Bayesian Algorithmic Modeling in Cognitive Science

Julien Diard

UNIVERSITÉ DE GRENOBLE

HABILITATION A DIRIGER DES RECHERCHES

École Doctorale Mathématiques, Sciences et Technologies de l'Information, Informatique

Spécialité : **Mathématiques Appliquées et Informatique**

Présentée par

Julien DIARD

préparée au sein du **Laboratoire de Psychologie et NeuroCognition**

Bayesian Algorithmic Modeling in Cognitive Science

Habilitation à diriger des recherches soutenue publiquement le **27 Novembre 2015**, devant le jury composé de :

Dr Frédéric ALEXANDRE
DR INRIA, INRIA Bordeaux Sud-Ouest
Examinateur

Pr Laurent BESACIER
PU Université Joseph Fourier, Laboratoire d'Informatique de Grenoble
Examinateur

Dr Pierre BESSIÈRE
DR CNRS, Institut des Systèmes Intelligents et de Robotique
Examinateur

Pr Philippe LERAY
PU Polytech'Nantes, Laboratoire d'Informatique de Nantes Atlantique
Rapporteur

Dr Pierre-Yves OUDEYER
DR INRIA, INRIA Bordeaux Sud-Ouest
Rapporteur

Pr Mark PITT
Professor Ohio State University, Language Perception Laboratory
Rapporteur

Contents

List of Figures

À Éva
qui, elle, sait écrire.

Acknowledgment

Writing a *habilitation* thesis is an odd exercise.

It consists of course of exposing scientific topics and results, in a domain which is hopefully well-known by the author. This part of the exercise is rather usual, with common trade-offs between details and conciseness, scientific rigor and care for pedagogical concerns, etc.

However, this thesis is also, in part, a biographical account, as it summarizes the scientific journey of the author and appears as a single-author manuscript. Attributing anything found in the remaining pages to me would be a gross mischaracterization at best, an agreed-upon lie. There are of course many other actors to consider, from the well-known shoulders of giants, to support and inspiration found in all places, most of them outside the lab. It is also conventional to acknowledge here all lab directors, team leaders, supporting agencies, many colleagues, friends and family; consider it done.

The silver lining to this biographical nature of this manuscript is that I took it as an opportunity to list and count the supervisors, collaborators and students I directly worked with. Supervisors and students were easy to count; for collaborators, I settled upon the list of those I have co-authored papers with. In this manner, I find that there are $N = 48$ people I have worked with over the years. Most, but not all of them, are mentioned in the following pages. Out of these, I find that about $n = 12$ of them are, to my eyes, exceptional scientists. Of course, I will not name names, so as to stay cordial to the others. Also, and more importantly, this estimation is really tongue-in-cheek; I would not presume to judge anyone here, lest they judge me in return, which would evidently be unpleasant. In any case, that number n is certainly larger that expected. Let me offer a back of the envelope estimation of my luck.

Assuming being "exceptional" is to lie beyond the $\mu + 2\sigma$ boundary, each colleague being exceptional has roughly probability $p = .025$. Having $n = 12$ out of $N = 48$ colleagues being exceptional has probability $\binom{N}{n} p^n (1-p)^{N-n} = 1.6691 * 10^{-9}$. Looser criteria for being exceptional of course increase this number, but even if it included ten percent of the population, this number would still be very small (around 10^{-3}).

Thus, I find that having such an exceptional collaboration network had a very small, ϵ probability to occur. I therefore find myself extremely lucky, and I have demonstrated it. As a consequence, I feel extremely thankful, but you will have to take my word for it, as, of course, you do not have direct access to my cognitive constructs.

It will be no surprise that probabilities and building models of presumed cognitive entities are themes that will permeate the following pages.

CHAPTER 1

Introduction

My research domain is **Bayesian modeling of sensorimotor systems**. I am interested both in natural and artificial sensorimotor systems. In other words, my research is multidisciplinary by nature, at the crossroads, on the one hand, of mathematics and computer science, and, on the other hand, of cognitive science and experimental psychology.

By "**sensorimotor systems**" above, I mean systems with sensors and actuators, like animals and robots, that perceive, reason and act, not necessarily in this order. Indeed, a classical, coarse-grained view of information processing in both experimental psychology and robotics considers a three-step process: first, information collection and pre-processing to acquire or update internal representations of external phenomena of interest (perception), second, computation involving these internal representations to prepare future actions (planning), and third, action execution and monitoring (action).

This very simple model of information processing is, like most models, both wrong and useful. Indeed, it is useful, for instance, as a rough guide to understand information flow in neurobiology, and at the same time, it is wrong because it hides away some of the complexity of said information flow (sensory prediction, temporal loop effects, top-down tuning of perception, etc.). I am interested in models of information flow that depart from the classical, three-piece "perceive, plan, act" schema, and instead include intermediate representations, modular subsystems, complex structures, multi-layered hierarchies, with intricate information flow. I will call these "representational cognitive models", or, in order to follow Marr's terminology, "algorithmic cognitive models" (Marr; 1982). To define such models, I need a mathematical language that is both flexible and powerful, and allows expressing assumptions in an transparent manner, so as to ease the design, communication, interpretation and comparison of models.

By "**Bayesian modeling**" above, I mean, first and foremost, using probabilities as a mathematical language for knowledge representation and manipulation. This is also known as the "subjectivist approach to probabilities", or "subjectivist Bayesianism" (Bessière et al.; 1998a,b, Colas et al.; 2010, Fiorillo; 2012). More specifically, I follow here the seminal works of Pierre Bessière, who himself followed Edwin T. Jaynes (2003) [1]. Let me acknowledge upfront, in this short introduction, that a thorough discussion of the various meanings of "Bayesian modeling",

[1] One of the unsung heroes of Artificial Intelligence, which is somewhat understandable since he was a physicist.

and their application to other layers of Marr's hierarchy are in order, but reserved for a later portion of this document (Chapter 5).

In the subjectivist approach to probabilities, probabilities are considered as an extension of logic calculus for uncertain knowledge representation and manipulation. More precisely, and thanks to Cox's theorem, probabilities can actually be shown to be the only formal system suitable for this task, with the word "suitable" given a precise technical meaning (Jaynes; 1988, 2003, van Horn; 2003). Since it is concerned with knowledge representation and manipulation, this approach considers probabilities as measures of states of knowledge of a given subject; hence the "subjectivist" part of the name, which has nothing to do with arbitrariness. Instead of defining "subjective" probabilistic models, however, and to avoid any unfair pejorative interpretation of the word, I will rather refer to "cognitive probabilistic models", or "Bayesian cognitive models". A crucial feature of this approach is that it transforms the irreducible incompleteness of cognitive models into quantified uncertainty. This uncertainty is thus explicit in the probability distributions representing knowledge, and manipulated explicitly during Bayesian inference.

Bayesian Programming is the name of the methodology I use for defining structured probabilistic models (Bessière et al.; 2013). In a nutshell, Bayesian Programming is a two-step methodology:

- first, one defines a joint probability distribution over variables of interest, possibly identifying its free parameters using learning,
- second, one defines tasks to be solved by computing probability distributions of interest, that are usually not readily available, but must be computed from the joint probability distribution using Bayesian inference.

Full-fledged computer languages could be developed to implement Bayesian Programming. So far however, implementations I have used have taken the form of APIs and libraries, instead; they are called "inference engines" (I used the Probability as Logic (PaL) Common LISP library, the ProBT C++ API (ProBayes, France), or ad hoc solutions). In all cases, the programmer first translates knowledge in formal terms, then asks queries, which are answered automatically, by the inference engine, using Bayesian inference.

In that sense, Bayesian Programming, if it were implemented as a language, would be a declarative one, and could be seen as a "probabilistic Prolog". This contrasts sharply with the recent wave of tools for "probabilistic programming", which are mostly imperative in nature (see e.g., Goodman et al. (2008), Gordon et al. (2014)); that is to say they provide structures and functions for defining and sampling probability distributions, but do not constrain the programmer to first formally define a probabilistic model. Therefore, in Bayesian Programming, focus is less on how inference is performed, than on what knowledge is involved in inference. In other words, we do not model processes directly, we model knowledge that yields processes.

From artificial intelligence and programming to cognitive science and modeling

Bayesian Programming was originally developed in Pierre Bessière's research group in various contexts in robotic programming (Bessière et al.; 2008), such as behavior programming (Lebeltel; 1999, Diard and Lebeltel; 1999, Lebeltel et al.; 2000, Diard and Lebeltel; 2000, Lebeltel et al.; 2004, Simonin et al.; 2005, Koike et al.; 2008), robotic perception (Ferreira et al.; 2008, Lobo et al.; 2009, Ferreira et al.; 2011, 2012a,b, 2013, Ferreira and Dias; 2014), CAD applications

(Mekhnacha; 1999, Mekhnacha et al.; 2001), automotive applications (Pradalier et al.; 2005, Coué et al.; 2006, Tay et al.; 2008), robotic mapping and localization (Diard et al.; 2004a,b, 2005, Ramel and Siegwart; 2008, Tapus and Siegwart; 2008, Diard et al.; 2010a). It was also applied to artificial agent programming (Le Hy et al.; 2004, Synnaeve and Bessière; 2011a,b, 2012a,b) and to cognitive modeling, in a broad sense (Serkhane et al.; 2005, Colas et al.; 2007, Laurens and Droulez; 2007, Colas et al.; 2009, Möbus and Eilers; 2009, Möbus et al.; 2009, Colas et al.; 2010, Eilers and Möbus; 2010, Lenk and Möbus; 2010, Möbus and Eilers; 2010a,b, Eilers and Möbus; 2011). Works I have been involved with, in this last domain, will be extensively discussed in this *habilitation*, and thus more appropriately cited elsewhere.

Bayesian programming and Bayesian modeling are, however, two different matters. Programming is an engineering endeavor, modeling is a scientific one. In engineering, the goal is to build something that did not exist before; in modeling however, the goal is to understand an object of study that is preexisting.

It turns out that modeling is much easier than programming. This is not to be understood as bragging; it is instead a real concern, with a technical cause and far-reaching epistemological consequences. The technical cause is the power of expression of Bayesian modeling. Indeed, recall that the subjectivist approach to probabilities defines them as an extension of logic calculus. It follows that all models and computations based on logic are special cases of Bayesian models and Bayesian inference. Therefore, any function that is computable in logic is computable in probabilities [2]. Bayesian modeling, in that sense, is too powerful, and can express any function whatsoever.

In other words, Bayesian modeling in itself is not a scientific proposition, as it certainly cannot be refuted. This would seem to mean that the Bayesian brain theory is not a valid scientific theory; however, this is not so obvious, and refinements of this proposition need to be discussed. At the moment, however, I just highlight the precaution that building Bayesian models of cognition is to be distinguished from claiming that the brain somehow would be Bayesian. I postpone developments of this discussion to a later portion of this manuscript, where I propose to analyze the recent debate, in the literature, about the status and contribution of Bayesian modeling; this debate stems from the same analysis I just introduced (Chapter 5).

A pragmatic consequence of this analysis is that, whatever the object of study, building *one* Bayesian model of it is never a problem, and therefore, never an interesting goal [3]. Unfortunately, Building *a good* Bayesian model is not a viable alternate objective, as model goodness seldom has a clear, non-equivocal, absolute meaning.

Contribution: comparison of Bayesian algorithmic cognitive models

If building *a* Bayesian model is useless and if building *a good* Bayesian model is a red herring, what should be our goal, then? I propose that a sensible goal is to build *a better* Bayesian model than another (Bayesian) model. In other words, I will focus on model evaluation and model comparison.

Over recent years, I have had the chance to work with colleagues and students on Bayesian modeling of several cognitive functions, most notably reading and writing on the one hand, and

[2] I sometimes wonder whether this implies the existence of a "probabilistic extension of the Turing machine", that would be different from non-deterministic Turing machines, and, if yes, what its properties would be.

[3] On several occasions, people have come to ask me: "I am very interested in X, Bayesian models appear fashionable, and you're the lab's Bayesian guy; do you think it would be possible for you to build a Bayesian model of X?". I have always answered "Why, yes of course it would be possible". Lack of response at that true, but useless, answer usually indicated without fail that collaboration would be near impossible.

speech perception and production on the other hand. In each case, we have defined structured probabilistic models, as discussed above, using Bayesian Programming. **Our first contribution is a set of Bayesian algorithmic cognitive models**. This has led us to mathematically formulate assumptions about the structures of cognitive processes and the possible internal representations involved. This step usually was fruitful in itself, as expressing hypotheses in a Bayesian model forced us over and again to make these assumptions technically precise, usually more than they were described in the literature.

Of course, defining complex structured models leads to methodological difficulties. For instance, given some theoretical assumption in a scientific domain, there is never a unique, non-ambiguous manner to translate it into probabilistic terms. Moreover, when a model grows, its number of free parameters also grows. We could not be satisfied with obtaining a single model, as its evaluation would ultimately rest on the plausibility of the probabilistic structure of the model (which could somewhat be interpreted) and the plausibility of the parameter values (which usually cannot be interpreted, since most parameters we consider do not have well-defined physical sense that we could calibrate from the literature, independently of our modeling endeavor at hand).

We have therefore been careful to systematically define variants of our model, and then study their comparison. Everything else being equal, we have studied the effect that a small variation in the model would have. This is standard fare in the scientific methodology, as it allows to, in all likelihood, impute some observed effect to the experimental manipulation, and not to background assumptions, which, however faulty, were held constant. **Our second contribution is a methodology for the comparison of Bayesian algorithmic cognitive models.**

We have particularly been intrigued by negative answers to such comparison, what we call indistinguishability theorems, a.k.a. "mimicry theorems" (Chater; 2005, Chater and Oaksford; 2013): even models that appear different, e.g., when they involve different internal representations, sometimes yield the same exact experimental predictions. This helps pinpoint so called "crucial experiments", that is to say, the experimental conditions necessary to distinguish models, and thus, to compare theoretical propositions at the origin of the models.

