22] In summary, Optimality Theory is a promising candidate for a formal model of religious rituals in the cognitive science of religion.[23]

To understand the idea of Optimality Theory, suppose that the languages of the world can be organised into the following three types according to their stress pattern:

1. Main stress on the *first* syllable (e.g., Hungarian, Central Norwegian Lappish, Czech, Ono in New Guinea, Debu on Loyalty Islands).
2. Main stress on the *last* syllable (e.g., Uzbek, Yavapai, Moghol, Atayal, Guarani).
3. Main stress on the *penultimate* syllable (e.g., Polish, Piro, Djingili, Mohawk, Albanian, Mussau).

This is only a simple example for educative purposes, and a high number of languages – including Latin, English and Dutch – with more complex (for example, syllable weight dependent) stress systems have been ignored. Nevertheless, it seems to be true that there are but few languages where the rule is to always put the stress on the second syllable. However, the second syllable of a word in other language types may be stressed: the last syllable of two-syllable words in the second type, and the second-last syllable of three-syllable words in the third type. Yet, no language would stress the second syllable of a four-syllable word in this language typology. Data on a large number of languages have been collected, so the lack of languages with the stress always on the second syllable is most probably not simply a random gap. Thus, if a model could describe this typology – that is, predict the existence of the existing types and the non-existence of the non-existing types – we could argue that this model has 'grasped' something about the essence of human language.

---

[22] Gigerenzer, Todd and the ABC Research Group, *Simple Heuristics That Make Us Smart*, p. 91. For the connection of the ABC Research Group's 'fast and frugal heuristics' to Optimality Theory, see also: Bíró, *Finding the Right Words.* A recent study comparing an OT-like model (called the 'lexicographic decision rule') to alternatives in order to account for empirical data of human ethical decisions is presented by Coenen and Marewski, 'Predicting Moral Judgments of Corporate Responsibility with Formal Decision Heuristics'.

[23] An early attempt to combine Optimality Theory with ethical decision-making in a religious context is presented by Parker and Parker, 'Optimality Theory and Ethical Decision Making'. Although Parker and Parker present a useful introduction to Optimality Theory with a fair example from religious ethics, their constraints are too *ad hoc* and too specific to a certain culture. Therefore, these constraints cannot be seen as belonging to a 'universal ethical grammar', which would be required if one wished to adapt the OT philosophy to a cognitive study of ethics and religion.

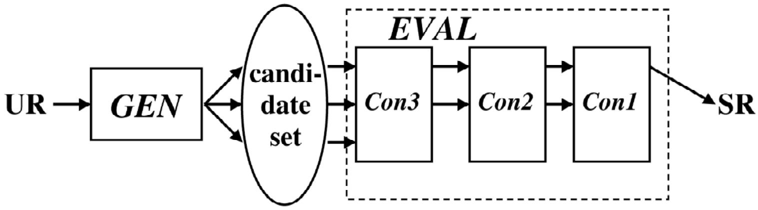


Figure 1. The basic architecture of an Optimality Theoretic grammar. The GEN (Generator) module maps the input, the underlying representation (UR) onto the set of candidates. Subsequently, the EVAL (Evaluator) module chooses the optimal member of this set, which is the output of the system, called the surface representation (SR). EVAL consists of a hierarchy of constraints, which act as filters. On this figure, Con3 >> Con2 >> Con1, that is, constraint Con3 is ranked the highest (is applied first), while Con1 is the lowest (is applied last).

Optimality Theory proposes such a model (Figure 1), postulating a *set of candidates* and a *ranked set of constraints*. The *candidates* are all of the imaginable possibilities, all of the potential forms that could be used in theory to express a given word or sentence in any language of the world. OT introduces two modules to compute the grammatical form of a certain word or sentence in a given language. First, the set of all candidates is generated by the GEN (Generator) module. Next, the best member of this set is chosen by the EVAL (Evaluator) module. Within EVAL, the relative 'goodness' of the candidates depends on the constraint ranking (constraint hierarchy), which is the source of why different languages produce different forms. According to the original philosophy of OT, the set of candidates and the set of constraints are universal, and only the *ranking* of the constraints is language specific. Consequently, it is the hierarchy that accounts for language types; thus, in OT, a grammar is in fact the constraint ranking.

To see how all this works in detail, let us build a model of word stress assignment. You can stress the first syllable, the second, or the third, and so on. One and only one syllable must be stressed.[24] If the input (underlying form, UR) is a four-syllable word (say, *American*), then the set of candidates is the set {*suuu*, *usuu*, *uusu*, *uuus*}, where *s* represents a stressed syllable and *u* stands for an unstressed syllable. For example, *suuu* is the four-syllable candidate whose first syllable is stressed (*Ámerican*), while *uusu* corresponds to stressing the penultimate syllable (*Amerícan*).

---

[24] We presently ignore secondary stress and focus exclusively on primary stress. Real phonological models aspire to account for both, and also for languages with more complex stress systems.

In the next step, we introduce the *constraints*, requirements such as:[25]

(1) EARLY: the stress must occur as *early* as possible in the word.[26]
LATE: the stress must occur as *late* as possible in the word.[27]
NONFINAL: the last syllable must not be stressed.[28]

These constraints are ordered and act as filters. The highest ranked constraint evaluates each candidate first and selects the best subset of its input.[29] Only those candidates that are not worse than some other candidates survive the first constraint. Then comes the second constraint, which similarly filters out some of the candidates that have survived the first constraint, and so on. If a candidate loses at some point, it can never come back into the game, even if it was very good with respect to lower ranked constraints. The output is the candidate (rarely, the candidates) that has survived all of the filters.

In other words, Optimality Theory postulates that the grammatical form, or the form produced by the human brain,[30] is the best (optimal) element of the candidate set with respect to all the constraints ranked by the given hierarchy.

---

[25] The convention is to spell their names with small capitals. The term *constraint* originally denoted hard constraints: requirements that a grammatical form *must* satisfy. Optimality Theory's innovation was the introduction of *soft constraints*: requirements that a grammatical form should satisfy *as much as possible*. As we will soon see, constraints are often violated even by the grammatical form, which fact has resulted in serious criticism from the part of Chomskyan linguists. In order to avoid this criticism (or misunderstanding), the constraints should rather be called basic evaluator functions. This term would also reflect the fact that most constraints do not simply accept or reject a candidate, but assign a number of violations to the candidate. The more violations a candidate is assigned, the worse it is with respect to that constraint.

[26] More precisely, this function returns the number of syllables between the beginning of the word and the stressed syllable (one violation mark per syllable between the beginning of the word and the stressed syllable).

[27] This function returns the number of syllables intervening between the stressed syllable and the end of the word.

[28] It returns *1*, if the last syllable is stressed, otherwise *0*.

[29] It assigns a number to each candidate and selects the candidates that are assigned the lowest value. The constraint filters out any candidate that is assigned a higher number than some other candidate.

[30] For a more precise formulation, as well as for a distinction between what is grammatical according to the static knowledge of language in one's brain and what is produced dynamically by the brain, see Bíró, *Finding the Right Words*. A similar idea is also hinted at in the discussion by Smolensky and Legendre, *The Harmonic Mind*, vol. 1. p. 226-228, on linguistic competence and performance.

The following *tableau*, as it is called in OT literature, summarises the behaviour of the four candidates for a four-syllable word with respect to the constraints mentioned above. Remember, *s* refers to a stressed syllable and *u* to an unstressed one.

(2)

| | EARLY | LATE | NONFINAL |
|---|---|---|---|
| *suuu* | good (0) | worst (3) | good (0) |
| *usuu* | medium (1) | bad (2) | good (0) |
| *uusu* | bad (2) | medium (1) | good (0) |
| *uuus* | worst (3) | good (0) | bad (1) |

Ranking the constraint EARLY the highest will make candidate *suuu* the winner: the other three candidates are worse with respect to EARLY, and are therefore immediately eliminated by this constraint, before the other two constraints could enter the game. Likewise, ranking the constraint LATE the highest will return candidate *uuus* as the single best candidate for constraint LATE, hence as the output of the whole grammar.