Model comparison takes many different forms, from well-quantified comparison of model goodness of fit, various criteria quantifying model complexity, to less formal evaluation about plausibility of assumed mechanisms and representations, based on neuronal plausibility, ability to reproduce classical effects, distinguishability from other models, etc. A major advantage of probabilistic modeling, in this regard, is that the same mathematical language is used to define models and formally express their comparison; this yields Bayesian comparison of Bayesian models. In this context, I have been interested in quantifying, in Bayesian terms, the distinguishability of Bayesian models; this led me to define a meta-model for Bayesian distinguishability of models. Unfortunately, as this point, I did not have the opportunity to apply this tool to the models I describe in this document; so far, this stays a side-project to my main research program (see Annex A.6).

Application of Bayesian Algorithmic Cognitive Modeling

Whatever its granularity, model comparison is, in our view, the necessary basis for answering scientific questions about the object of study. In this document, I will aim to illustrate this using five examples of comparison of Bayesian algorithmic cognitive models, from our recent research:

1. In the domain of reading and writing of cursive letters, we have been interested in the pos-

sible function of motor activations that are observed during letter identification. To answer this question, **we have defined the Bayesian Action-Perception (BAP) model.** In this model, we could compare recognition of letters with and without internal simulation of movement. This led us to the answer that internal simulations of movement would be redundant in normal conditions, but useful in difficult conditions, e.g., when the stimulus was partly corrupted.

2. In the domain of visual word recognition, we have been interested in the possible function of visual attention. To answer this question, **we have defined the Bayesian word Recognition using Attention, Interference and Dynamics (BRAID) model.** in this model, we aim to compare word recognition with and without attentional limitations, in order to test the hypothesis that attention would be a crucial component in reading acquisition, and that visual attention deficits would slow down word recognition.

3. In the domain of phonological system emergence, we have been interested in the properties that communicating agents would require to explain the observed regularities in phonological systems of languages (the so-called "language universals"). To answer this question, **we have defined the Communication of Objects using Sensori-Motor Operations (COSMO) general model architecture, and applied it in the emergence case, with the COSMO-Emergence model.** In this model, we could compare phonological evolution of communities of agents with or without concern for communication pressure. This led us to the answer that purely motor agents would not be able to converge towards common codes, contrary to sensory or sensorimotor agents; realistic simulations of vowel and syllable emergence yielded the wanted regularities.

4. In the domain of speech perception, we have been interested in the possible function of motor activations that are observed, in some cases, during speech perception. To answer this question, **we have defined the COSMO-Perception model, a variant of the COSMO model architecture for syllable perception.** In this model, we could compare purely auditory, purely motor and sensorimotor speech perception. This led us to the answer that motor and auditory perceptions would be, in some learning conditions, perfectly indistinguishable; this helps explore experimental conditions (e.g., imperfect learning, adverse conditions) in order to make then distinguishable, and study their functional complementarity (motor perception would be wide-band, auditory perception would be narrow-band).

5. Finally, in the domain of speech production, we have been interested in the origin of intra-speaker token-to-token variability in the production of sequences of phonemes. To answer this question, **we have a defined the COSMO-Production model, a variant of the COSMO model architecture for the production of sequences of phonemes.** We could compare COSMO-Production with GEPPETO, a twin model, that is to say, built upon the exact same assumptions, but mathematically expressed in the optimal control framework. This led us to the answer that COSMO-Production contains GEPPETO as a special case; it provides a formal equivalence between a Bayesian Algorithmic Model, where every piece of knowledge is encoded as a probability distribution, and an optimal control model involving a deterministic cost function. It also provides a way to model token-to-token variability from representational constraints in a principled manner.

Roadmap of this *habilitation*

The main objective of this *habilitation* manuscript is to summarize the main results introduced above, and then discuss their place in the current panorama of Bayesian modeling in cognitive science. To do so, the rest of this document is structured as follows.

Chapter 2 provides a primer to Bayesian Programming, discussing as intuitively as possible the main components and properties of Bayesian Programs, while still introducing necessary mathematical notations and elementary constructs, such as coherence variables.

Chapters 3 and 4 summarize the five cognitive models that constitute our main contribution; Chapters 3 concerns the study of reading and writing, and thus the BAP and BRAID models, whereas Chapter 4 concerns the study of speech perception and production, and thus the COSMO model and its variants, COSMO-Emergence, COSMO-Perception and COSMO-Production. In each case, we focus our presentation on the coarse-grained model structure, on the manner Bayesian inference was used to solve cognitive tasks, and on the way model comparison was used to explore scientific questions. Each model presentation is prefaced by biographical notes and bibliographical information, to guide the curious reader to further material.

Chapter 5 contains a discussion of the usual acceptation of the term "Bayesian modeling" in cognitive science, contrasting it with our approach. We also analyze the epistemological status and relevance of Bayesian modeling as a research program for the scientific study of cognition. This chapter should clarify some issues that have been merely hinted at in the current introductory chapter.

After a final chapter discussing perspectives for future work (Chapter 6), I provide as annex material first a bullet-list description of other research projects that I have been involved with over recent years (Annex A), and second an up-to-date *curriculum vitæ*, in French, containing information about grant support and activities other than research, such as student supervision (Annex B).

CHAPTER 2

Bayesian Programming

In this Chapter, I provide a cursory introduction to probabilities and Bayesian Programming, with the aim to make accessible the descriptions of Bayesian cognitive models of subsequent chapters. Portions of this Chapter are adapted from previous material (Diard; 2003, Colas et al.; 2010, Gilet et al.; 2011). More detailed presentations of Bayesian Programming are available elsewhere (Lebeltel et al.; 2004, Bessière et al.; 2013).

2.1 Preliminary: Probabilities and probabilistic calculus

The mathematical background required for this *habilitation*, and, one could argue to some extent, to practice Bayesian Programming and Bayesian modeling, is rather light. Indeed, since our focus is on modeling, and thus expressing knowledge in a mathematical form, we are interested in mathematical constructs with building ingredients that are easily interpreted, and easily manipulated. Our mathematical "vocabulary" will thus be, in most cases, limited to classical probability functions. Our mathematical "syntax" will be the rules of probability calculus, that is to say the sum rule and the product rule. We quickly recall these ingredients here, as a mathematical warmup session.

2.1.1 Subjective probabilities

Concerning the notion of probability itself, we provide no formal definition here, and assume the reader somewhat familiar with the notion. However, we underline a specificity of the subjectivist approach to probabilities in this matter. In the subjectivist approach to probabilities, formally, the probability $P(X)$ of a variable X does not exist. Because we are modeling the knowledge a subject π has about variable X, we must always specify who the subject is. Therefore, we will always use conditional probabilities: $P(X \mid \pi)$ describes the sate of knowledge that the subject π has about variable X.

This allows to formally reason about states of knowledge and assumptions. For instance, when two subjects π_1 and π_2 have different knowledge about X, the probability distributions $P(X \mid \pi_1)$ and $P(X \mid \pi_2)$ differ. Given a series of observations about X, Bayesian inference then allows to formally reason about whether the knowledge and assumptions of π_1 or π_2 better

explain these observations. This is not just a technicality: this is the essence of Bayesian inference in general, and the formal basis for model comparison.

This need for specifying the modeled subject radically separates the subjective approach to probability from the frequentist approach. Indeed, in the frequentist approach, the probability is a property of the event itself, whoever the observing subject is. Frequentist probabilities are, in that sense, ontological, whereas subjective probabilities are epistemic (Phillips; 2012). In the frequentist conception of probability, then, it is natural to note $P(X)$ as a property of X solely. This somewhat limits model comparison, which is more natural to express in the subjectivist approach to probability.

Of course, even a die-hard subjectivist realizes that notation is made more cumbersome if all right-hand sides of probability terms must refer to the subject being modeled. That is why, of course, it will be left implicit when possible.

Other notation simplifications and unorthodoxies must be mentioned here. For instance, in the statistical notation, it is customary to note series of variables using commas between variables. However, we do not follow this convention here, as series of variables are formally conjunctions of variables (see below, Section 2.2.1), with the classical logical conjunction operator $\wedge$ becoming a space when left implicit; therefore we will note $P(X_1\ X_2 \mid \pi)$ instead of $P(X_1, X_2 \mid \pi)$.

Also, because our models mostly consider discrete variables, and sometimes mix discrete and continuous variables, and since our focus is not on the implementation of Bayesian inference, we will simplify notations and consider the discrete case as default. As a consequence, contrary to the usual notation that distinguishes and uses $P(\cdot)$ for the probability distributions in the discrete case and $p(\cdot)$ for probability density functions in the continuous case, all probability terms will be noted with the $P(\cdot)$ notation. In the same manner, all integrals will be denoted with the discrete summation symbol, $\sum$.

2.1.2 Probability calculus

Probability calculus only relies on two mathematical rules, called the sum rule and the product rule (a.k.a., the chain rule).

The sum rule states that probability values over a variable X sum to 1:

$$\sum_X P(X \mid \pi) = 1 \ . \tag{2.1}$$

The product rule states how the probability distribution over the conjunction of variables X_1 and X_2 can be composed of the probability distributions over X_1 and X_2:

$$P(X_1\ X_2 \mid \pi) = P(X_1 \mid \pi)P(X_2 \mid X_1\ \pi) = P(X_2 \mid \pi)P(X_1 \mid X_2\ \pi) \ . \tag{2.2}$$

The product rule is better known as a variant that derives from it, called Bayes' theorem:

$$P(X_1 \mid X_2\ \pi) = \frac{P(X_1 \mid \pi)P(X_2 \mid X_1\ \pi)}{P(X_2 \mid \pi)}, \text{ if } P(X_2 \mid \pi) \neq 0 \ . \tag{2.3}$$

Most of the time, the distinction between the product rule and Bayes' theorem is not used, and both names and equations are used interchangeably.

From the sum rule and product rule, we can derive a very useful rule, called the marginalization rule:

$$P(X_1 \mid \pi) = \sum_{X_2} P(X_1\ X_2 \mid \pi) \ . \tag{2.4}$$

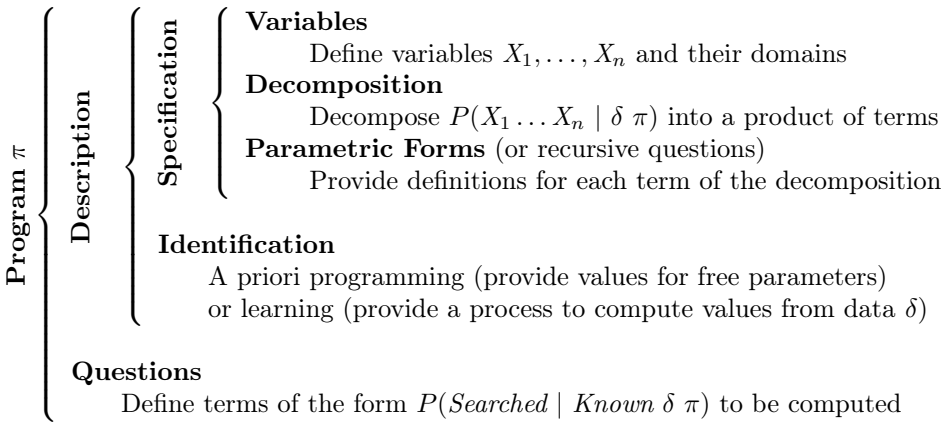

Figure 2.1: **Commented structure of the Bayesian Programming methodology.**

2.2 Bayesian Programming methodology

Bayesian Programming is a methodology for defining Bayesian Programs. It is a structured guideline, following the architecture shown Figure 2.1.

A Bayesian Program (BP) contains two parts:

- the first is declarative: we define the *description* of a probabilistic model;
- the second is procedural: we specify one or several *questions* to the probabilistic model described in the first part.

A description itself contains two parts:

- a *specification*: we define the joint probability distribution of the model, that encodes the modeler's knowledge about the phenomenon of interest;
- an *identification* process: we define methods for computing values of the free parameters of the joint probability distribution.

Finally, a specification contains three parts:

- a selection of *relevant variables* to model the phenomenon;
- a *decomposition*, whereby the joint probability distribution on the relevant variables is expressed as a product of simpler distributions, possibly including conditional independence assumptions to further simplify terms;
- the *parametric forms* in which each of the terms of the decomposition is associated with either a given mathematical function or a recursive question to another BP.

A Bayesian Program is formally uniquely identified by a pair of symbols: the first represents the set of preliminary knowledge used for defining this precise model (usually denoted π in our practice, or some subscripted variant of π, e.g., π_{BAP} for the BAP Bayesian Program), the second represents the experimental data used during the identification phase (usually denoted δ in our practice, or some variant of it). Therefore, and as before (Section 2.1.1), being formally rigorous would require noting the δ, π symbols in all right-hand sides of all probabilistic terms of

Bayesian Programs, for instance in the decomposition, in parametrical forms, etc. However, of course, they are usually unambiguous and thus left implicit, which simplifies greatly notations. We will follow this practice in most of the remainder of this document, except for situations like recursive questions to sub-models, or model comparison, where they become technically necessary.

This overview shows that for a modeler to define a BP, five steps must be followed in sequence: first, define the relevant variables, second, decompose the joint probability distribution, third, associate parametric forms or recursive questions to each term of the decomposition, fourth, define how free parameters are to be identified, and fifth and finally, define questions of interest to compute using Bayesian inference. We now turn to each of these five steps, to provide technical details.

2.2.1 Probabilistic variables

Because the subjectivist approach to probabilities can be seen as an extension of logic, it is grounded in logical propositions and logical operations. We define a probabilistic variable X as a set of k logical propositions $\{[X = x_1], [X = x_2], \ldots, [X = x_k]\}$, with the following two properties:

1. exhaustivity: at least one of the propositions $[X = x_i]$ is true;

2. mutual exclusion: no two propositions $[X = x_i], [X = x_j], i \neq j$ can be true simultaneously.

We call the symbols $x_1, x_2, \ldots, x_k$ the values of the probabilistic variable X, the set of these values is the domain of the variable, and k is its cardinal. Notation is usually simplified by replacing the proposition $[X = x_i]$ by x_i, whenever it is unambiguous that x_i is the value taken by variable X. For instance, $P(X \mid y)$ is a single probability distribution, $P(X \mid Y)$ is a set of probability distributions (one for each possible value of variable Y), and $P(x \mid y)$ is a probability value. In the notational convention we follow, lowercase symbols usually refer to values, and names that start with uppercase letters refer to variables.

Note that this definition of probabilistic variables can be extended to the continuous case, as the limit when the number of logical propositions grows to infinity. This is not without pitfalls however, and usually it is without much practical interest, especially when the goal is computer simulation of a model, which ultimately requires a discretization process (except for the rare cases which have analytical solutions). In most of our contribution therefore, discrete models will be presented, so as to tackle explicitly measurement precision and representational accuracy.

The conjunction $X \wedge Y$ of two probabilistic variables X and Y, of domains $\{x_1, \ldots, x_k\}$ and $\{y_1, \ldots, y_l\}$, is also a probabilistic variable: it is the set of the $k * l$ propositions $[X = x_i] \wedge [Y = x_j]$, $\forall i, j$. This is the case because the exhaustivity and mutual exclusion properties still hold, which is straightforward to prove. The conjunction of probabilistic variables corresponds to the intuitive notion of multi-dimensional variables, whose domains are the product of domains of each dimension.

However, the disjunction $X \vee Y$ of probabilistic variables is not a probabilistic variable, as the mutual exclusion property does not hold in the set of all propositions of the form $[X = x_i] \vee [Y = x_j]$. For this reason, variable disjunction is never employed in our work; this is not because of a technical impossibility, but more out of convenience.

2.2.2 Joint probability distribution decomposition

The goal of the declarative part of a Bayesian Program $\langle \pi, \delta \rangle$ is to fully define the joint probability distribution $P(X_1\ X_2\ \ldots\ X_n \mid \pi\ \delta)$. In most cases, this multidimensional probability distribution is not easily defined as is. The product rule allows to decompose this joint probability distribution into a product of lower-dimensional probability distributions. Instead of providing a formal definition of this process (Bessière et al.; 2008), we illustrate it on a four-variable example. For instance:

$$P(X_1\ X_2\ X_3\ X_4 \mid \pi\ \delta) = P(X_1\ X_2 \mid \pi\ \delta)P(X_3 \mid X_1\ X_2\ \pi\ \delta)P(X_4 \mid X_1\ X_2\ X_3\ \pi\ \delta)\ . \tag{2.5}$$

This product can be further simplified, by stating conditional independence hypotheses, that allow to cross off variables of right-hand sides of terms. For instance, assuming that X_3 is independent of X_2, conditionally on knowing X_1, and that X_4 is independent of X_1 and X_3, conditionally on knowing X_2, yields:

$$P(X_1\ X_2\ X_3\ X_4 \mid \pi\ \delta) = P(X_1\ X_2 \mid \pi\ \delta)P(X_3 \mid X_1\ \pi\ \delta)P(X_4 \mid X_2\ \pi\ \delta)\ . \tag{2.6}$$

Note that, formally, Eq. (2.5) and (2.6) have different semantics. In Eq. (2.5), the equality sign is a real mathematical equality, so that, even though a product of low-dimensional terms appears, the complexity of the model is not reduced (we replaced one many-dimensional term by many low-dimensional terms). In contrast, in Eq. (2.6), the equality sign is a "physicist" equality, as it translates simplifying assumptions into the model, which may be adequate or may be completely wrong. When chosen wisely, they drastically break down the complexity of the model without much loss in model accuracy.

Of course, there are many ways to decompose a given joint probability distribution; the modeler is usually driven, however, by the will to make appear, in the chosen product, terms that are easily interpreted, easily defined, or easily learned, and the trade-off between model accuracy and model complexity. This is a crucial part of the modeler's art.

2.2.3 Parametric forms

Once the joint probability distribution is decomposed into a product of terms, each term must be given a formal definition. This is done by associating each term with a probability law; the modeler has a wide array of choices, from the usual and most common probability laws (e.g., uniform distributions, normal (Gaussian) distributions, conditional probability tables, etc.) to more specific and exotic choices (e.g., discrete truncated Gaussian distributions, von Mises distributions (Diard et al.; 2013a), Laplace succession laws (Diard et al.; 2013b), mixture models, etc.). An exhaustive list of suitable probability distributions would of course be beyond the scope of this manuscript; instead, in the following Chapters, we will recall the necessary probability distributions if and when they appear.

We note that the result of this step of Bayesian Programming is to associate a parameter space to each term of the decomposition. For instance, provided that X_3 is a continuous variable over $\mathbb{R}$, and that X_1 is a discrete variable of cardinal k, defining the term $P(X_3 \mid X_1\ \pi\ \delta)$ of Eq. (2.6) as a set of Gaussian probability distributions implies that there are $2k$ free parameters to this term. They are the means μ_i and variances σ_i^2 of Gaussian probability distributions, one for each of the k values of X_1.

Sometimes, making these parameters explicit in the notation allows to model the learning process; for instance, $P(X_3 \mid X_1\ \pi\ \delta)$ becomes $P(X_3 \mid X_1\ \mu_1\ \ldots\ \mu_k\ \sigma_1^2\ \ldots\ \sigma_k^2\ \pi\ \delta)$, and is complemented in the model by a prior distribution over parameters $P(\mu_1\ \ldots\ \mu_k\ \sigma_1^2\ \ldots\ \sigma_k^2 \mid \pi\ \delta)$. This enables inference about parameter values given subsequent observations of variables $\langle X_3, X_1 \rangle$. In other words, this yields a model of the learning process as Bayesian inference.

2.2.4 Identification of free parameters

However, when the learning process is not the focus of the Bayesian model, the free parameters are left implicit in the probabilistic notation, and their update according to experimental data is treated algorithmically. This is the aim of the "identification" phase of Bayesian Programming. Here, the modeler either describes an algorithmic learning process, or manually defines parameter *a priori*. Terms that do not depend on experimental data can then formally omit the δ of their right-hand side.

Either way, at this end of this stage, all free parameters are set, so that all terms of the decomposition are fully defined, so that the joint probability distribution $P(X_1\ X_2\ \ldots\ X_n \mid \pi\ \delta)$ is also fully defined. This concludes the declarative phase

2.2.5 Probabilistic questions and Bayesian inference

Once the joint probability distribution $P(X_1\ X_2\ \ldots\ X_n \mid \pi\ \delta)$ is defined, the declarative phase of Bayesian Programming is completed. We interpret the joint probability distribution as the cognitive model, that is to say the mathematical expression of the knowledge available to the cognitive subject. We then enter the procedural phase, in which the cognitive model is used to solve cognitive tasks. To do so, we assume that a cognitive task is modeled by one or several *probabilistic questions* to the model, which are answered by Bayesian inference, without further involvement of the modeler. This is a particular stance regarding cognitive modeling: in our approach, the modeler's goal is not to directly model cognitive processes, but to model a set of knowledge that yields processes. This ensures mathematical coherence between the resulting cognitive processes.

We call a probabilistic question any probability term of the form $P(\mathit{Searched} \mid \mathit{Known}\ \pi\ \delta)$, where the sets of variables *Searched* and *Known*, along with a third set noted *Free*, form a partition of $X_1\ X_2\ \ldots\ X_n$ (with $\mathit{Searched} \neq \emptyset$). *Searched*, *Known* and *Free* respectively denote the variables we are interested in, the variables whose values are observed at the time of inference, and the remaining, unconstrained variables.

A theorem states that, given a joint probability distribution, any probabilistic question can be computed using Bayesian inference, in a systematic, automatized manner. This is demonstrated by a constructive proof, that is to say, by showing how any question is answered using Bayesian inference, by referring only to the joint probability distribution (Bessière et al.; 2003, Bessière et al.; 2008). The proof is as follows.

Given any partition of the n variables $X_1\ X_2\ \ldots\ X_n$ into three subsets *Searched*, *Known* and

Free, the question $P(Searched \mid Known\ \pi\ \delta)$ is computed from $P(X_1\ X_2\ \ldots\ X_n \mid \pi\ \delta)$ by:

$$\begin{aligned} P(Searched \mid Known\ \pi\ \delta) &= \frac{P(Searched\ Known \mid \pi\ \delta)}{P(Known \mid \pi\ \delta)} \\ &= \frac{\sum_{Free} P(Searched\ Known\ Free \mid \pi\ \delta)}{\sum_{Searched,\ Free} P(Searched\ Known\ Free \mid \pi\ \delta)} \\ P(Searched \mid Known\ \pi\ \delta) &= \frac{\sum_{Free} P(X_1\ X_2\ \ldots\ X_n \mid \pi\ \delta)}{\sum_{Searched,\ Free} P(X_1\ X_2\ \ldots\ X_n \mid \pi\ \delta)}\ . \end{aligned} \tag{2.7}$$

This derivation successively involved Bayes' theorem and the marginalization rule.

Note that this derivation, of course, only holds if $P(Known \mid \pi\ \delta) \neq 0$. This corresponds to automatically avoiding probabilistic questions that assume a set of observations with probability 0. In other words, probabilistic questions that start from a set of impossible assumptions are mechanically excluded by the formalism, yielding a mathematical dead-end; this is a remarkable built-in safeguard of Bayesian inference.

When computing $P(Searched \mid Known\ \pi\ \delta)$, $P(Known \mid \pi\ \delta)$ can be seen as a constant, as it only depends on the values of *Known*, which are fixed in this probabilistic question. In other words, the denominator of Eq. (2.7) is a constant value. Eq. (2.7) provides a manner to compute this value. Another manner is to compute everything up to a constant Z, and normalize afterwards, which is sometimes faster. We use the $\propto$ symbol to denote equality in proportionality:

$$\begin{aligned} P(Searched \mid Known\ \pi\ \delta) &= \frac{1}{Z}\sum_{Free} P(X_1\ X_2\ \ldots\ X_n \mid \pi\ \delta) \\ &\propto \sum_{Free} P(X_1\ X_2\ \ldots\ X_n \mid \pi\ \delta)\ . \end{aligned} \tag{2.8}$$

The joint probability distribution of Eq. (2.8) is itself defined as a product of terms, so that any inference amounts to a number of sum and product operations on probability terms. Of course, this brute force inference mechanism sometimes yields impractical computation time ans space requirements, as Bayesian inference in the general case is $\mathcal{NP}$-hard (Cooper; 1990).

Recall that computational complexity actually concerns the worst possible case in some class of problems, and does not say anything about the common case. It is true that Bayesian inference is intractable sometimes, but this concerns unstructured, flat models that describe brutally a high-dimensional state space; such models are usually not interesting (making the point of Kwisthout et al. (2011) technically correct but somewhat moot in practice).

In most usual cases however, as in the ones in the present manuscript, the model is highly structured, which helps inference. That is to say, the first practical step after Eq. (2.7) is to replace the joint probability distribution by its decomposition, which usually results in an expression with a sum (over *Free*) of a product of terms. In a symbolic simplification phase, these sums and products are reordered, factoring out terms, and the resulting expression is simplified whenever possible.

Follows a numerical computation phase, where a variety of classical techniques are available to exactly or approximately compute the terms, depending on the model structure and properties. At this stage, recognizing that the Bayesian Program at hand is of a known family (e.g., a Kalman filter, a Hidden Markov Model, etc) sometimes provides specific and efficient inference algorithms. Furthermore, since most of our models include Dirac probability distributions, they can be replaced by deterministic functions that break down Bayesian inference in several independent processes, articulated by algorithms that transmit variable values only. In other words, we

sometimes define formally our inferences in a fully probabilistic framework, and then implement them using a combination of deterministic and probabilistic programs, for efficiency.

All of the inferences described in the remainder of this manuscript have either been carried out either using a general purpose probabilistic engine, ProBT (ProBayes, Grenoble, France), which also integrates custom methods for representation and maximization of probability distributions (Bessière; 2004, Mekhnacha et al.; 2007), or custom code in various languages (e.g., Mathematica, Matlab, C).

2.2.6 Decision model

The output of Bayesian inference is a probability distribution of the form $P(Searched \mid Known\, \pi\, \delta)$, that models a cognitive task. In some cases however, such a distribution is not an adequate model of the known output format of the observed cognitive process. That is the case whenever the process clearly outputs a unique value; in that case, the process appears to end in a *decision*.

If we assume that this decision is based on the available knowledge, that is to say, on the answer to a probabilistic question, then to model this decision step, we need to describe a manner to go from a probability distribution to a single value. There are two classical decision models: one is to assume that the value with maximum probability is output, the other is to assume that values are drawn according to their probabilities. These two decision models have different properties.

When the process only outputs the value of maximum probability, it ensures that the "best" solution, according to available knowledge, is selected [1]. If knowledge does not vary, the selected solution does not vary either, which leads to repeated output of the same value, when the process is observed over time. This stereotypy may or may not correspond to the observed process; in the study of natural cognitive system, however, absolute repeatability of behavior is not the norm.

The alternative decision model is to draw, at each decision time, a value according to its probability. Values drawn are not the "best" solutions, and repeated decisions yield different values. Repeating this process allows the observer to reconstruct the knowledge that the drawn values originated from. Indeed, random sampling, when observed over time, is the only strategy guaranteed to reflect the original probability distribution.

Decision process modeling is a wide area of research; this short analysis does not come close to making it justice. Except in a recent research project that was specifically investigating this topic (see Section 4.4), we must admit that the choice of decision process, for our models, is usually driven by coarse-grained considerations. Indeed, we consider a drawing policy to be more biologically plausible, because it trades stereotypy, a rare trait in cognitive systems, with variability.