Furthermore, the hierarchy NONFINAL >> LATE >> EARLY yields candidate *uusu* as optimal. Namely, first it is candidate *uuus* that meets its Waterloo when the highest ranked constraint is NONFINAL; and then *uusu* is relatively the best among the surviving three candidates with respect to the second highest ranked constraint, LATE. The following tableau visualises this competition:

(3)

| | NONFINAL | LATE | EARLY |
|---|---|---|---|
| *suuu* | good (0) | worst (3)! | good (0) |
| *usuu* | good (0) | bad (2)! | medium (1) |
| ☞ *uusu* | good (0) | medium (1) | bad (2) |
| *uuus* | bad (1) ! | good (0) | worst (3) |

Here, the order of the constraints reflects the hierarchy: the highest ranked, NONFINAL, is leftmost, followed by the second highest, LATE, while the lowest ranked, EARLY, is rightmost. The OT tradition is to use the ! symbol to mark the point where a candidate loses the battle. The cells on the right to this point do not play any role in the computation of the winner, so they are marked by shading. Candidate *uuus* leaves the battlefield in the first round, since it is worse for the constraint NONFINAL than its competitors. Consequently, only three cells are white in the next column, and clearly candidate *uusu* is relatively the best. Since it is the only surviving candidate, that is, there is only one white cell in the last column, the last constraint does not influence the computation. The famous hand symbol ☞ points to the winner candidate *uusu*.

Importantly, the Optimality Theoretic constraints are *violable* (or *soft*): it is possible that the winner candidate violates certain constraints. For example, candidate *uusu* won the competition in (3) despite its violation of the constraints LATE and EARLY. The best candidate nevertheless wins because other candidates also violate these constraints, and/or these constraints are ranked relatively low. The violability of the constraints is an innovation in OT, while in previous and alternative theories the winner must satisfy *all* constraints.

To sum up, this model is able to account for the observed linguistic typology, as each observed language type corresponds to some constraint hierarchy (constraint ranking). The model is further corroborated by the fact that it correctly predicts even the significant gap in language typology mentioned earlier: none of the six possible constraint rankings return candidate *usuu*, that is, no OT grammar with these constraints puts the stress on the second syllable as a rule.[31]

A further issue in contemporary linguistics is the *learnability* of a grammar framework, that is, working out algorithms that can automatically learn a language.[32] Suppose that the learner (a child learning a mother tongue, an adult learning a second language or a software in language technology) knows that the set of possible grammars is $\{G_1, G_2, G_3,\ldots\}$. Then, the learner is given certain learning data, that is, utterances produced by the target grammar $G_t$, a member of the set of possible grammars. The task of the learner is to find (or at least to approximate) this target grammar $G_t$ based on the learning data. The existence of such a learning algorithm is necessary for the suggested grammar framework to be a cognitively adequate model of the human linguistic competence. The same requirement also applies to any model of learnt cultural phenomena, including religion.

For Optimality Theory, a number of learning algorithms have been proposed. Now the task is to find the hierarchy of the (universal) constraints that returns the observed forms as the optimal ones. Suppose, for example, that the learner first hears that the initial syllable is stressed in a two-syllable word (*su*). This fact establishes that the constraint LATE cannot be ranked the highest, otherwise the last syllable would be stressed. However,

[31] The reader is invited to check this statement with pen and paper at this point. It will facilitate the understanding of the forthcoming argument.

[32] For a formal introduction to language learning (and to the effect of learning on language change), including ample bibliographical references, a recommended starting point is Niyogi, *The Computational Nature of Language Learning and Evolution*.

the rule can still be that the first syllable is stressed, or that the second last syllable is stressed. A further piece of learning data, such as *usu*, might subsequently lead the learner to the correct conclusion that the grammar of the language to be learned is NONFINAL >> LATE >> EARLY. A formal learning algorithm describes how this could be done by a dull computer or by a mechanically working set of brain neurons, without reference to the human intuition I have just expected the reader to use in the three previous sentences. As we will see, learning processes will become the main driving forces behind the ritual dynamics in our account.

In what follows, we will introduce an Optimality Theoretic system to describe religious rituals, and then suppose that humans attempt to learn the 'grammar' of superhuman agents. The dynamics in the three-dimensional space of rituals suggested by McCauley and Lawson should follow from the learning procedure.

## 3. *Optimality Theory and human behaviour*

### 3.1 *Eating optimally: a first example*

As a first step towards the application of OT to religious rituals, let us analyse a non-religious form of human behaviour, namely, food consumption. A person entering a restaurant at dinnertime faces a set of possibilities, including eating vegetables, fish, chicken, beef, pork, dog and horse. However, importantly, not eating anything is also an option; let us call this option the *null candidate*. These possible forms of behaviour generated by the input 'entering a restaurant' define the *candidate set* analogous to the candidate sets presented in the previous, linguistic example.

Subsequently, a number of *constraints* driving one's choice can also easily be identified.[33] The constraint DONTSTARVE is unquestionably a universal constraint, but it might be useful to differentiate between two versions of it: DONTDIE and DONTSTARVE. The first constraint is very highly ranked, as proven by the fact that an average European will most probably even eat dog if the only other option would be to die of hunger. However, if the situation were not so extreme, the same person would probably rather stay hungry than eat dog. Moreover, if you are offered something that you do not really like but are used to eating, say, spinach, you still might consume it if you have no better option: if otherwise you may stay hungry or run into unpleasant social situations, such as offend the host, for example.

[33] These constraints can also be seen as a formalisation of the 'preference structures' that Fred Keijzer, in his contribution to the present volume, suggests employing to understand the role of religion.

To account for an aversion to dog and spinach, we introduce further constraints prohibiting the consumption of culturally forbidden and personally disfavoured food: CULTURALLYFORBIDDEN and PERSONALTASTE.

The situations described can be summarised by the following Optimality Theoretic tableau, where the *violation mark* * means that choosing that candidate would violate that constraint:

(4)

| | DONTDIE | CULTURALLY FORBIDDEN | DONTSTARVE | PERSONAL TASTE |
|---|---|---|---|---|
| *die of hunger* | * | | | |
| *stay hungry* | | | * | |
| *dog* | | * | | * |
| *spinach* | | | | * |
| *chicken* | | | | |

For example, the candidate *die of hunger* violates the constraint DONTDIE; but it satisfies the two constraints prohibiting certain types of foods, as well as the constraint DONTSTARVE (after having died, one cannot starve). Eating dog is forbidden in European culture and also happens to violate personal taste. We also suppose for the sake of the example that the person in question does not like spinach.

While introducing the constraints, we have also argued for the constraint ranking to be:

(5) DONTDIE >> CULTFORBIDDEN >> DONTSTARVE >> PERSONALTASTE

The central part of an Optimality Theoretic analysis of some phenomenon is the argument for a certain constraint hierarchy. Techniques, algorithms and computer packages exist to support the linguist in doing so. In our case, one can simply check that a different ranking would not yield the expected behaviour. It is true that candidate *chicken* will win for *any* ranking (because it violates no constraint), provided that the candidate set includes *chicken*. Nevertheless, we can find restricted candidate sets (scenarios where chicken is not on the menu) that will help refute alternative hierarchies. For example, a model that places the constraint CULTURALLYFORBIDDEN above DONTDIE makes the wrong prediction that most people will prefer dying to eating dog in the case where only these two options are present. This prediction can be checked using the following tableau, which employs the ! and ☞ symbols, as well as shading in a similar sense to tableau (3):

(6a)

| | CULT FORBIDDEN | DONTDIE | DONT-STARVE | PERSONAL TASTE |
|---|---|---|---|---|
| ☞ *die of hunger* | | * | | |
| *dog* | *! | | | * |

Note that ranking CULTURALLYFORBIDDEN >> DONTDIE describes the case of the religiously fanatic person or the martyr, who would rather die than eat prohibited food.

Another alternative hierarchy to (5), PERSONALTASTE >> DONTSTARVE is the ranking that depicts those few who prefer starving to eating spinach, if no chicken is offered:

(6b)

| | DONTDIE | CULT FORBIDDEN | PERSONAL TASTE | DONT-STARVE |
|---|---|---|---|---|
| *die of hunger* | *! | | | |
| ☞ *stay hungry* | | | | * |
| *dog* | | *! | * | |
| *spinach* | | | *! | |

However, even this person would most probably not choose death to eating spinach, hence constraint PERSONALTASTE cannot dominate DONTDIE. Likewise, most Europeans would prefer to remain hungry rather than eat dog, which proves that CULTURALLYFORBIDDEN >> DONTSTARVE. To sum up, we have shown that if we ignore the case of the martyr and of the extremist spinach hater, the behaviour of an average Westerner is described by constraint hierarchy (5).

At the same time, the remarks regarding the martyr and the spinach hater demonstrate how different hierarchies formed by the same cultural constraints can explain different types of people or different types of behaviour, similarly to the different rankings of linguistic constraints accounting for different language types.

On a more general level, the background philosophy of Optimality Theory postulates that:

1. The set of inputs are unrestricted and universal (the 'Richness of the Base' principle).
2. The set of candidates generated for a certain input is universal (the Generator function, also called GEN, is universal).
3. The constraints are universal; the set of constraints is universal.
4. The only variable parameter, that is, the only source of cross-linguistic variation, is the constraint hierarchy.

Does our proposed model of cultural behaviour meet these criteria? In theory, each individual from any culture can face any situation. Hence, the 'Richness of the Base' holds, despite the fact that many people in several cultures will never in their life enter a restaurant that offers dog (or pork, for that matter). (Note that a similar restriction is also present in linguistics: even though theoretically any string could be a possible underlying form in any language, the lexicons of the languages are restricted to a finite number of words.) Furthermore, it is also true that any person will generate the same set of candidates in the same situation. The only parameter accounting for differences in behaviour is the constraint ranking, as we have just seen in the case of fanatics or martyrs.

However, in the model presented, the third criterion seems to be false, because the constraint CULTURALLYFORBIDDEN is culturally defined, whereas PERSONALTASTE varies for each individual. Every person has such constraints, but they are not the same – they do not assign the same violation marks – across the entire human race. Yet, a large body of recent work in OT linguistics suggests the introduction of language-specific constraint families: constraints whose general idea is universal, but whose precise content is language-specific. Correspondingly, in our case the general idea of specific types of culturally forbidden and personally disliked food is universal, even if the particular content is culturally or individually defined.

### 3.2 *Eating even more optimally: a second approach*

An alternative approach is to replace these two constraints – CULTURALLYFORBIDDEN and PERSONALTASTE – with universal constraints such as DONTEATDOG, DONTEATPORK, DONTEATCHICKEN, DONTEATSPINACH and so forth. For each substance X, our cognitive system automatically generates a constraint DONTEATX. We then derive notions such as personal taste and culturally forbidden foods from the ranking of these constraints. If constraint DONTEATDOG is very highly ranked across the members of a group, then we can speak of a cultural prohibition. Ranking DONTEATCHICKEN above DONTEATBEEF, even if both are ranked low, means that the person prefers beef to chicken, but has no problem eating the latter if the former is not an option. If DONTEATSPINACH is ranked relatively high by an individual, then he or she has a very strong aversion to spinach.

This second approach can thus explain several levels of aversion; but also several levels between the individual and the larger group. For example, an aversion to broccoli can be present at a family level. Hence, there is no need to define *a priori* what a culture is for the purpose of a constraint such as CULTURALLYFORBIDDEN. Culture can be defined secondarily as a tendency towards a particular constraint ranking shared by the members of the group. For different purposes we can allow larger and narrower groups at the same time.

This second option has further advantages. It can also account for cross-individual variation. Take two individuals, both of whom often eat beef and chicken, but where one prefers beef and the other prefers chicken if given the choice. Applying our model requires ranking both the constraint DONTEATBEEF and the constraint DONTEATCHICKEN very low (relative to other constraints such as DONTEATDOG or DONTSTARVE) for both individuals. However, their relative rank is different: one ranks the constraint DONTEATCHICKEN higher than DONTEATBEEF, and the other vice versa:

(7a) Prefers beef:

| | . . . | DONTEATCHICKEN | DONTEATBEEF |
|---|---|---|---|
| *chicken* | | *! | |
| ☞ *beef* | | | * |

(7b) Prefers chicken:

| | . . . | DONTEATBEEF | DONTEATCHICKEN |
|---|---|---|---|
| ☞ *chicken* | | | * |
| *beef* | | *! | |

Even individuals can display certain variability: one day choosing beef, but the next day preferring chicken. We can therefore stipulate a temporal reordering of the constraints close to each other, due to reasons such as 'I did not like the beef yesterday' (so I slightly promote DONTEATBEEF in the hierarchy), or 'now I miss chicken' (that is, demoting DONTEATCHICKEN). These random minor temporal promotions and demotions of constraints – sometimes resulting in the reordering of neighbouring ones – are realised in a principled way in Paul Boersma's Stochastic Optimality Theory.[34] In the

[34] Boersma and Hayes, 'Empirical Tests of the Gradual Learning Algorithm'.

next section our model will also require such shifts in the ranking of constraints.

The last advantage of the second model is its simplicity and naturalness. The constraints used are directly related to the elements of the candidate set, that is, to the reality of the world, whereas more complex notions, such as individual and cultural preferences, become derived concepts.

Furthermore, the constraints can be seen as basic, very much physiologically motivated cognitive factors: 'don't die!', 'don't starve!', 'don't eat X!'. Remember that in linguistics constraints originally meant non-violable constraints, and OT's innovation was to allow violable constraints. Similarly, the factor 'don't eat X' is hard, non-violable (ranked very high) if X is really poisonous, whereas soft, violable (ranked relatively low) if X is not very healthy or should be avoided in too great a quantity. Perhaps it is in order to avoid substances that are edible but dangerous in high quantities that our cognitive system temporarily promotes constraint DONTEATX after having consumed X (hence, we would like to eat something different the next day), even if X did not cause nausea. In turn, if Middle Eastern cultures have ranked DONTEATPORK very high (and this hierarchy is stable across the population and in time), then I propose that this phenomenon uses (or misuses, is parasitic on) the same cognitive mechanisms that are employed by a human or animal population to avoid highly poisonous food.

What mechanisms are they? We are speaking of two intertwined learning mechanisms: (1) individuals learn from experience to promote certain constraints above others, and then (2) pass on this knowledge to other individuals through cultural learning. The first learning procedure arises from the interaction between an individual and their environment, and translates personal experience into culture. For example, a good or bad physiological experience following the consumption of substance X is translated into a piece of cultural knowledge; in our model, into the promotion of the constraint DONTEATX, if substance X was unpleasant, and into the demotion of the same constraint in the opposite case. The second learning process, however, takes place *within* the population and *within* the given 'domain' (culture, in our case). The second individual learns from the first without actually experiencing, for example, the consequences of eating substance X. This learning requires the usual mechanisms by which the next generation acquires their mother tongue or their culture. Similar learning mechanisms are also responsible for spreading new linguistic or cultural features (memes, if you wish) among members of the same age group. The advantage of this secondary learning is obvious: you do not have to experience the bad taste of X yourself to learn to avoid it.

However, not all pieces of cultural knowledge – not all of the details of constraint ranking, in our model – can be derived from direct biological experience. The peer-to-peer learning mechanism, or even the architecture itself, can very easily create unexpected consequences.[35] At this point the cognitive mechanism starts to have its own life and opens the door to by-products.

Even though many rationalist minds have tried to connect the Jewish and Islamic aversion to pork to hygienic and climatic explanations, much more complex mechanisms must be present, otherwise one could not explain its stability over time and geographical location, and even less so its traditional native justifications. According to the native justification, namely, it is not the interaction with nature but the interaction with the superhuman agent that caused the promotion of the constraint DONTEATPORK to the position equivalent to pork being highly poisonous. Whatever occurred historically, we probably indeed require an interaction with the superhuman agent to explain why this constraint is still ranked extremely high among Jews and Muslims. The natural context explains the biological evolution of the architecture, whereas the social context may explain the historical evolution of the by-products.[36]

The following section therefore employs the social context rather than the natural context when explaining the dynamics of rituals.