Random sampling also appears to be more satisfying from a methodological standpoint. Indeed, replacing a probability distribution by a stereotyped value is equivalent to replacing a probability distribution by a Dirac probability distribution. This operation decreases the uncertainty, and thus, the entropy, of the knowledge representation. From this perspective, random sampling appears as a process that correctly reflects the information available in the model, whereas probability maximization mathematically adds information to the model. If this

[1] We set aside Bayesian Decision Theory on purpose here. Indeed, in Bayesian Decision Theory, a probabilistic model is supposed to be combined with a deterministic function (loss function, reward function) assigning values to states or actions. This contrasts with our approach, where every bit of available knowledge is translated in probabilistic terms. A further discussion of this is to be found in another research project (see Section 4.4).

added information makes sense, then the modeler should consider making it explicit in the model; otherwise, this added information is unwarranted.

For these reasons, we usually consider random sampling decision processes in our models. However, because it is seldom the focus of our research questions, or even a necessary component in our models, we sometimes dismiss this choice altogether, and content ourselves with computing probability distributions.

2.3 Tools for structured Bayesian Programming

With the Bayesian Programming methodology as described so far, one can build models in a probabilistic form, and manipulate them using Bayesian inference. We claimed previously that Bayesian Programming would allow building highly modular, hierarchical Bayesian models, as in classical structured programming. To do so, two tools are required, which we describe now: the first is recursive Bayesian Program calls, the second is coherence variables.

2.3.1 Recursive Bayesian Program call

When we presented how parametric forms could be assigned to terms of the joint probability distribution decomposition, we intentionally left out one possibility (Section 2.2.3), for pedagogical purpose.

Indeed, instead of directly defining a probabilistic term of Bayesian Program $\langle \pi, \delta \rangle$ by using a mathematical form, the modeler can define it by asking a question to another Bayesian Program $\langle \pi_2, \delta_2 \rangle$. Consider for example the last term of Eq. (2.6), and assume that another Bayesian Program provides relevant information about X_2 and X_4; then one can write:

$$P(X_4 \mid X_2 \ \pi \ \delta) \stackrel{\text{def}}{=} P(X_4 \mid X_2 \ \pi_2 \ \delta_2) \ . \tag{2.9}$$

This closely mirrors subroutine calls in structured programming (and since the *Searched* set of probabilistic questions cannot be empty, function calls to be more precise). For instance, in the above example, X_2 can be thought of as an input variable to the function, X_4 would be an output variable, there could be an arbitrarily complex computation process encapsulated away (possibly involving internal variables, other than X_4 and X_2). Also, the same pitfalls exist: for instance, when $P(X_4 \mid X_2 \ \pi_2 \ \delta_2)$ itself uses a recursive call to $\langle \pi, \delta \rangle$ internally, it can either create inference loops that require adapted inference algorithms, or even non-technically sound models.

Subtleties about such recursive calls also mirror ones in classical programming. Note that probabilistic variables are formally "local" to Bayesian Programs that defines them; in other words, the X_4 variable of $\langle \pi, \delta \rangle$ is not the same as in $\langle \pi_2, \delta_2 \rangle$. Their names could be different, as long as they share their domains, then they can be linked by recursive calls. This is similar to the notion of formal and actual parameters of subroutine calls [2].

Whether such subroutine proceeds by call by value, or by reference, and whether side effects are supported, are properties of the implementation of the Bayesian inference engine, not of the mathematical framework. To the best of our knowledge, a thorough formal analysis of Bayesian Programming as a declarative programming language is yet to be developed [3].

[2] As in deterministic programming, having formal and actual parameters with the same name appears technically as bad form in Bayesian Programming; however, the humble diffusion of Bayesian Programming in the community has limited the need for enforcing good form. Let us wait for Bayesian Programming practitioners first, and then worry about turning them into good practitioners.

[3] Although, a very recent paper by De Raedt and Kimmig (2015) may be a first step in this direction, for the more general case of probabilistic programming languages.

At this point however, we note that recursive calls are not the only classical control structure of which we have probabilistic analogues: for instance, probabilistic conditional switches can be implemented using mixture models, probabilistic temporal loops using Bayesian filters (see Colas et al. (2010) for an introduction).

Of course, the major difference between recursive calls in the deterministic and probabilistic case is that, instead of a single value, the output of a subroutine call is a probability distribution. Unfortunately, there is an asymmetry in the mathematical notation, which prevents easily feeding a subroutine with a probability distribution as input (*a.k.a.*, soft evidence). Indeed, in a probabilistic question, the *Known* variable refers to a set of observed values; one cannot ask a question of the form $P(\mathit{Searched} \mid P(\mathit{Known})\ \pi\ \delta)$. To do just that, the modeler needs to resort to another tool than recursive calls, instead structuring the global model using coherence variables.

2.3.2 Coherence variables

Reasoning with soft evidence is one of the entry points into coherence variables. Another is to consider them as Bayesian switches (Gilet et al.; 2011) or a tool for behavior fusion programming (Pradalier et al.; 2003). Instead of providing a general treatment of coherence variables, which can be found elsewhere (Bessière et al.; 2013), we illustrate them on an example around a term $P(B \mid A)$, with two goals: the first is to handle soft evidence in Bayesian inference, so as to compute something with the semantics of $P(B \mid P(A))$, and the second is to have a tool to control whether a subpart of the model is connected or not to the rest.

We start with the technical definition of a coherence variable. In a Bayesian Program, a probabilistic variable Λ [4] is said to be a coherence variable when:

- its domain is Boolean (noted $\{0, 1\}$ here),
- it appears in the decomposition of the joint probability distribution in a term of the form $P(\Lambda \mid X\ X')$ (i.e., Λ is alone on the left-hand side, and two or more variables are on the right-hand side),
- the term $P(\Lambda \mid X\ X')$ is defined using a Dirac distribution [5], with value 1 if and only if some relation over right-hand sides variables holds (although generalizations of this exist). In most cases, an equality relation, such as $X = X'$ is considered: $P([\Lambda = 1] \mid X\ X') = \delta_{X=X'}(\Lambda)$.

We now introduce our model example, structured around a $P(B \mid A)$ term and a coherence variable connected to A. Using coherence variables yields variable duplication; in our example, variable A is duplicated. The joint probability distribution is $P(B\ A\ \Lambda\ A')$, decomposed as:

$$P(B\ A\ \Lambda\ A') = P(B \mid A)P(A)P(\Lambda \mid A\ A')P(A')\ . \tag{2.10}$$

We further assume that $P(A)$ is defined as a Uniform probability distribution, whereas $P(A')$ is not. The precise forms of $P(A')$ and $P(B \mid A)$ are not important for this example.

[4] Although we try to follow a convention where variables are capitalized and values are not, this is often not followed strictly in practice and in some of our papers. This is especially the case for variables with Greek symbols, such as coherence variables which unfortunately, often appear as λ.

[5] To avoid notational confusion, and because Dirac distributions are noted with a δ symbol, we silently drop the $\langle \pi, \delta \rangle$ symbols of probability terms of this section, which are unambiguous anyway.

We first show how Bayesian inference with soft evidence can be achieved. To do so, we consider the probabilistic question $P(B \mid [\Lambda = 1])$. Bayesian inference yields:

$$P(B \mid [\Lambda = 1]) \propto \sum_{A,A'} P(B \mid A)P(A)P(\Lambda \mid A\ A')P(A') \ .$$

The double summation is simplified because the coherence term is 0, unless $A = A'$, so that:

$$P(B \mid [\Lambda = 1]) \propto \sum_{A'} P(B \mid [A = A'])P(A') \ .$$

This has the desired semantics: whatever the distribution $P(A')$, when computing $P(B \mid [\Lambda = 1])$, the whole probability distribution $P(A')$ influences the $P(B \mid A)$ term, without having to consider a particular value for variable A'. This is reasoning with soft evidence.

The above computation can also be interpreted as having closed the Bayesian switch made of coherence variable Λ: by setting $[\Lambda = 1]$, the whole submodel about variable A' was connected to the model $P(B \mid A)$. Let us now verify that the Bayesian switch can be set in the "open" position. This is implemented by simply not specifying a value for Λ:

$$\begin{aligned}
P(B) &\propto \sum_{A,A',\Lambda} P(B \mid A)P(A)P(\Lambda \mid A\ A')P(A') \\
&\propto \sum_{A} P(B \mid A)P(A) \sum_{\Lambda} P(\Lambda \mid A\ A') \sum_{A'} P(A') \\
&\propto \sum_{A} P(B \mid A)P(A) \\
P(B) &\propto \sum_{A} P(B \mid A) \ .
\end{aligned}$$

In this inference, whatever the probability distribution $P(A')$, it vanishes because the coherence term can be simplified, as the summation over Λ yields a factor of 1. In that sense, from the point of view of variable B, submodel $P(A')$ was disconnected.

Note that, in the above example, we assumed $P(A)$ to be a Uniform probability distribution, so that it vanished from mathematical derivations. However, this is not necessary for the functioning of coherence variable, either in the soft evidence or in the Bayesian switch interpretation. In the general case, $P(A)$ can be any probability distribution, which need not be identical to $P(A')$. Therefore, a final use of coherence variables is to have two probability distributions about the same variable co-exist in a single model. This is not possible when defining a model by a direct decomposition of its joint probability distribution, as applying the product rule forbids any variable to appear more than once on the left-hand side of probability terms.

In other words, using coherence variables is a manner to "bypass" the product rule, and flatten out models; instead of having hierarchical constructs and recursive calls, a model can be re-written to articulate pieces of knowledge in an arbitrary manner. This is a powerful tool to express structured models, that will be used extensively in Chapters 3 and 4. It trades power of expression with the safety net provided by the constraint of the product rule; as such, it should be used with caution.

2.4 Relation with other probabilistic frameworks

We have already analyzed the relationship between Bayesian Programming and other, more widespread probabilistic frameworks elsewhere (Diard; 2003, Diard et al.; 2003b), proposing a

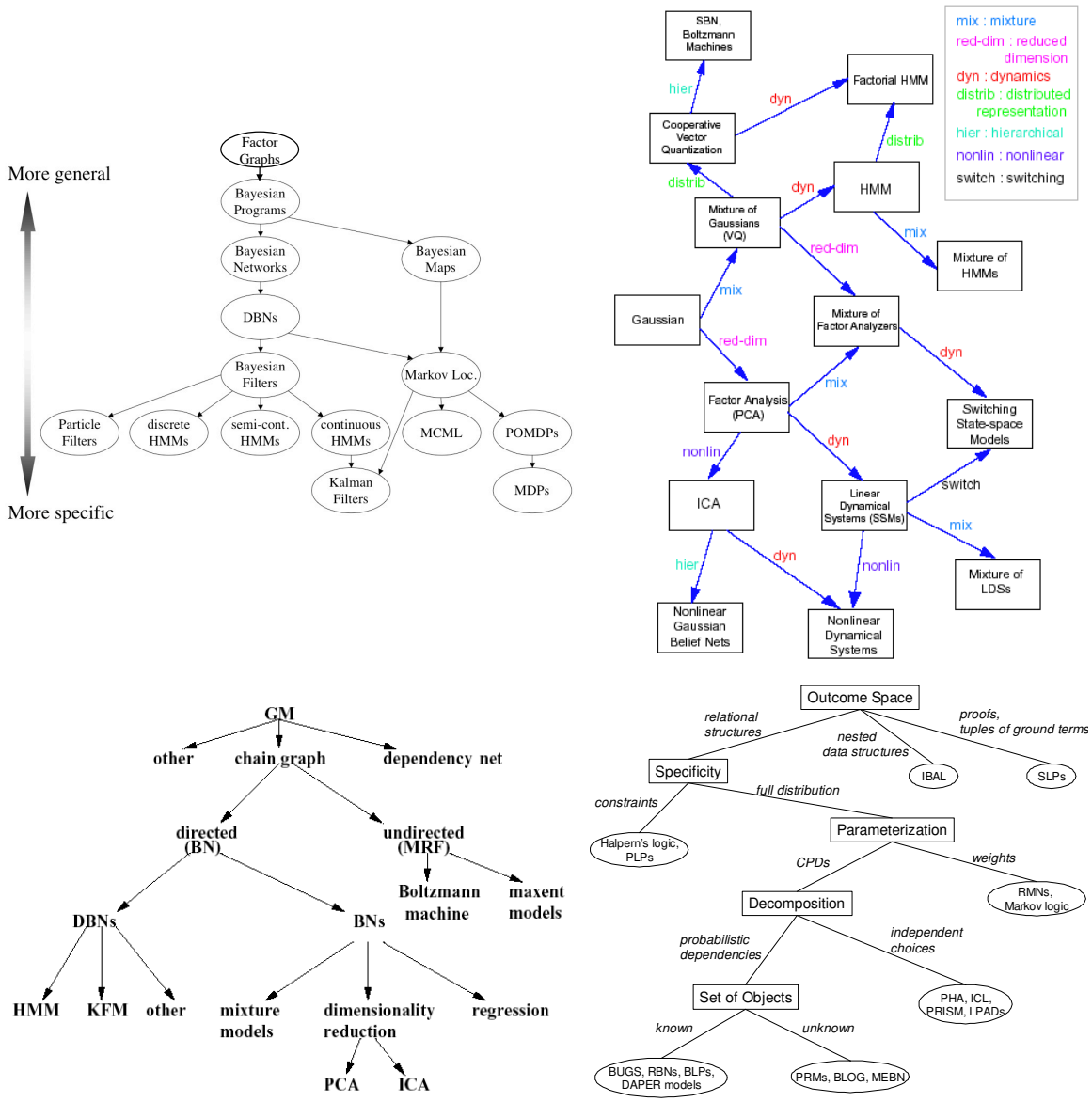

Figure 2.2: **Taxonomies of probabilistic frameworks.** Probabilistic frameworks are partially ordered using various ordering criteria. Usual acronyms are used; of note for the current manuscript are: BN (Bayesian Networks), DBN (Dynamic Bayesian Networks), HMM (Hidden Markov Models). **Top, left:** Adapted from Diard (2003), Diard et al. (2003b). **Top, right:** linear-Gaussian models and their specializations, taken from Roweis and Ghahramani (1999). **Bottom, left:** Taken from Murphy (2002). **Bottom, right:** first-order probabilistic languages, taken from Milch and Russell (2007).

taxonomy that is consistent with and complementary to others in the literature (Roweis and Ghahramani; 1999, Murphy; 2002, Milch and Russell; 2007); they are shown Figure 2.2. Let us recall some elements of this analysis here, aimed for the reader that has some familiarity with the usual technical definitions and vocabulary of the domain; other readers can safely skip or skim this section on their way to the next Chapter.