## 4. *Optimality Theory and rituals*

### 4.1 *Ritual dynamics: (selected) observations to be explained*

In this section we not only seek to model the rituals themselves, but also to explain their dynamics, their diachronic changes. Partly following and partly critically reformulating the list of possible religious developments in McCauley and Lawson,[37] our starting point will be the following set of observations:

1. When one performs a religious ritual for the first time, emotional arousal is relatively high.[38]

---

[35] See Bíró, *Finding the Right Words*, for examples of the erroneous functioning of the language production system.
[36] A different cognitive explanation for food taboos – following a helpful overview of earlier rationales – is presented by Fessler and Navarrete, 'Meat is good to taboo'.
[37] McCauley and Lawson, *Bringing Ritual to Mind.*
[38] This applies not only to the person performing a ritual for the first time in their life (first communion, first Torah reading, etc.), but also to someone performing a ritual to invoke rain for the fist time in that year, for example. In the case of a first communion, a bar mitzvah or a wedding ceremony, a number of non-religious factors also influence the levels of excitement: stage fright, fear of entering a new

2. Repetition leads to a decrease in emotional arousal if the ritual is a *human-action-only* ritual. This decrease converges towards a low arousal attractor position, around which fluctuations are possible.
3. Repetition leads to an increase in emotional arousal if the ritual is a *superhuman-reaction* ritual, and the reason for repetition is the failure of the previous ritual (measured as the lack of observable superhuman reaction). Due to serial repetitions, the level of arousal reaches a ceiling, followed by a breakdown.
4. A ritual system that has only *human-action-only* rituals fluctuating around the low-arousal attractor position(s?) will generate outbursts of high-arousal rituals at random moments in time. In the most extreme case, this outburst creates a splinter group; in milder cases, the system itself increases the arousal of some rituals – or introduces higher arousal rituals – in order to avoid the *tedium effect*.

Here, we differentiate between *human-action-only* rituals and *superhuman-reaction* rituals, corresponding more or less to McCauley and Lawson's special patient/instrument rituals and special agent rituals respectively.[39] The reason for the new terminology is that we are not so much concerned with the inner *structure* of the rituals, but rather their *function*. In the case of a *superhuman-reaction* ritual the congregant performs a ritual in order to coerce the superhuman agent to act, for example, to bring rain or a good crop, to validate the status of two people as married, or to bring along the period of the companies. In the case of *human-action-only* rituals, however, no such immediate response is expected. The only goal is to maintain good contact with the superhuman agents (or fellow congregants).

Our long-term aim is to *quantitatively* account for the differences between these two types of rituals, to describe the nature of the low-arousal attractor (including the fluctuations around it), the maximal arousal ceiling and the outbursts, but also to analyse the factors influencing this dynamics. In what follows we will try to establish the first qualitative steps in this research programme, leaving computer simulations and mathematical analysis to the future.

---

form of life, the long preparation process, etc. It is an essential question whether these factors (present also in similar, but non-religious situations) should be accounted for by a theory of religious rituals.

[39] For an introduction to McCauley and Lawson's model, see Risto Uro's article in the present volume.

## 4.2 *Theology as a behavioural grammar*

Our starting point – which in many respects resembles that of Stark and Bainbridge[40] but does not aim at deriving a socioeconomic system – is that humans not only have their own *behavioural grammar* driving their own actions (modelled as an Optimality Theoretic system described in the previous section); but they also have a hypothesis about the behavioural grammar driving fellow agents in society (supposing healthy adults with a Theory of Mind). As the fellow agents include gods, spirits and ancestors, humans should also have a theory of the superhuman agents' minds. Thus, in addition to the OT behavioural grammar driving my own behaviour, and the OT grammars that my mind assumes for each fellow human, I also retain a set of additional OT grammars for each superhuman agent that I and my culture postulates. Let us call this latter theory or grammar of a superhuman (counterintuitive) agent's mind an *intuitive theology*.

A congregant predicts using this *theology* how the superhuman agents would react to a human behaviour *b*. To each *b* the theology assigns a value *r*, the *reaction* of that particular god to human action *b*. In other words, the theology grammar is a mapping from possible values *b* to the corresponding values *r*, or a set of possible (*b, r*) pairs. Let us call this set a *theological language*.[41] Human actions that do not entail the reaction of a counterintuitive agent do not belong to the realm of religion, whereas divine actions that have no influence on present society are purely mythological. What concerns us here is the interaction between the two spheres, and more specifically, the superhuman actions that are direct or indirect reactions to human actions (such as rituals and other religious acts,[42] or secular acts with some positive or negative religious-moral values). Even facts independent of human actions that have a direct effect on human life, such as 'each autumn the goddess calls the rain to water the earth', will be considered as mythology and will be omitted, unless the event is described as a function of human behaviour, such as 'each autumn the goddess calls the rain to water the earth, providing we have presented the correct sacrifice and behaved well'.

---

40 Stark and Bainbridge, *A Theory of Religion*.

41 In Chomsky's terms, this set is an E-language ('external language'): the infinite set of possible utterances that are grammatical with respect to the finite grammar, that is, to the I-language ('internal language').

42 The distinction between rituals and religious actions were introduced by Lawson and McCauley (*Rethinking Religion*, p. 127). Religious actions are actions whose structural descriptions involve at least one element of a religious conceptual system. Rituals are special religious actions: those whose structural descriptions also have a patient (object) position, in addition to the agent (subject) position. Nevertheless, we will not consistently follow this distinction in this article.

Consequently, a religious person does not simply face a set of possible behaviours $\{b_1, b_2, \ldots\}$, as was the case in Section 3. He or she has to optimise the elements of the *theological language*, that is, a set $\{(b_1, r_1), (b_2, r_2), \ldots\}$. When planning our acts, we also take into account the possible reactions of the superhuman agents. Note that this observation applies not only to religious acts, but also to any action in society that involves a reaction from the part of other agents. In fact, the theological evaluation of actions seems to reuse the system evolved for social purposes.

In our food-selection example, certain alternatives might trigger positive or negative reactions, for example, depending on whether you decide to cook the your partner's favourite meal for dinner, or you happen to eat the last piece of their favourite cake. The possible reactions of the other agents are also features that should be evaluated by the constraints of our behavioural grammar. Such situations can be described in terms *bi-directional Optimality Theory*.[43] Here we take into account both the speaker's and the hearer's perspectives ('if I say it this way, he would understand it that way'; 'if she had meant that, she would have formulated her utterance in another way').

However, to simplify our discussion, we will skip all the details of bi-directional Optimality Theory. What we will do is first restrict the set of all behaviour-reaction pairs $(b, r)$ to those which conform to the theological language: for each human behaviour $b$ we let the intuitive theology calculate what reaction $r$ by the superhuman agent is assumed. Then, we find the best element of this restricted set. In this second evaluation process, our own behavioural grammar may include constraints on $b$ (which behaviour I prefer myself), but also on $r$ (what reaction I prefer or would like to avoid).

Let us suppose that theology teaches that certain human behaviours $b_{sin}$, such as stealing, involve a superhuman reaction $r_{punishment}$. Thus, the candidate set will be restricted to the candidates in the left column of (8a). The candidate (*stealing, no punishment*) is eliminated by theology teaching automatic divine punishment, and the candidate (*no stealing, punishment*) by the belief in divine justice. Human behaviour will then optimise for the pros and cons of the different possible action-reaction pairs $(b, r)$: behaviours $b_{sin}$ will be avoided, unless some other factor outweighs the tendency for punishment avoidance. Such a model can be written simply in terms of Optimality Theory:

[43] Blutner, 'Some aspects of optimality in natural language interpretation'.

(8a)

| | NOPUNISHMENT BYGOD | WEALTH | NOJAIL |
|---|---|---|---|
| (*stealing, punishment*) | * | | * |
| (*no stealing, no punishment*) | | * | |

Here, we have introduced the following constraints, depending on either behaviour *b* or reaction *r*: NOPUNISHMENTBYGOD is violated by candidates (*b*, *r*) such that the divine reaction *r* involves punishment. The constraint WEALTH is violated if the human behaviour *b* does not guarantee material wealth. Finally, NOJAIL is violated if the reaction of society or other non-superhuman agents (not shown in the tableau) is a jail sentence.

Different hierarchies will account for different behaviours in society. The reader is invited to check – following the techniques used to compute earlier tableaux – that if the highest ranked constraint is WEALTH, the individual in question becomes a thief; but not otherwise.[44] The relative rank of the two other constraints in different individuals in society reflects the efficacy of religious education and the secular legal system. A person influenced by religious values will not steal because of the highly ranked constraint NOPUNISHMENTBYGOD, while in other cases the practical principle of jail avoidance will dominate.

Further constraints might also be added, such as stealing can be a behaviour considered as 'cool' or as 'bad' by peers, or divine reward and punishment in this life or after death can be discerned as having different weights, etc. – as you wish. Constraints may differentiate between different levels of reward. So the constraints EARN$1000 and EARN$10 are satisfied only if the behaviour earns at least $1000 and $10 respectively; not stealing or stealing less violates the constraint, resulting in a star in the tableau. Imagine that the constraint EARN$1000 is ranked much higher than constraint EARN$10, and that the constraint NOPUNISHMENTBYGOD is ranked between them in the behavioural grammar of a person. Tableaux (8b) and (8c) prove that this person will steal if there is a possibility that they will earn $1000, but not if the sum is much lower. The first column in (8b) is the interesting

[44] If the reader is encountering Optimality Theory for the first time, it might really be worth stopping here for a while. Please take a piece of paper and pencil, and write down the six possible tableaux corresponding to the constraints in (8a). Make sure you understand which constraints assign a violation mark to the two candidates. Then, in each case, find the highest ranked constraint that makes a difference between the candidates. Mark this violation by an exclamation mark. Finally, add the shading to the cells behind the exclamation mark. This little exercise should illuminate much of the paper!

case of both candidates violating the highest ranked constraint and therefore neither of them losing.

(8b) A chance to steal $10

| /$10 in front of you/ | EARN$1000 | NOPUNISH BYGOD | EARN$10 |
|---|---|---|---|
| (*stealing, punishment*) | * | *! | |
| ☞ (*no stealing, no punishment*) | * | | * |

(8c) A chance to steal $1000

| /$1000 in front of you/ | EARN$1000 | NOPUNISH BYGOD | EARN$10 |
|---|---|---|---|
| ☞ (*stealing, punishment*) | | * | |
| (*no stealing, no punishment*) | *! | | * |

Nevertheless, if you would like to argue for a model that is more than a mere play with OT tableaux, you should propose well-founded constraints, which are universal, which can account for a number of behavioural types, and which might be deduced from basic biological, ethological or psychological observations. This last example (8) does not aspire to be such a model; it is rather an illustration.

4.3 *The grammar of transactions*

After these simpler models, let us focus on a more abstract one that should bring us closer to understanding the dynamics of religious ritual systems. Note that the abstract constraints below can always be replaced with specific constraints such as those just described.

In what follows, we focus on the grammar of positive *transactions*. In such transactions, people engage in some form of self-sacrifice (such as spending time on prayer, suffering pains, or bringing offerings) in order to coerce the divine being to act favourably towards them. This is the type of ritual we have called *superhuman-reaction* ritual. In the case of *human-action-only* rituals, the divine reaction does not play an immediate role, even though, as we shall see, it might influence the long-term dynamics.

The basic structure of such a transaction between human and superhuman agents can be described thus: the human agent invests a price $p$ (the human behaviour is $b = p$), and subsequently a reaction $r$, ascribed to the superhuman agent, is experienced. In a simplifying way, the model supposes that the price, payable through such things as assets, money, material goods, energy, time, human suffering and pain, can be measured as a single non-negative number $p$. These resources have an upper limit; so let $L$ denote this limit. The gods, in turn, either fulfil the request (the divine reaction is $r = 1$) or they do not ($r = 0$). In sum, the following scenarios are possible:

| | |
|---|---|
| $(b = 0, r = 0)$ | $(b = 0, r = 1)$ |
| $(b = 1, r = 0)$ | $(b = 1, r = 1)$ |
| $(b = 2, r = 0)$ | $(b = 2, r = 1)$ |
| … | … |
| $(b = p, r = 0)$ | $(b = p, r = 1)$ |
| … | … |
| $(b = L, r = 0)$ | $(b = L, r = 1)$ |

For example, scenario $(p = 4, r = 0)$ corresponds to the case of a human agent presenting a sacrifice of a price of 4 but without the superhuman agents listening to the offer. Similarly, another scenario $(p = 0, r = 1)$ expresses a case in which gods fulfil the wish of the humans without any sacrifice on the human side. The opposite scenario is $(p = L, r = 0)$, in which case the gods do not react at all, even though a person has done their utmost. Let us add a further scenario, which is only a theoretical one, but which will become important: $(p = L+1, r = 1)$ will describe the breakdown of the system. Namely, this scenario will be the winner when humans assume (when their intuitive theology predicts) that they should do more than they can (that is, more than their limit $L$) in order to have the gods fulfil their wish.

The set of scenarios described – the candidate set of our OT-like model – is of course a simplification. Human investment might be more complex than what can be modelled as a single number. Imagine that different kinds of assets, time and pain are involved at the same time. Furthermore, the perceived reaction of the superhuman agents can also be of several kinds: it may be perceivable over a short or long term, be perceivable by all agents or only by priests, be dependent upon the interpretation of the events, and so forth. Nevertheless, simplifying the problem to a one-dimensional price paid by humans and to a yes-or-no reaction by the superhuman agents will be useful in a first approximation, and at a later stage the model can be made more complex.

It has long been supposed that humans are driven by a series of laziness constraints, euphemistically called ECONOMY, widely used in linguistic literature:

(9) ECONOMY$_z$ (ECO$_z$, $*p > z$): do not pay more than $z$!

This constraint is satisfied only by candidates (*p, r*) such that $p \leq z$, that is, by scenarios in which human agents pay a price not higher than *z*. If $p > z$ then the candidate (*p, r*) violates the constraint $\text{ECO}_z$. The abbreviation *p > z refers to this property: a violation mark * is assigned to all candidates whose *p* is greater than *z*. ECONOMY is a *constraint family* because it includes a number of constraints for different *z* values. By their very nature, these constraints must be ranked as follows:

(10) Human behavioural grammar:
$\text{ECONOMY}_L >> \text{ECONOMY}_{L-1} >> \ldots >> \text{ECONOMY}_2 >> \text{ECONOMY}_1$

In other words, the idea of not being willing to pay more than a high price is always more influential on one's decisions than the idea of not being willing to pay more than a low price. A different ranking would render some of the constraints superfluous. Given these constraints, the winning candidate is always the one with the lowest price. For example, if *L = 4*:

(11)

| | $\text{ECONOMY}_4$ | $\text{ECONOMY}_3$ | $\text{ECONOMY}_2$ | $\text{ECONOMY}_1$ |
|---|---|---|---|---|
| (*b* = 5, *r*) | *! | * | * | * |
| (*b* = 4, *r*) | | *! | * | * |
| (*b* = 3, *r*) | | | *! | * |
| (*b* = 2, *r*) | | | | *! |
| ☞ (*b* = 1, *r*) | | | | |

In other words, if no other factor is present, humans will always choose the cheapest possible option. Why then are we investing in interactions at all? Notwithstanding cases of altruism, the motivation is that we would like to reach our goals through the interaction, that is, we would like our partner to react. Consequently, we need a theory of the partner's mind, which is what we have called an *intuitive theology* in the case of a superhuman partner.

The intuitive theology consists of a family of constraints called DONTGIVE, similar to ECONOMY, as well as the additional constraint REACT that describes the wish of the superhuman to also engage in the interaction.