As we defined it, Bayesian Programming can be regarded as the most general framework for defining probabilistic models that are consistent with the product rule (Diard et al.; 2003b). Note that when we considered probabilistic variables and logical operators to combine then, we did not

consider first-order logical operators (such as $\forall$ and $\exists$), which do not have much interest when the probabilistic variable domains are discrete and finite, as is common in our practice. First-order probabilistic logic and relational statistical learning are exciting research domains (Koller and Pfeffer; 1997a, Friedman et al.; 1999, Milch and Russell; 2007, Kersting and De Raedt; 2007), leading to object-oriented variants of the programming paradigm we use, like object-oriented Bayesian networks (Koller and Pfeffer; 1997b). We humbly acknowledge that we have no clue as to what using such frameworks would bring to the matter of algorithmic cognitive models; we limit ourselves to "propositional-based" probabilistic frameworks.

In this context, the two closest neighbors of Bayesian Programming are the well-known Bayesian Networks, which are more specific than Bayesian Programming, and probabilistic factor graphs, which are more general. The general-to-specific measure we refer to here considers the set of all models that can be written using each formalism.

Firstly, Bayesian Programming being more general than Bayesian Networks means that there are some probabilistic models that are consistent with the product rule that cannot be expressed as a Bayesian Network. Indeed, one such model corresponds to our previous example, in Eq. (2.6), that we recall here:

$$P(X_1\ X_2\ X_3\ X_4 \mid \pi\ \delta) = P(X_1\ X_2 \mid \pi\ \delta)P(X_3 \mid X_1\ \pi\ \delta)P(X_4 \mid X_2\ \pi\ \delta)\ .$$

A Bayesian Network is a model of the form $P(X_1\ X_2\ \ldots\ X_n) = \prod P(X_i \mid \mathcal{P}a(X_i))$, with $\mathcal{P}a(X_i)$ the set of "parent" variables of X_i, i.e., a subset of $\{X_1, \ldots, X_{i-1}\}$. Our counterexample of Eq. (2.6) does not fit this constraint, as variable $X_1 \wedge X_2$ was to be considered as a single, multidimensional variable for the first term, but the term over X_3 only had a portion of variable $X_1 \wedge X_2$ as parent.

Secondly, Bayesian Programming being less general than probabilistic factor graphs means that there are some probabilistic models that can be written with probabilistic factor graphs that are not consistent with the product rule. Indeed, recall that factor graphs are models of the form $g(X_1\ X_2\ \ldots\ X_n) = \prod f_i(S_i)$, with S_i any subset of $\{X_1, \ldots, X_n\}$, and that a probabilistic factor graph is a factor graph where functions $g(\cdot)$ and $f_i(\cdot)$ are probability distributions. Note that, with this definition, any product of probability distributions constitute a probabilistic factor graph, independently of whether it corresponds to a valid decomposition of the joint probability distribution using the product rule.

We must now discuss the above analyses, and acknowledge that they can be disputed. Indeed, concerning the Bayesian Network counterexample, one could argue that there is a way to rephrase Eq. (2.6) in order to make it fit the Bayesian Network definition, without changing the model. In this case, terms $P(X_3 \mid X_1)$ and $P(X_4 \mid X_2)$ can be rewritten respectively as $P(X_3 \mid X_1\ X_2)$ and $P(X_4 \mid X_1\ X_2)$, and the conditional independence hypotheses can be re-introduced implicitly later, by having probability terms that, in effect, do not depend on the added variable (in the discrete case, that would just cost some memory space and computation time).

In a similar manner, one could argue that coherence variables, which have been shown to be able to "bypass" the product rule constraint when defining Bayesian Program, in effect allow the programmer to write any probabilistic factor graph in the form of a Bayesian Program.

Such arguments effectively would conflate Bayesian Programming with other probabilistic formalisms, that we presented as strictly distinct neighbors. However, arguments of this nature would also allow conflating every formalism with every other one; for instance, any static model is actually a Dynamic Bayesian Network with a single time-slice, any Factorial HMMs can be cast as a vanilla HMM with a single mega-state variable (Murphy; 2002), etc. From that point

of view, probabilistic frameworks are not strict mathematically well-defined sets of probabilistic models, they also integrate useful tools for interpreting and representing these models.

2.5 Graphical representation of Bayesian Programs

Such a useful tool for interpreting probabilistic models is, of course, graphical visualization of models. Even though the "real" definition of a Bayesian program is provided in algebraic terms, we recognize the pedagogical importance of accompanying this definition by a graph; communicating and discussing over graphs is usually much easier than over equations.

In our practice, we have adopted the graphical representation of Bayesian Networks, even though, in some cases, the graph does not represent exactly the model (when it is not in the class of Bayesian Network models). In the following chapters, we will present models we designed using both notations. Remember that the "ground truth" notation is the mathematical expression, and the graphs only a pedagogical tool.

Also, graphs sometimes lead to intuitive interpretations which, unfortunately are misleading. For instance, in graphical representations of Bayesian Programs, it is tempting, but wrong, to interpret nodes (variables) as semantically constrained pieces of models. Variables merely define spaces; it is the probability distributions they appear in, in the context of probabilistic inferences, that constrain their semantics.

For instance, consider variable P, representing percepts in a perception model [6] of the form $P(P\,S) = P(S \mid P)P(P)$. P can be interpreted as the result of a perception process in the question $P(P \mid S)$, or to the starting point of a sensory prediction process in the question $P(S \mid P)$. These two processes may have different neurobiological correlates, and so the P variable would actually be a model of different neural areas, depending on the question asked to the model. This goes contrary to the "localist" interpretation of graphs, where the structure sometimes resembles models of neural pathways. Instead, the Bayesian Program above models information representation and manipulation: it would yield the experimentally testable prediction that both processes rely on the same knowledge. The clear-cut separation between knowledge representation and knowledge manipulation, which is a specificity of Bayesian Programming, is somewhat lost when using only a graphical representation of the joint probability distribution. Recall that, in Bayesian Programming, we never model processes directly, but we model knowledge that yields processes.

[6] Many reviewers have already mentioned to us they found the use of a variable denoted P troublesome, because it could be confused with function $P(\cdot)$, omnipresent in a probabilistic framework. With glee, we remark that, if a compiler is able to distinguish overloaded symbols by counting their arity, so should they.

CHAPTER 3

Bayesian modeling of reading and writing

In this and the next Chapter, we describe Bayesian models of cognitive functions that we have studied over the last few years. For each, we focus our presentation on the model itself; the interested reader will find discussions about the scientific context of each topic in the original publications, referred to in introductory captions. To present each contribution, we first provide the main model, first by defining its joint probability distribution, second by describing the method for parameter identification, third and finally by showing how it solves cognitive tasks using Bayesian inference. We then define a variant model, which is compared to the main model in order to answer a scientific question of interest.

3.1 Bayesian modeling of letter reading and writing: BAP

Biographical note

Collaborators: Pierre Bessière (ISIR, Paris), Richard Palluel-Germain (LPNC)

Supervised students: Estelle Gilet (Ph.D., LIG, Grenoble, defended in 2009), Maxime Frecon (Master 1)

Publications: Gilet et al. (2011, 2010), Gilet (2009), Gilet et al. (2008b,a)

This section adapts material from Gilet et al. (2011)

We have been interested in the cognitive processes involved in perception and action, and, more precisely, in the tasks of reading and writing, which we have decided to study jointly. The most common approach, in the study of this action–perception loop, is to consider the influence that the prediction of future perceptions has on the current choice of action; we have instead focused on modeling the influence of motor knowledge on perception. To capture this influence, we have developed the Bayesian Action–Perception (BAP) model, whose main feature is an internal motor simulation loop, which may be recruited in perception tasks.

We restricted ourselves to the case of isolated letters to limit lexical, semantic and other top-down effects related to the global perception of words. Furthermore, we treated the case of online recognition, where the presented trajectories contain both spatial and sequence information. In other words, we considered perception tasks where the letter is perceived as it is being traced.

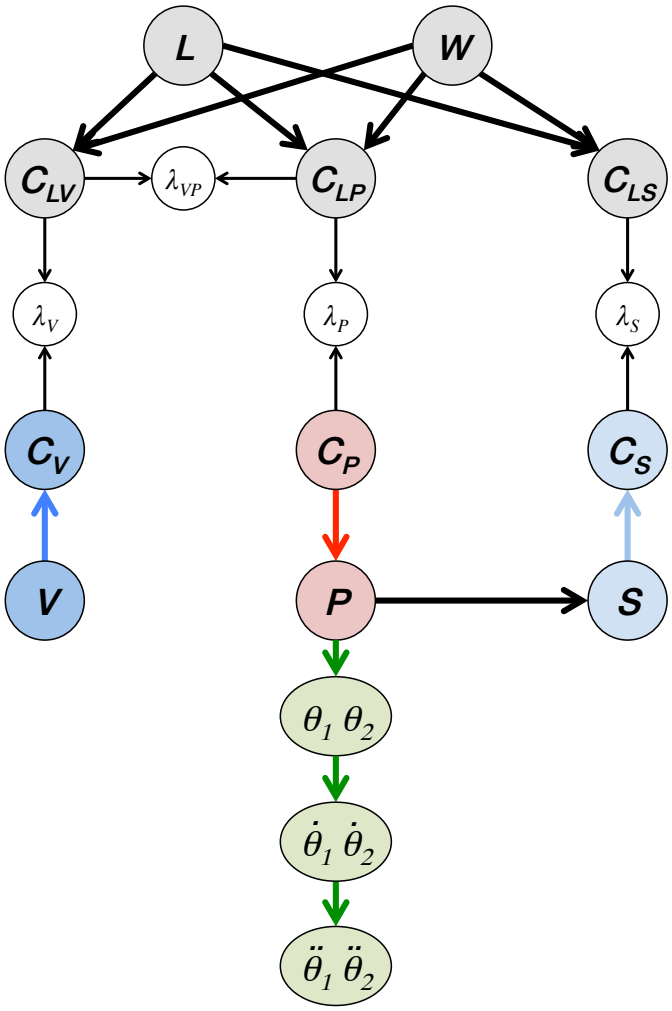

Figure 3.1: **Graphical representation of the structure of the BAP model.** The top part connects the letter L and writer W variables to via-point sequences C_{LV}, C_{LP} and C_{LS} (in black). The bottom part contains visual knowledge from visual input V to via-point sequence C_V (in blue), simulated visual knowledge from internal visual representation S to via-point sequence C_S (in light blue), and motor knowledge, made of two components: movement planning from via-point sequence C_P to complete trajectory P (in red) and effector dependent trajectory generation, from trajectory P to articulatory variables θ_1, θ_2 and their derivatives (in green). These components are linked using various λ coherence variables, acting as Bayesian switches (in light gray). Each C_x variable is 64 dimensional, P is $2(T+1)$ dimensional, articulatory variables are $T+1$ dimensional, from time step 0 to T.

3.1.1 BAP model definition

The BAP model was defined by the following decomposition of the joint probability distribution:

$$
\begin{aligned}
&P(L\ W\ C_{LV}\ C_V\ C_{LP}\ C_P\ C_{LS}\ C_S\ V\ P\ E\ S\ \lambda_L\ \lambda_V\ \lambda_P\ \lambda_S) \\
&= \left(\begin{array}{l}
P(L)P(W) \\
P(C_{LV} \mid L\ W)P(C_{LP} \mid L\ W)P(C_{LS} \mid L\ W) \\
P(V)P(C_V \mid V) \\
P(P \mid C_P)P(C_P)P(E \mid P) \\
P(S \mid P)P(C_S \mid S) \\
P(\lambda_V \mid C_{LV}\ C_V)P(\lambda_P \mid C_{LP}\ C_P)P(\lambda_S \mid C_{LS}\ C_S)P(\lambda_L \mid C_{LV}\ C_{LP})
\end{array} \right) . \qquad (3.1)
\end{aligned}
$$

A graphical network representing this decomposition is shown Figure 3.1. This is a simplified version of the whole joint probability distribution, as most variables it features are actually multi-dimensional. We first describe the main architecture of the model, then traverse it, from top to bottom, introducing more precisely the variables, the terms of the decomposition and their interpretation, and the associated parametrical forms, when necessary.

Four hypotheses form the basis of the architecture of the BAP model. First, there are several distinct internal representations of letters: one for the perception model, one for the action model, and one for the simulated perception model. Second, these representations are of the same nature (Cartesian space) and are based on the same encoding. Third, this encoding consists in

summarizing letter trajectories by sequences of via-points, which lie at points where the tangent is either vertical or horizontal, and at cusps. Finally, a feedback loop from the generated trajectories back to the internal representation of letters implements an internal simulation of movements.

The most abstract variables in the BAP model are L and W, which respectively represent letter and writer identity. Both are defined as discrete, finite variables, that is to say, they are simply sets of values, without any ordering between them.

We note the representations of letters as C_{LV} for the (visual) perceptual representation, C_{LP} for the (production) motor representation, and C_{LS} for the simulated perception representation. Because there are several pieces of knowledge that would require having such variables on left-hand sides of probability terms, they have been duplicated (adding C_V, C_P and C_S) and connected with coherence variables (λ_V, λ_P and λ_S). A fourth coherence variable, λ_L, connects C_{LV} and C_{LP}.

Each letter representation variable is a multi-dimensional and temporal series of variables. We chose a representation model that could be presumed to be relevant for both recognition and production processes; we assumed that letters are represented by a sequence of via-points, placed where either the x derivative ($\dot{x}$) or the y derivative ($\dot{y}$), or both, is zero. This therefore assumes that letters are encoded in a Cartesian, x, y space, isomorphic to the workspace. Via-points are four-dimensional, collecting position C_{Lx}, C_{Ly} and velocity information $C_{L\dot{x}}$, $C_{L\dot{y}}$. We denote the set of via-points for a given trajectory as $C_L^{0:N}$, with N, the maximum number of via-points, set to 16, which is quite sufficient for all trajectories we considered.