(12) $\text{DONTGIVE}_z$ ($\text{DG}_z$, *p < z): do not react to a proposal lower than *z*!
REACT (R, *r = 0): react to any proposal!

The constraint $\text{DONTGIVE}_z$ is violated by scenarios (*p, r*) in which $p < z$ and simultaneously $r = 1$. However, candidate scenarios in which either the price paid is high enough ($p \geq z$) or the superhuman does not react to the human sacrifice ($r = 0$) satisfy this constraint. The idea is that the vendor should not sell (i.e., $r = 1$) for a too low a price (i.e., $p < z$).

The constraint REACT penalises interactions in which the superhuman does not react to the offer: it is satisfied by candidates $(p, r)$ such that $r = 1$, and violated if $r = 0$.

In the intuitive theology – describing the mind of the superhuman agent, and maintained in the mind of the human agent – members of the DONTGIVE family are ranked by their nature in decreasing order. This is because any vendor would avoid a transaction whose price is less than a low value $z$, but it is much less of an issue to avoid a transaction whose price is less than a high $z$. The constraint REACT is ranked somewhere between two constraints of this family, say, between $\text{DONTGIVE}_n$ and $\text{DONTGIVE}_{n+1}$:

(13) Grammar of intuitive theology:
$\text{DONTGIVE}_0$ >> $\text{DONTGIVE}_1$ >> … >> $\text{DONTGIVE}_n$ >> REACT >>
>> $\text{DONTGIVE}_{n+1}$ >> … >> $\text{DONTGIVE}_L$ >> $\text{DONTGIVE}_{L+1}$

The value $n$ in theology (13) will become central: different theologies differ in terms of where they rank the constraint REACT, that is, what value the parameter $n$ (the index of the DG constraint just preceding REACT) takes.

How does an intuitive theology work? The superhuman agent is offered a sacrifice of price $p$, and has two options: either to react or not to react. Hence, according to the human agent's theory of the superhuman agent's mind, the latter will judge between two options. Candidate ($b = p$, $r = 0$) stands for the option with no superhuman reaction and candidate ($b = p$, $r = 1$) describes the case of a positive reaction. Will the superhuman agent accept the offer?

The crucial point in determining the answer is which of $p$ (the price of the offer) and $n$ (the place of the constraint REACT in hierarchy (13) describing the intuitive theology) is larger. Consider the following two examples:

(14) Intuitive theology $n = 2$:
$DG_0$ >> $DG_1$ >> $DG_2$ >> REACT >> $DG_3$ >> $DG_4$

(14a) $n = 2, p = 2$:

| | DG0 | DG1 | DG2 | REACT | DG3 | DG4 |
|---|---|---|---|---|---|---|
| ($b = 2$, $r = 0$) | | | | *! | | |
| ☞ ($b = 2$, $r = 1$) | | | | | * | * |

(14b) $n = 2, p = 1$:

| | DG0 | DG1 | DG2 | REACT | DG3 | DG4 |
|---|---|---|---|---|---|---|
| ☞ ($b = 1, r = 0$) | | | | * | | |
| ($b = 1, r = 1$) | | | *! | | * | * |

In the cases of both tableaux, the $n$ used to rank REACT is $n = 2$ (the constraint REACT is ranked between the constraints DONTGIVE$_2$ and DONTGIVE$_3$). If the human agent offers a price of $p = 2$ or higher, as in (14a), the superhuman should accept it, because the constraint REACT will eliminate the other option (not reacting). If, however, the price of the sacrifice presented is lower than $n$, namely $p = 1$, in tableau (14b), then the superhuman is expected not to react, because some highly ranked DONTGIVE constraints – DG$_2$ in (14b) – prevent the reaction before the evaluation of the two candidates reaches the constraint REACT.

Therefore, if a human agent presupposes the intuitive theology (13), he or she must bear in mind only the options of paying a price not lower than $n$, since otherwise it can be expected that the sacrifice will be rejected by the superhuman partner. However, the ECONOMY constraints in his or her own behavioural grammar (10) force the payment to be as little as possible, as shown by tableau (11). The candidate set evaluated according to the behavioural grammar thus includes only the candidates in relation to which it is expected – based on the intuitive theology – that there will be a positive reaction from the superhuman:

(15) Behavioural grammar (10) employed to the candidates whose $r = 1$ and which are winners for theology (14).

| | ECONOMY$_4$ | ECONOMY$_3$ | ECONOMY$_2$ | ECONOMY$_1$ |
|---|---|---|---|---|
| ($b = 5, r = 1$) | *! | * | * | * |
| ($b = 4, r = 1$) | | *! | * | * |
| ($b = 3, r = 1$) | | | *! | * |
| ☞ ($b = 2, r = 1$) | | | | * |

In this set – the positive reaction subset of the theological language, $\{(b = n,\ r = 1), (b = n + 1,\ r = 1), \ldots (b = L,\ r = 1)\}$ – the optimal candidate is $(b = n,\ r = 1)$. Consequently, the human agent performs a ritual investing a price $n$, and not a penny more. Remember that this value depends on the human agent's intuitive theology, that is, at what lowest price the agent believes the superhuman would react.

4.4 *Deriving the dynamics*

Having reached this point, let us return to the four observations made at the beginning of Section 4. In what follows, I propose a preliminary explanation of these observations using the transaction grammar introduced in the previous subsection.

When a ritual is presented for the first time, the person's theology assumes a more or less random position for the constraint REACT in theology (13), which translates into a medium-large value for *n*. In turn, the ritual is presented for the first time with a medium level of emotional energy invested. As it is the first time ever (such as in the case of a first communion or a first call to the Torah), or the first time after a long period (first prayers for rain in that year), there is no personal experience as yet, and this medium value of *n* is defined by 'general intuitive theological knowledge' in the semantic memory.[45]

The constraint family ECONOMY is probably derived from experience held in the episodic memory, whereas the intuitive theology is a combination of elements from the semantic and the episodic memory (compare this to Whitehouse's two modes of religiosity).[46] Before one performs the ritual for the first time, only semantic information is present, which is learnt culturally from older co-religionists and which does not contradict personal experiences. This is probably why the leaders of many initiation rituals can easily convince novices to endure so much pain: the novice simply postulates a high *n* if told to do so. Subsequently, as this is the first time the novice's behaviour violates the highly ranked constraint $\text{ECONOMY}_n$, the event remains memorable because it provides precious autobiographical information (experience) on this important constraint. Namely, the novice learns that even this constraint is violable: one can survive such pain and endure it for the sake of fellow group members. Therefore, constraints such as those ensuring group solidarity must be ranked even above $\text{ECONOMY}_n$.

The developments following the first performance depend on the superhuman agent's 'feedback', as experienced by our human congregant. If the latter feels or believes that the superhuman agent has reacted, then there is no reason to alter the intuitive theology.

If, however, the congregant has the impression that the gods have not accepted the sacrifice, then there is a need to revise the theory concerning

[45] A ritual not performed for a year might be an intermediate category. Only traces of the experience accumulated in the past remain in the episodic memory. Therefore, its first repetition involves more emotions than the repetition of a ritual performed the previous day, but fewer emotions than the performance of a ritual for the very first time.

[46] Whitehouse, *Inside the Cult*; Whitehouse, *Modes of Religiosity*.

the superhuman agent's mind. A *learning* process commences based on the following piece of information: from the options ($b = n$, $r = 0$) and ($b = n$, $r = 1$), the superhuman agent has actually chosen the former (not to react to the sacrifice of price $n$). The constraint REACT must be *demoted* in the hierarchy (13), because this is how the congregant can only explain the rejection of the sacrifice by the superhuman agent.[47] In other words, the constraint REACT moves more to the right in the hierarchy. The value of $n$, the threshold measuring the place of the constraint REACT within the hierarchy, is increased. The revised theology will predict that gods only accept more expensive sacrifices, and the congregant entertaining this revised theology will repeat the ritual with a higher emotional arousal.

In the example of hierarchy (14), repeated here as (16a), the human agent wrongly expected the superhuman agent to react positively to an offer of price 2. The lack of reaction to an actual sacrifice is seen as a piece of learning data, based on which the human agent modifies their mental model of the superhuman agent's mind, resulting in the new hierarchy (16b). This adjusted theology correctly accounts for the lack of a reaction to an offer of price 2, but at the same time predicts a reaction to a higher price in (16c).

(16a) Intuitive theology ($n = 2$):
$DG_0 >> DG_1 >> DG_2 >> \text{REACT} >> DG_3 >> DG_4$

| | DG0 | DG1 | DG2 | REACT | DG3 | DG4 |
|---|---|---|---|---|---|---|
| ($b = 2$, $r = 0$) | | | | *! | | |
| ☞ ($b = 2$, $r = 1$) | | | | | * | * |

(16b) Intuitive theology ($n = 3$):
$DG_0 >> DG_1 >> DG_2 >> DG_3 >> \text{REACT} >> DG_4$

| | DG0 | DG1 | DG2 | DG3 | REACT | DG4 |
|---|---|---|---|---|---|---|
| ☞ ($b = 2$, $r = 0$) | | | | | * | |
| ($b = 2$, $r = 1$) | | | | *! | | * |

[47] The most basic learning algorithms for Optimality Theory – *Error Driven Constraint Demotion* (EDCD) and *Recursive Constraint Demotion* (RCD) – are summarised in: Tesar and Smolensky, *Learnability in Optimality Theory*. Another widespread approach is Boersma's *Gradual Learning Algorithm* (GLA; cf. Boersma and Hayes, 'Empirical Tests…'). Note that the bibliographical references point to the most comprehensive and accessible publications, but earlier versions can be also found in the Rutgers Optimality Archive (http://roa.rutgers.edu).

(16c) Intuitive theology ($n = 3$):
$DG_0 >> DG_1 >> DG_2 >> DG_3 >> $ REACT $>> DG_4$

| | DG0 | DG1 | DG2 | DG3 | REACT | DG4 |
|---|---|---|---|---|---|---|
| ($b = 3, r = 0$) | | | | | *! | |
| ☞ ($b = 3, r = 1$) | | | | | | * |

If the congregant experiences failure for the second time, the candidate set will be further restricted, because a second learning process takes place based on this new piece of evidence. He will then repeat the ritual with an even higher price. This procedure goes on as long as the annual rainfall or the period of the companies does not arrive. One can easily see that this model results in what McCauley and Lawson[48] intuitively predicted: a *superhuman-reaction* ritual (a special-agent ritual) is performed usually only once with a relatively high level of arousal; but if it has to be repeated several times due to the failures experienced, then it will be repeated with an increasing level of investment (emotional arousal), and it will sooner or later reach a breaking point (a ceiling).[49]

We expect the breaking point to arrive when the only option is candidate ($b = L + 1$). This happens after the ritual has been performed by paying the highest possible price $L$. Since it has been rejected, our poor congregant must conclude that the correct theology prefers ($b = L$, $r = 0$) to ($b = L$, $r = 1$), that is, the constraint REACT must be further demoted below the constraint $\text{DONTGIVE}_{L+1}$. Then, however, the only price to which the superhuman agent is supposed to react positively is candidate ($b = L + 1$), but such a price surpasses human capabilities.

What happens in cases of rituals where no *direct* superhuman reaction is expected (*human-action-only* rituals, more or less the special patient/special instrument rituals in the terminology of McCauley and Lawson)? In such cases no positive or negative feedback – no information on success or failure – influences the intuitive theology. As the agents of the action are the congregants, their behaviour directly determines success or failure. It is like a one-directional communication towards the superhuman agents with a repeated $r = 1$ positive reaction from the divine. Starting with a medium level $n$, the agent's mind will slightly alter theology (13) due to random fluctuations. Whenever the constraint REACT happens to be slightly promoted, $n$ decreases, and the congregant decreases the price or arousal invested in the ritual. Such a constraint promotion may occur, for example, as the consequence of a learning process following that the congregant has

48 McCauley and Lawson, *Bringing Ritual to Mind.*
49 McCauley and Lawson, *Bringing Ritual to Mind.*

witnessed somebody else performing the ritual 'successfully' with less investment. As no failure will bring evidence against this new theology, which is more convenient to the lazy human being, the latter will deduce that the gods are content even with this lower price.

In turn, slowly but surely the theology will be altered so that less and less effort will be required on the part of the congregant in the ritual. Hence, we have accounted for the low level of arousal (or other price) typical to human-action-only rituals. Additionally, we predict that the more often they are performed without any superhuman reaction, the lower the price people invest in them.

What prevents the congregant from becoming maximally lazy? How can this person maintain an amount of ritual investment (for example, a level of emotional arousal) that might fluctuate, but is constant on average over the longer term? We argue that accidental events which restore the theology – that is, personal experiences that demote the constraint REACT in hierarchy (13) – are crucial. Several factors are imaginable, including social influences and pressures, random private events and high-arousal outbursts, including those undergone in splinter groups.

A balanced ritual system also contains rituals involving the 'measurable' reaction of superhuman agents, and these rituals will help restore the theology through the mechanism described earlier. The balance of a balanced ritual system lies between superhuman-reaction rituals, which increase the *n* value of the intuitive theology, and between human-action-only rituals, which decrease the same value. Note that the current proposal also explains why the arousal levels of different rituals are interconnected at all: the 'glue' of the ritual system is the intuitive theology, which is independent of the particular rituals, but influences all of them. In fact, the notion of an intuitive theology is unrelated to any theory of rituals: it emerges automatically from the psychological observation that children and adults have a representation of other agents' mind and from the view in CSR that gods and ancestors are culturally postulated (counterintuitive) agents.

If there are no superhuman-reaction rituals, splinter group outbursts might serve the same purpose. (Note that our theory explains the role and the importance of these outbursts, but not their source: that is, why certain individuals suddenly begin to increase the ritual price paid whenever the price has reached a low level, that is, in a situation of tedium.) Often, observing the high-arousal ritual performed by another congregant is in itself sufficient, probably due to an empathy mechanism. Communication is a similar phenomenon: listening to the personal testimony or to the teaching of another individual will influence one's intuitive theology.[50]

---

[50] At least in Judaism, intensive study of the laws of rituals can have the effect of

Social pressure to invest more in rituals is another, external way to exclude low-price candidates from the candidate set in (15). In this case, it is not theology but the theory of other, human agents' minds that excludes unpleasant candidates from the candidate set.

Finally, I conjecture, an important role is played by random events that are interpreted as positive reactions *r* of the superhuman agents to an action *b* of the human agent with a relatively higher investment *p*. It randomly occurs that the congregant experiences something good and this event coincides more or less with a ritual that is remembered as being performed with a higher price. The increased investment can be the result of social pressure, or one might just subjectively recall this particular performance as being more intense. Stochastic variants of Optimality Theory, corresponding to a stochastic behaviour of the mind, also enable us to add random fluctuations in performance intensity to the model. In all of these cases, the congregant will interpret the positive event as a reaction to that ritual performance with increased investment, even though the ritual was originally not performed in order to achieve that specific goal. This self-observation becomes an important piece of learning data: namely, the information that so much investment is successful in coercing such a divine reaction (in general). Moreover, suddenly the other performances of the ritual with a lower level of *p* are retrospectively reinterpreted as being unsuccessful in bringing about this reaction. In turn, the usual learning algorithm starts working and will restore the intuitive theology to a higher *n* value.

In summary, the intuitive theology maintained by the congregant will fluctuate due to a number of factors. The equilibrium or attractor position (or positions) of these low-arousal rituals depend on, among other elements, the frequency of these random reinterpretations of events. The more often they occur, the higher the *n* of the attractor position – as future computer simulations should also demonstrate.