As terms of the form $P(C_{Lx}^{0:N}\ C_{Ly}^{0:N}\ C_{L\dot{x}}^{0:N}\ C_{L\dot{y}}^{0:N} \mid L\ W)$ have high dimensionality (64 dimensions for $N = 15$), we use conditional independence hypotheses to decompose them into a product of smaller distributions. The joint probability distributions over such sets of variables are defined as:

$$
\begin{aligned}
&P(C_L^{0:N} \mid L\ W) \\
&= \left(\begin{array}{l}
\left(\begin{array}{ll}
P(C_{Lx}^0 \mid L\ W) & P(C_{Ly}^0 \mid L\ W) \\
P(C_{L\dot{x}}^0 \mid L\ W) & P(C_{L\dot{y}}^0 \mid L\ W)
\end{array} \right) \\
\prod_{n=1}^{N} \left(\begin{array}{ll}
P(C_{Lx}^n \mid C_{Lx}^{n-1}\ L\ W) & P(C_{Ly}^n \mid C_{Ly}^{n-1}\ L\ W) \\
P(C_{L\dot{x}}^n \mid C_{L\dot{x}}^{n-1}\ L\ W) & P(C_{L\dot{y}}^n \mid C_{L\dot{y}}^{n-1}\ L\ W)
\end{array} \right)
\end{array} \right) .
\end{aligned}
\tag{3.2}
$$

In other words, in the BAP model, letters are represented as sequences of via-points, with a first order temporal Markov assumption (we assume that the positions or velocities of a via-point (index n) depend only on the positions or velocities of the previous via-point (index $n-1$)) and a naïve Bayes fusion model for the four dimensions of position and velocity information. Each of the terms of the form $P(C_{Lx}^n \mid C_{Lx}^{n-1}\ L\ W)$ are conditional probability tables (CPT), whose parameters are identified from an experimental database provided to the model during a learning phase.

The top-part of the model (Figure 3.1) consists in encoding, for each letter L and writer W, such a probabilistic model. This probabilistic database is triplicated, and each is connected to a perception, production, and simulated perception model [1]. These models connect the via-

[1] This multiplicity of representations is for syntactic purposes of the construction of the Bayesian Program; it is the classical solution for correctly defining Bayesian models containing loops. It is certainly not an assumption about possible neural correlates. Recall that we build algorithmic models, coarsely inspired from large-scale knowledge of perception and action processes in the brain, but without strong claims about neurobiological implementation models. For instance, our multiple representations could model activities, at different time steps, of a single neurobiological area.

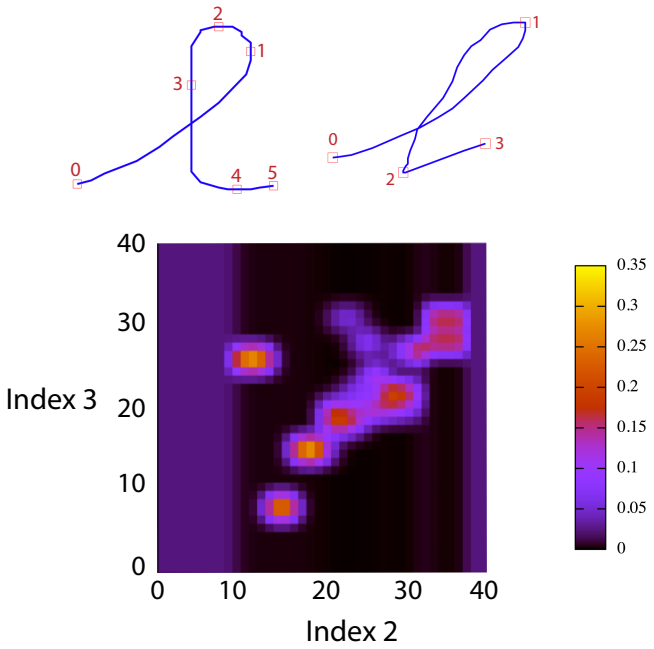

Figure 3.2: **Example of sample traces and the learned probability distributions in the letter representation model. Top:** Two allographs of the letter l written by writer Julienne. **Bottom:** Probability distributions of the abscissa of the third via-point of the letter (l) from the writer Julienne, as a function of the abscissa of the second via-point: $P(C^3_{Lx} \mid C^2_{Lx}\ [L = l]\ [W = \text{Julienne}])$. Each column is a probability distribution and sums to 1.

point based representations to complete trajectories in the Cartesian workspace, sampled with a fine-grained temporal resolution.

For instance, the perception model connects the visual input which is a complete trajectory of x, y coordinates, $V = V_X^{0:M} \wedge V_Y^{0:M}$, with M the length of the input trajectory, to the via-point sequence C_V in the term $P(C_V \mid V)$. The perception and simulated perception models are identical, and simply consist in a deterministic algorithm f for via-point extraction from a complete trajectory, wrapped in a probabilistic setting (using a functional Dirac model).

The action model is decomposed into two sub-models: the trajectory generation (or planning) model and the effector model. Trajectory generation is, like the perception models, a deterministic algorithm g wrapped in the probability distribution $P(P \mid C_P)$, and produces a complete trajectory $P = P_X^{0:T} \wedge P_Y^{0:T}$ from a set of via-points C_P, using a classical acceleration minimization constraint, which has a polynomial solution.

The effector model translates a trajectory $P = P_X^{0:T} \wedge P_Y^{0:T}$ into effector space E, with the probability distribution $P(E \mid P)$. In our main experiments, we simulated a two-degree of freedom articulated arm, with $E = \theta_1 \wedge \theta_2$ representing shoulder and elbow angles (with their velocity and acceleration derivatives noted $\dot{\theta}_1, \dot{\theta}_2, \ddot{\theta}_1, \ddot{\theta}_2$). The effector model is composed of three terms: first, $P(\theta_1^{0:T}\ \theta_2^{0:T} \mid P_X^{0:T}\ P_Y^{0:T})$ the inverse kinematic model, along with two terms $P(\dot{\theta}_1^{0:T}\ \dot{\theta}_2^{0:T} \mid \theta_1^{0:T}\ \theta_2^{0:T})$, $P(\ddot{\theta}_1^{0:T}\ \ddot{\theta}_2^{0:T} \mid \dot{\theta}_1^{0:T}\ \dot{\theta}_2^{0:T})$ for computing successive derivatives using a finite difference method. Overall, this system can be seen as encapsulating a deterministic algorithm h in a probabilistic manner.

3.1.2 Parameter identification in BAP

With the decomposition of the joint probability distribution of Eq. (3.1) above, the BAP model is almost fully specified. The only free parameters concern the terms of the form $P(C_L \mid L\ W)$, which are the probabilistic models of letters, as a function of letter and writer identity.

Given a small experimental database of 40 labeled sample traces, we identified the parameters of the associated CPTs in a straightforward manner. We illustrate this by showing both sample

traces and learned probability tables, Figure 3.2.

Letters have many possible forms – called allographs – because of fluctuations in handwriting (see Figure 3.2, top). The representation of letters is robust to this within-writer variability. Indeed, the learned CPTs incorporate it, and implicitly encode several allographs in one distribution. For instance, Figure 3.2 (bottom) presents the probability distribution of the third via-point of a letter, given the position of the second via-point. The two allographs of Figure 3.2 (top) respectively correspond to the series of peaks below the diagonal (the third via-point is to the left of the second via-point, as in the upward l), and the main peak above the diagonal (the third via-point is to the right of the second via-point, as in the slanted l).

3.1.3 Inference in BAP

We have shown that the BAP model could solve a wide variety of cognitive tasks related to reading and writing. We simulated six cognitive tasks: i) letter recognition (purely sensory), ii) writer recognition, iii) letter production (with different effectors), iv) copying of trajectories, v) copying of letters, and vi) letter recognition (with internal simulation of movements).

Perception tasks: Letter and writer recognition

The cognitive task of letter recognition consists of identifying the presented letter. In other words, the question is: "given a trajectory produced by a known writer, what is the letter?" In probabilistic terms, this corresponds to computing the probabilistic question:

$$P(L \mid [V_x^{0:M} = v_x^{0:M}]\ [V_y^{0:M} = v_y^{0:M}]\ [W = w]\ [\lambda_V = 1])\ , \tag{3.3}$$

where $v_x^{0:M}, v_y^{0:M}$ is the input trajectory, w is the given writer, and $\lambda_V = 1$ activates only the perception and letter representation parts of our model. Bayesian inference yields:

$$\begin{aligned} &P(L \mid [V_x^{0:M} = v_x^{0:M}]\ [V_y^{0:M} = v_y^{0:M}]\ [W = w]\ [\lambda_V = 1]) \\ &\propto\ \ P([C_{LV} = f(v_x^{0:M}, v_y^{0:M})] \mid L\ [W = w])\ . \end{aligned} \tag{3.4}$$

To evaluate the performance of letter recognition, we split our database of trajectories into a training set and a testing set, using 35 samples for training and 5 samples for testing. Training consisted of parameter identification, as previously described, and testing consisted of computing the probability distribution over letters L and using this distribution to draw randomly a value for L. This selected value, the answer to the recognition task, was then used to assess whether the model had succeeded in recognizing the presented letter.

We repeated this procedure, varying the samples that were used for training and testing, applying classical K-fold cross-validation (Russell and Norvig; 1995). We obtained full confusion matrices, from which averaging correct values yielded a satisfying recognition rate of 93.36%. Some misclassifications arise because of the geometric similarities of some letters, and the small size of our learning database. As performance is not our focus here, we used this experiment merely to validate that the model could yield letter recognition [2].

[2]For another project, where the BAP model was adapted to the context of eye writing, with a focus on applicative purposes, refer to Section A.4.

A more complicated variant of letter recognition is obtained when writer identity w is not provided as input:

$$\begin{aligned} & P(L \mid [V_x^{0:M} = v_x^{0:M}]\ [V_y^{0:M} = v_y^{0:M}]\ [\lambda_V = 1]) \\ & \propto \sum_{w \in W} P([C_{LV} = f(v_x^{0:M}, v_y^{0:M})] \mid L\ [W = w])\ . \end{aligned} \tag{3.5}$$

In this case, we still observed a high accuracy rate of 92.72%. An even more difficult case is to test letter recognition using the model with a new writer, by using testing trajectories provided by a writer who was not used in the training trajectories. In this case, the correct recognition rate drops to 49.68%, still well above chance level.

In a symmetric manner, writer recognition, based on some input trajectory, and with or without providing letter identity as input, is performed by computing:

$$\begin{aligned} & P(W \mid [V_x^{0:M} = v_x^{0:M}]\ [V_y^{0:M} = v_y^{0:M}]\ [L = l]\ [\lambda_V = 1]) \\ & \propto P([C_{LV} = f(v_x^{0:M}, v_y^{0:M})] \mid [L = l]\ W)\ , \end{aligned} \tag{3.6}$$

$$\begin{aligned} & P(W \mid [V_x^{0:M} = v_x^{0:M}]\ [V_y^{0:M} = v_y^{0:M}]\ [\lambda_V = 1]) \\ & \propto \sum_{l \in L} P([C_{LV} = f(v_x^{0:M}, v_y^{0:M})] \mid [L = l]\ W)\ . \end{aligned} \tag{3.7}$$

Production task: writing letters

Given a letter l to write, and a writer w to imitate, what are the accelerations required to trace the letter? This writing task is translated, mathematically, by computing:

$$\begin{aligned} & P(\ddot{\theta}_1^{0:T}\ \ddot{\theta}_2^{0:T} \mid [L = l]\ [W = w]\ [\lambda_P = 1]) \\ & \propto \sum_{C_{LP}} \begin{pmatrix} P(C_{LP} \mid [L = l]\ [W = w]) \\ P(\ddot{\theta}_1^{0:T}\ \ddot{\theta}_2^{0:T} \mid [\dot{\theta}_1^{0:T}\ \dot{\theta}_2^{0:T} = h(g(C_{LP}))]) \end{pmatrix}. \end{aligned} \tag{3.8}$$

Instead of explicitly computing the costly summation of Eq. 3.8, we drastically approximate it, in a Monte-Carlo inspired manner, which can be seen as a two-step algorithm. First, the model of letter representation is used to draw randomly positions and velocities of via-points according to $P(C_{LP} \mid [L = l]\ [W = w])$. Second, from the drawn via-points, the trajectory generation model is used to determine the complete trajectory between them, and the effector model finally translates the Cartesian coordinates of points in the trajectory to joint coordinates and accelerations to apply.

Writing with the BAP model yields between-writer and between-trial variabilities. We first illustrate between-writer variability, showing Figure 3.3 trajectories for *a*s generated using the writing styles of Estelle and Christophe, and the corresponding writing styles in the learning data.

If we ask the writing question several times to the model, we observe within-writer inter-trial variability; that is, the resulting trajectories are not identical (see Figure 3.4). Indeed, as the positions and the velocities at via-points are drawn according to a probability distribution, the obtained trajectories vary. This result is, of course, in agreement with the everyday observation that every time we write, we are not producing exactly the same trajectory.