## 5. *Conclusion*

In this paper, religious rituals, or religious actions in general, were seen as transactions between a human agent and a culturally postulated superhuman agent. We have hypothesised that the human agent maintains a *behavioural grammar* driving their own actions, as well as a theory of mind for each of

---

'revitalising' the 'automated' ritual practice. See: Naccache, *Quatre exercices de pensée juive pour cerveaux réfléchis*, Chapter 2, especially p. 74 and p. 86. Related techniques to overcome the tedium effect in Judaism were discussed in my lecture 'Is Judaism Boring? The role of symbols in "imagistic" Jewish movements in the nineteenth century' (http://www.csr-arc.com/view.php?arc=17).

their transaction partners, which also has similar form as a grammar. In the case of a superhuman transaction partner, this second grammar was called the *intuitive theology*. I have argued that the cognitive science of religion should develop formal models of these grammars in order to run simulations that will subsequently be able to account for the observable dynamics of religious rituals. However, before arriving at this stage, an adequate model of behavioural grammar and intuitive theology must be found.

The word *grammar* in the expression *behavioural grammar* (such as one's own behavioural grammar and intuitive theology) does not mean that behaviour in general and religious practices in particular are reduced to language, or that we argue for a close connection between language and religion. The word simply refers to the source of the idea in the history of the discipline, the only hypothesis being that both language and religious behaviour are driven by higher cognition. As language was the first element of higher cognition that was approached from a cognitive perspective, the methodologies employed and the theories developed can serve as starting points in the cognitive science of religion. The same applies to *universal moral grammars*, recently argued for in a similar vein, a proposal that follows Chomsky's path very closely, both in its formalism and argumentation.[51]

Hence, we developed a model of behavioural grammar, and considered that the model of intuitive theology – a special type of behavioural grammar – would be very similar. The model had to be formal enough to be implemented on a computer, because computational implementations of cognitive architectures have been successfully applied to demonstrate the role of cognitive processes in social phenomena.[52] In this contribution we have chosen a cognitive model barely employed beyond linguistics, namely, Optimality Theory (OT), for several reasons.

First, following the argument of Lawson and McCauley[53] concerning the parallels in language and religion, we argue that a successful language model has the potential to also be useful in the cognitive science of religion. Since Optimality Theory is certainly an adequate model to describe lan-

---

[51] Mikhail, 'Universal moral grammar'. For a criticism, see Dupoux and Jacob, 'Universal moral grammar: a critical appraisal'. I have been recently referred to the work of Shweder et al., who introduced OT-like universal constraints more than ten years ago to account for sleeping arrangements in different cultures. The reference, which I could not check, may be Shweder, Jensen, Goldstein, 'Who sleeps by whom revisited', in: Goodnow et al., eds., *Cultural Practices as Contexts for Development*.

[52] See for instance the articles in Sun, ed., *Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation*.

[53] Lawson and McCauley, *Rethinking Religion*.

guage,[54] as demonstrated by the several hundred papers accumulated in the online Rutgers Optimality Archive in the last decade and half,[55] OT is a natural choice to describe behavioural grammars and intuitive theologies. Yet, it has to be emphasised that only the mathematical framework of OT has been borrowed, not its language-specific elements, such as the concrete constraints mentioned in Section 2.

Second, Optimality Theory is one of the linguistic architectures with most research having been carried out on its formal – mathematical and computational – properties, and especially on its learnability. Therefore, future work with OT in the cognitive science of religion can also rely on all these results concerning the framework itself, although of course, not the linguistic content.

Finally, Smolensky and Legendre (2006) convincingly argued that Optimality Theory is a model that is interpretable both on the lower levels of neuronal (connectionist) computations and on the symbol-manipulating higher levels used in complex tasks, such as language. At present, OT is probably the best model that can bridge the gap between the neurons of the brain and the manipulation of symbols by higher cognition. Therefore, if one is able to formulate a model of religion using Optimality Theory, then its realisation in a neurologically credible connectionist way is straightforward.

Moreover, a uniform handling of different cultural phenomena such as language and religion reinforces the cognitive plausibility of the models of both. This uniform handling might indeed adequately describe the way our mind/brain works in general, while details of the model can remain domain specific.

Even if coming from a different tradition, Optimality Theory can also be seen as a special variant of Rational Choice Theory, developed mainly in economics, which was introduced to the sociology of religion by Stark, Iannoccone and Bainbridge, among others. However, an important difference is that Rational Choice Theory is concerned with the group, and therefore

---

[54] Coincidentally, the term used by Smolensky and Legendre to describe their language model which combines OT with a connectionist implementation is the *cognitive science of language*. This expression should, however, not be confused with *cognitive linguistics*, the anti-Chomskyan approach mentioned in the introduction.

[55] The Rutgers Optimality Archive (ROA) and the related Optimality List can be found at http://roa.rutgers.edu. As an extremely useful research tool with respect to finding papers, but also in the dissemination of drafts and increasing the visibility of articles, ROA probably had an important role in popularising Optimality Theory in the 1990s. The Archive for Religion and Cognition (ARC), available at http://www.csr-arc.com, was modelled on ROA.

defines an initially oversimplified, then gradually improved model of the individual's choice, whereas Optimality Theory aims primarily at the cognitive bases of the behaviour of the individual. Hence, Optimality Theory needs a more complex model: crucially, its target function that is to be optimised is not real-valued.[56] Even if not presented as such in this article (for the sake of clarity), the target function to be optimised in Optimality Theory is generally much more abstract than the utility of Rational Choice Theory. The former, rooted in connectionism and describing fast, automatic and unconscious 'decisions', is believed to be a function of the physical state of the brain (its energy in a physical sense, if taking connectionism to the extreme). The utility function in Rational Choice Theory, accounting for more explicit personal decisions, is usually based on human concepts such as monetary value, social position or other interests. Unquestionably, the latter can also be reduced to states of the brain, and then the two theories may be combined.

After having introduced Optimality Theory in Section 2, we illustrated how to employ it as behavioural grammars. In particular, we demonstrated in Section 3 how to introduce constraints related to eating taboos and how to account for different traditions by establishing the permutations of the ranking of these constraints. Two kinds of learning must be distinguished: the constraint prohibiting the consumption of a certain substance can be promoted either after personal empirical experience or as a consequence of social learning. The second kind of learning provides knowledge without the individual having to experience negative situations themselves, but it also explains how food taboos can develop as by-products of the mechanism.

Subsequently, Section 4 aimed at explaining ritual dynamics by introducing a *transaction grammar* between human and superhuman agents. The key to ritual dynamics was that the congregant alters their intuitive theology based on experience: a perceived reaction of the superhuman agent to a specific human action, or the lack of such a reaction. Here, we have exploited the fact that Optimality Theory comes with an elaborate model of learnability.

Most of the models described could have been formulated using a simpler formalism. For example, the essence of tableaux (14) was that the intuitive theology maintained by a human agent about the postulated mind of the superhuman agent is simply an integer *n*: an offer is expected to be accepted if and only if its value is at least *n*. Why then introduce so many con-

[56] Bíró, *Finding the Right Words.* Note that Harmonic Grammar, the precursor of Optimality Theory, which is becoming popular once again, is closer to Rational Choice Theory in this respect.

straints? Why is the threshold *n* not an adequate grammar in itself? Thus far, it seems it is. However, the complex formalism in Optimality Theory has the advantage of being able to combine very different types of (non-quantifiable) factors in a non-trivial way. A more elaborate OT model can describe how one balances different types of pain, emotion, sacrifice of goods or the investment of energy and time in order to achieve different types of goals. This latter scenario cannot be described simply by introducing a single integer as the model of the intuitive theology in the congregant's mind.

In summary, we argue that Optimality Theory is a promising framework to explain religious rituals, or even human behaviour in general. The explanation should also cover the observable variations and dynamics. We have seen that Optimality Theory accounts for different types of language or human behaviour by showing that these types are produced by different constraint hierarchies (*factorial typology*). Likewise, the dynamics emerge from learning processes, that is, from re-ranking the same constraints based on experience.

This contribution only attempts to make a start in this field. Some technical details could not be explained due to lack of space, and at some points we could only indicate new directions. Future research should develop the model further.

*Acknowledgements*

I would like to thank Joseph Bulbulia, István Czachesz and Thomas Lawson for their comments on earlier versions of this paper.