Perception and production task: copying trajectories and letters

We now turn to a cognitive task that involves the representation of letters, and the perception and action branches of the model. It consists in copying input trajectories, that is, we provide

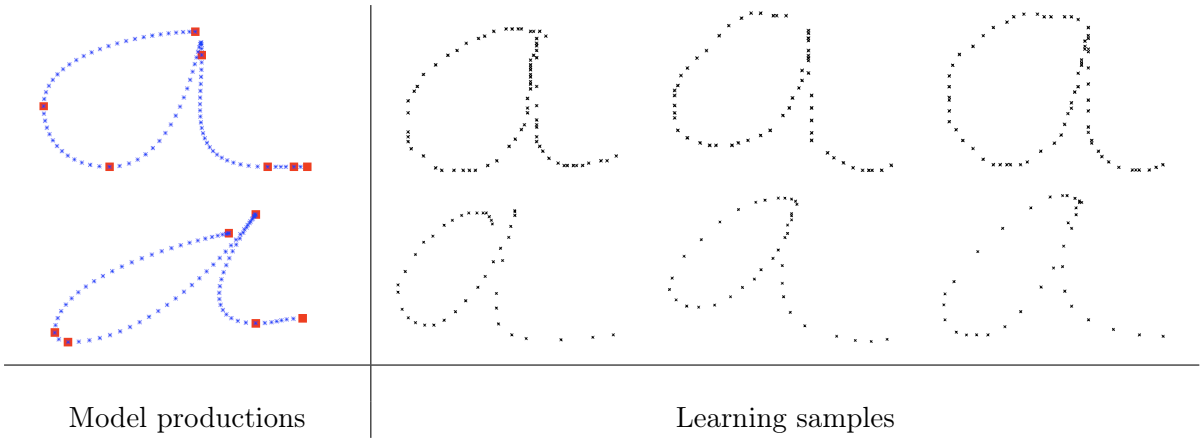

Figure 3.3: **Illustration of between-writer variability when writing *a*s. Left column:** Traces produced by the model when given $[W = \text{Estelle}]$ (top) or $[W = \text{Christophe}]$ (bottom) as a constraint. **Right column:** Sample trajectories of writers Estelle (top row) and Christophe (bottom row), in the learning database. Estelle's *a*s are more rounded, whereas Christophe's *a*s are more slanted; this is captured and reproduced by the model.

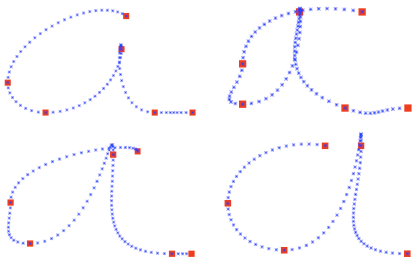

Figure 3.4: **Illustration of inter-trial variability when writing *a*s.** Four trajectories obtained by computing $P(\ddot{\theta}_1^{0:T}\ \ddot{\theta}_2^{0:T} \mid [L = a]\ [W = \text{Estelle}]\ [\lambda_P = 1])$.

an input trajectory and ask the model to compute the corresponding accelerations to apply to the simulated arm. This is translated mathematically and solved by Bayesian inference in the following manner:

$$P\left(\ddot{\theta}_1^{0:T}\ \ddot{\theta}_2^{0:T} \;\middle|\; \begin{array}{l}[V_x^{0:M} = v_x^{0:M}]\ [V_y^{0:M} = v_y^{0:M}]\\ [\lambda_V = 1]\ [\lambda_L = 1]\ [\lambda_P = 1])\end{array}\right) \tag{3.9}$$
$$\propto \left(\begin{array}{l} P(\ddot{\theta}_1^{0:T}\ \ddot{\theta}_2^{0:T} \mid \dot{\theta}_1^{0:T}\ \dot{\theta}_2^{0:T}\ \theta_1^{0:T}\ \theta_2^{0:T})\ P(\theta_1^{0:T}\ \theta_2^{0:T} \mid P)\\ P(P \mid [C_P = f(v_x^{0:M}, v_y^{0:M})])\end{array}\right).$$

Note that, in the inference, the internal representation of letter model was "bypassed", by setting $[\lambda_L = 1]$ in the probabilistic question. The via-points extracted from the input trajectory, by the term $P(P \mid [C_P = f(v_x^{0:M}, v_y^{0:M})])$ are directly fed into the production model. As a consequence, with this mathematical translation of the task, any type of trajectory can be copied, not only those for known letters.

Dropping the $[\lambda_L = 1]$ constraint reconnects the internal representation of letter model, so that it is not the input trajectory that is copied, but the letter recognized from the input trajectory: instead of trajectory copy, this performs letter copy. We report this in a slight variation of the model, where variable W, representing the writer, has been duplicated into W_V from W_P, to distinguish the writer that traced the input letter from the writer style used to trace the recognized letter. Letter copy corresponds to the following probabilistic question and

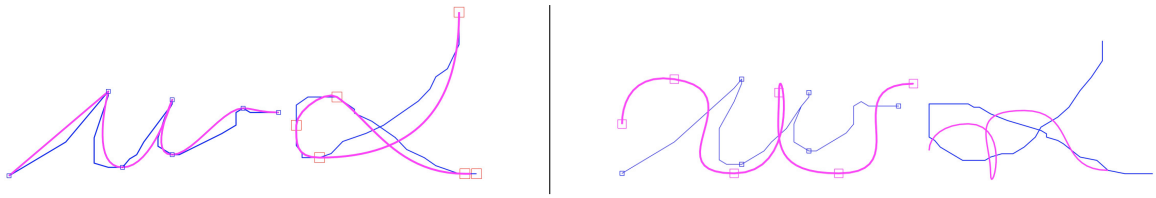

Figure 3.5: **Examples of trajectory and letter copying.** The input trajectories are in blue; the copied, output trajectories are in pink. **Left:** trajectory copy. The model can copy trajectories corresponding to known letters (e.g., w) and those corresponding to unknown symbols, outside of the learned repertoire (e.g., α). **Right:** letter copy. The graphical forms between input and output trajectories can be quite different, provided that the writing styles of the input and output writers are different; thus, giving different values to W_V and W_P allows performing forgery.

inference:

$$P\left(\ddot{\theta}_1^{0:T}\ \ddot{\theta}_2^{0:T} \;\middle|\; \begin{array}{l}[V_x^{0:M}=v_x^{0:M}]\ [V_y^{0:M}=v_y^{0:M}]\\ [\lambda_V=1]\ [\lambda_P=1]\ [W_P=w_p]\ [W_V=w_v]\end{array}\right)$$
$$\propto \left(\begin{array}{l} P(\ddot{\theta}_1^{0:T}\ \ddot{\theta}_2^{0:T} \mid \dot{\theta}_1^{0:T}\ \dot{\theta}_2^{0:T}\ \theta_1^{0:T}\ \theta_2^{0:T}) P(\theta_1^{0:T}\ \theta_2^{0:T} \mid P)\ P(P \mid C_{LP})\\ \sum_L \left(\begin{array}{l} P([C_{LV}=f(v_x^{0:M},v_y^{0:M})] \mid L\ [W_V=w_v])\\ P(C_{LP} \mid L\ [W_P=w_p])\end{array}\right)\end{array}\right). \tag{3.10}$$

Figure 3.5 shows examples of letter trajectory and letter copy.

Perception, internal production and simulated perception task: letter recognition

The final cognitive task that we investigated is letter recognition, but, contrary to the previous case where only the perception and representation of letters sub-models were involved, here, the entire BAP model is activated. In other words, this task can also be seen as an extension of trajectory copying, where, instead of being executed, the planned trajectory is fed to the internal simulation of movement loop. This task is translated and solved by Bayesian inference using (assuming the writer w given):

$$P\left(L \;\middle|\; \begin{array}{l}[V_x^{0:M}=v_x^{0:M}]\ [V_y^{0:M}=v_y^{0:M}]\ [W=w]\\ [\lambda_V=1]\ [\lambda_L=1]\ [\lambda_P=1]\ [\lambda_S=1]\end{array}\right)$$
$$\propto \left(\begin{array}{l} P([C_{LV}=f(v_x^{0:M},v_y^{0:M})] \mid L\ [W=w])\\ P([C_{LS}=h(g(C_{LV}))] \mid L\ [W=w]))\end{array}\right). \tag{3.11}$$

This is the product of two terms, the first of which, $P([C_{LV}=f(v_x^{0:M},v_y^{0:M})] \mid L\ [W=w])$, is exactly Eq. (3.4). In other words, this first term amounts to letter recognition in the reading task, where the motor and simulation parts of the model are not activated. The second term of Eq. (3.11) is $P([C_{LS}=h(g(C_{LV}))] \mid L\ [W=w]))$. This also corresponds to letter recognition but using via-points that are the result of a longer circuit inside the model. First, via-points are extracted from the input trajectory, and then these are forwarded to the trajectory generation motor model, which generates a complete simulated trajectory. This is then forwarded to the simulated perception branch of the model, which extracts from it another set of via-points. These via-points are then compared in the letter representation model with the memorized letter representations.

We experimentally tested the model under the same conditions as in the reading task using only the perception sub-model. We obtained an overall recognition rate of 90.22%.

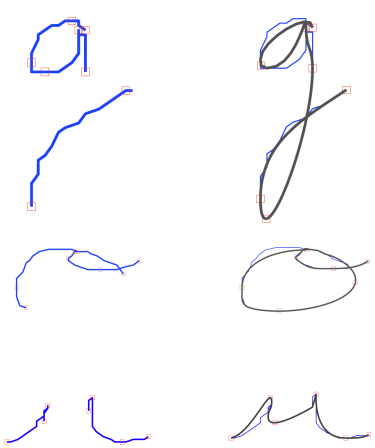

Figure 3.6: **Example cases where motor knowledge helps identify the correct letter. Left:** incomplete trajectories presented to the model (in blue) and corresponding via-points extracted by perception (in red). **Right:** the same trajectory superposed with the trajectory generated by internal motor simulation (in black) and resulting via-points extracted by simulated perception (in red).

3.1.4 Model variants and their comparison

The scientific issue we were interested in concerned the accumulated evidence of cortical motor activations observed during perception tasks, such as reading letters (Longcamp et al.; 2003, 2006). Indeed, the question of the functional role of these motor activations remain open: are they simply by-products, or do they participate in some manner in the perception processes? In the BAP model, we were able to explore this question mathematically, by comparing the letter recognition task, without (Eq. (3.4)) and with (Eq. (3.11)) activation of the motor and simulated perception knowledge.

A first result of this comparison is purely formal: in the BAP model, perception involving motor and simulated perception knowledge is the product of two terms, one of which is the purely perception task. It demonstrates that motor activations adds knowledge to the perception process; if motor and perceptual knowledge are not wildly inconsistent, combining them by a product of probability distributions will lower variance of the final estimate. In other words, adding knowledge usually should yield better perceptual estimations.

A second result comes from the quantitative comparison of performance scores in the two tasks. In our experiment, correct recognition scores were comparable (90 vs. 93%). Some differences appeared in the confusion matrices, with some errors being corrected by adding internal simulation of movements, and others appearing. What was surprising was that adding motor knowledge did not, in this case, increase performance.

A third result comes from the analysis of situations where motor knowledge would increase performance. We found out that the previous result was due to a "ceiling" effect, that is, stimuli were too easy to recognize, independently of whether motor knowledge was involved. We have therefore designed more difficult stimuli: instead of presenting complete trajectories as inputs, we designed truncated versions of trajectories where we erased a set of consecutive points. We have found several cases where reading without motor simulation would fail but reading with motor simulation would succeed. We illustrate a few of such cases in Figure 3.6. This suggest a possible role of motor knowledge in adverse situations: it would help recover missing information from the stimulus.

This last result yields an experimental prediction, whereby, in a perception task, motor knowledge would be more useful in adverse conditions as compared to nominal conditions. Assuming that involving motor knowledge necessitates motor cortical areas activations, our results suggest higher activations of motor cortical areas in adverse perceptual conditions. Alternatively, motor

activations might occur after behavioral decisions in nominal conditions, but would be required in adverse conditions to reach decision thresholds, and thus would precede decision. Testing such experimental predictions would require neuroimaging studies; another protocol would impair motor processes during perception tasks, for instance by overloading motor areas with a concurrent task. Using such an experimental manipulation, James and Gauthier (2009) found effects on performance that are consistent with our predictions.

3.1.5 Discussion

We discuss one feature of the BAP model as a Bayesian Algorithmic Cognitive Model. During presentation of the cognitive tasks and their resolution by Bayesian inference, we have routinely interpreted equations as ordered sequences of steps in the treatment of information, as in a pseudo-algorithm. But this is just for pedagogical purposes. For instance, consider Eq. (3.11), and its product of two terms. The commutativity of the product obviously forbids the conclusion that the first term is computed before or after the second term.

Moreover, just because the graphical representation of the BAP model (Figure 3.1) shows spatially distinct subparts of models, this does not mean that we would expect spatially distinct corresponding areas in the central nervous system (CNS). More precisely, although we require mathematically distinct perception and simulated perception branches in the model, it could be the case that, in the CNS, there is only one set of areas that deal with both perception and simulated perception, with possibly temporally distinct or overlapping activations. The situation is identical concerning the multiplication of C_x variables, due to coherence variables: this certainly does not imply multiple copies of mental representations of letters.

If we correctly restrict ourselves to the algebraic notation, the model does not provide directly any prediction about spatial or temporal properties of possible neural correlates. Additional assumptions about the implementation would have to be made. This is a normal feature of algorithmic models: they do not allow directly to claim properties at the implementation level, and only concern mathematical description of information processing systems.

3.2 Bayesian modeling of word recognition and visual attention: BRAID

Biographical note

Collaborator: Sylviane Valdois (LPNC)

Supervised students: Thierry Phénix (Ph.D., ongoing), Svetlana Meyer (Ph.D., ongoing)

In the study of reading, we have recently begun to develop a more thorough Bayesian model of the perceptual processes involved. The overall aim is to provide a mathematical framework able to account for reading acquisition, and, hopefully, reading acquisition disorders. Our starting point, in this prospect, is a Bayesian model of word recognition, that should form a viable basis in order to be extended towards lexicon acquisition and phonological processes.

Because of our specific view on Bayesian modeling, not as an optimal modeling framework, but as a Bayesian Algorithmic Modeling framework, we have strived to provide as detailed as possible an account of the rich perceptual process underlying word recognition. We have thus combined components describing interference effects between neighboring letters, dynamic effects

of perceptual evidence accumulation and memory decay, and visual attentional effects. We have obtained the Bayesian word Recognition using Attention, Interference and Dynamics (BRAID) model.

We provide here a rapid introduction to this model, that is still a work in progress; more precisely, we have so far designed the model and begun to study its parameter space, but have not developed yet all simulations required to demonstrate BRAID's ability to account for all expected effects in the literature (such as the superiority effect, the lexical density effect, etc.). We thus describe here some illustrative experimental results, instead.

3.2.1 BRAID model definition

The BRAID model was defined by the following decomposition of the joint probability distribution:

$$
\begin{aligned}
&P(W^{0:T}\ L_{1:N}^{1:T}\ \lambda_{L_{1:N}}^{1:T}\ A^{1:T}\ C_{1:N}^{1:T}\ P_{1:N}^{0:T}\ \lambda_{P_{1:N}}^{1:T}\ \Delta I_{1:N}^{1:T}\ I_{1:N}^{1:T}\ S_{1:N}^{1:T}\ G^{1:T}) = \\
&P(W^0)\left[\prod_{n=1}^{N} P(P_n^0)\right]\prod_{t=1}^{T}\left[\begin{array}{l}
P(W^t \mid W^{t-1})\prod_{n=1}^{N} P(L_n^t \mid W^t) \\
\prod_{n=1}^{N} P(\lambda_{L_n}^t \mid L_n^t\ P_n^t) \\
P(A^t)\prod_{n=1}^{N} P(C_n^t \mid A^t) P(P_n^t \mid P_n^{t-1}\ C_n^t) \\
\prod_{n=1}^{N} P(\lambda_{P_n}^t \mid P_n^t\ I_n^t) \\
P(G^t)\prod_{n=1}^{N}\left[P(S_n^t) P(\Delta I_n^t) P(I_n^t \mid S_{1:N}^t\ \Delta I_n^t\ G^t)\right]
\end{array}\right] .
\end{aligned}
\tag{3.12}
$$

Subscript indexes $X_{1:N}$ refer to spatial position in a left-to-right letter sequence, superscript indexes $X^{1:T}$ refer to time evolution of variable X from time index 1 to T. Apart for the first few terms of the product, which represent a temporal prior distribution over words $P(W^0)$, and over letters $\prod_{n=1}^{N} P(P_n^0)$, the bulk of BRAID is a product, over time, of a stationary model at time t. A graphical network representing this model at time t is shown Figure 3.7, assuming a sequence of $N = 5$ positions.

The model at time t is composed of three sub-models: a lexical knowledge model (first line of Eq. (3.12)), a visual short-term memory and attention model (third line of Eq. (3.12)) and a low-level visual letter recognition model (fifth line of Eq. (3.12)), articulated by two layers of coherence variable (second and fourth line of Eq. (3.12)). This forms the five lines of the innermost product of Eq. (3.12). We now quickly describe each of the three sub-models, starting from the bottom of both Eq. (3.12) and Figure 3.7.

Low-level visual letter recognition sub-model

This sub-model, in a nutshell, links variables $S_{1:N}^t$, representing the N images forming the visual stimulus, to variables $I_{1:N}^t$, representing the letters in the visual stimulus. In a precise model of this relationship, variables $S_{1:N}^t$ would encode images or geometrical description of shapes, and map these to a discrete categorical space encoding the possible letters of an alphabet. Such a model would include intermediate layers encoding contour detection, feature detection, geometrical configuration of features, and so on.

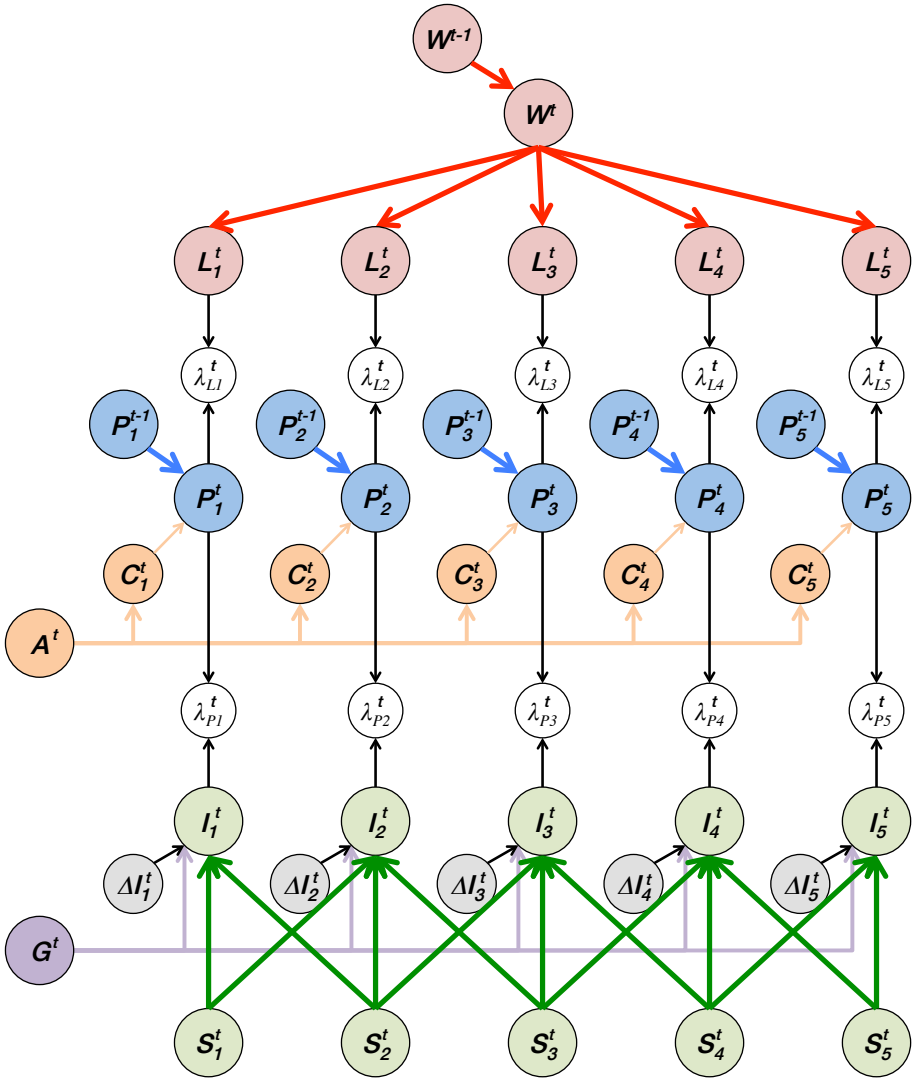

Figure 3.7: **Graphical representation of the structure of the BRAID model.** From bottom to top: a visual stimulus, made of a sequence of letter images $S^t_{1:N}$ is first decoded, providing likely letter identities $I^t_{1:N}$ (in green); this process depends on gaze position G^t (in purple) and involves interference from neighboring stimuli, whose magnitude is represented by $\Delta I^t_{1:N}$; letter identities are then consolidated to memory $P^t_{1:N}$, in which a temporal memory decay (in blue; from $P^{t-1}_{1:N}$ to $P^t_{1:N}$) is counterbalanced by attention $C^t_{1:N}$, piloted by attention repartition A^t (in orange); finally, a lexical model encodes knowledge of letter sequences $L^t_{1:N}$ forming word W^t (in red; also subject to memory decay, from W^{t-1} to W^t). Coherence variables λ connect the three sub-models. Subscript indexes $X_{1:N}$ refer to spatial position, superscript indexes X^t refer to time evolution of variable X.

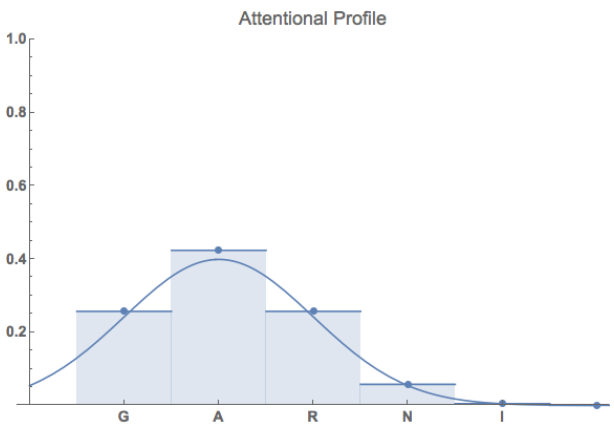

Figure 3.8: **Spatial repartition of visual attention in BRAID.** The x-axis represents spatial positions over a five letter stimulus, the French word "GARNI" (*garnished*). The plots represent a Gaussian probability distribution encoding the repartition of visual attention over positions, and its discrete approximation $P(A^t)$. It maps to control variables C_n^t, allowing to block memory decay. Here, attention is centered on the second letter, and the limited variance, modeling visual attention span, does not allow to maintain information of the fifth letter well.

Here, we summarize this whole process by an overall confusion matrix, from letter identity S_n^t to letter identity I_n^t, so that the domains of both S_n^t and I_n^t are discrete sets $\{a, b, c, \ldots, z\}$. Such a confusion matrix is easily obtained experimentally. In our experiments, we use a classical confusion matrix of letter recognition (Townsend; 1971), mathematically degraded by a free parameter a controlling temporal resolution. In other words, depending on a, information is acquired more or less rapidly, so that more or less time is needed to correctly identify letters.

This visual recognition model is further refined in two ways. Firstly, we assume that neighboring stimuli influence the recognition of letter at position n; this is modeled by lateral "interference", so that letter recognition I_n^t depends on stimuli S_{n-1}^t, S_n^t and S_{n+1}^t. Eq. (3.12) and the present summary feature a simplified notation so as not to differentiate between border letters, that only have a neighbor on one side, and inside letters, that have neighbors on each side. The strength of this lateral interference mechanism is defined by a discrete probability distribution over variable ΔI_n^t with equal interference from left and right flankers (e.g., $\{0.2, 0.6, 0.2\}$), and with the central value numerically controlled by a free parameter θ_I.

Secondly, we model gaze position with a Dirac probability distribution over variable G^t, centered on fixation position g^t. Given g^t, we assume that stimuli directly under this position are recognized faster than peripheral stimuli. This affects parameter a that degrades the confusion matrix of letter recognition, so that a grows linearly with eccentricity, modeling a linear acuity gradient.

Visual short-term memory and attention sub-model

Whereas the first component of BRAID manages the acquisition of information about letter stimuli, the second component of BRAID allows to store and maintain that information. It involves a visual-short term memory whose decay is counterbalanced by visual attention.

Variable P_n^t also has the set $\{a, b, c, \ldots, z\}$ as domain. When coherence variables $\lambda_{P_{1:N}}^t$, acting as Bayesian switches, are "closed", the low-level visual recognition model propagates its information to visual short-term memory, so that the probability distribution over variable P_n^t is the same as the one over variable I_n^t. The dynamics of memory decay, encoded into the transition model $P(P_n^t \mid P_n^{t-1})$, tend to dilute acquired information, so that, without external stimuli, the probability distribution over P_n^t would tend towards a uniform probability distribution (assuming letter frequency is not captured at this level of the model). The speed of memory decay is controlled by a free parameter θ_P.

The memory decay model is refined by a binary control variable C_n^t, which either allows

(when $C_n^t = 0$) or blocks (when $C_n^t = 1$) memory decay. A perfect memory system could thus be modeled: setting all control variables to 1 would ensure no information loss.

However, we limit the system's capability, by modeling a total amount of possible memory retention. This is represented by attention variable A^t, whose domain is an interval of spatial positions $[1, N]$, and $P(A^t)$, a (discrete approximation of a) Gaussian probability distribution of parameters μ_A, σ_A, the position and spread of visual attention. Parameter σ_A can be seen as a formal description of visual attention span (VAS). Visual attention repartition is illustrated Figure 3.8. Finally, probability distribution $P(C_n^t \mid A^t)$ operates the link between visual attention and the control over memory decay, so that each memory slot does not decay in direct proportion of $P(A^t)$.

We note that, thus formulated, the BRAID model allows an independent control of gaze and attention position, by the parameters of probability distributions $P(G^t)$ and $P(A^t)$. We also note that our model of attention does not allow, as is, a varying level of attention, to model superior alertness or impaired attentional resources.

Lexical knowledge sub-model

The third and final component of BRAID encodes orthographic, lexical knowledge in the system. For each word w of set W, a probability distribution over its letters is encoded, assuming independence of the knowledge of letters conditioned on the word, as a product $\prod_{n=1}^{N} P(L_n^t \mid [W^t = w])$. For an expert reader, these distributions are very close to Dirac probability distributions; that is to say, they are the combination of a Dirac probability distribution centered on the correct letter at each position, and ϵ probabilities uniformly distributed over all other letters. Were these ϵ probabilities set to 0, the model would not be able to recognize incorrectly spelled words; a non-zero ϵ value can thus be seen as a crude error model, allowing for any misspelling, and considering all spelling errors equally likely.

The transition model $P(W^t \mid W^{t-1})$ is structurally identical to, and plays a similar role as, the memory decay model over letters $P(P_n^t \mid P_n^{t-1})$: without stimulation, the acquired information about word identity decays, and tends towards a resting state probability distribution $P(W^0)$ which represents word frequency. The dynamics of this decay are controlled by a free parameter θ_W.

3.2.2 Parameter identification in BRAID

As we have seen, there are many parameters in the BRAID model, from the numerical values of the confusion matrix for letter recognition, to the parameters controlling the strength of interference from crowding flankers, to the parameters controlling gaze position and attention repartition, the memory decay parameters of letter percepts and word percepts, etc. Thankfully, most of these parameters have direct, physical interpretations, which suggest how they should be set.

For instance, the parameters of the confusion matrix of course depend on letter stimuli; whether they are capital letters, cursive letters, the complexity of the font, etc. Calibrating the parameters of BRAID on any given experiment in the literature makes the model correspond to the circumstances of that experiment. We do not expect this to be an issue, as the process of word recognition at large should not depend much on such specificities. Moreover, our aim is not to use the BRAID model to provide a mechanistic account of isolated letter recognition, predicting the observed confusion matrices of the literature. Instead, we calibrate our model to

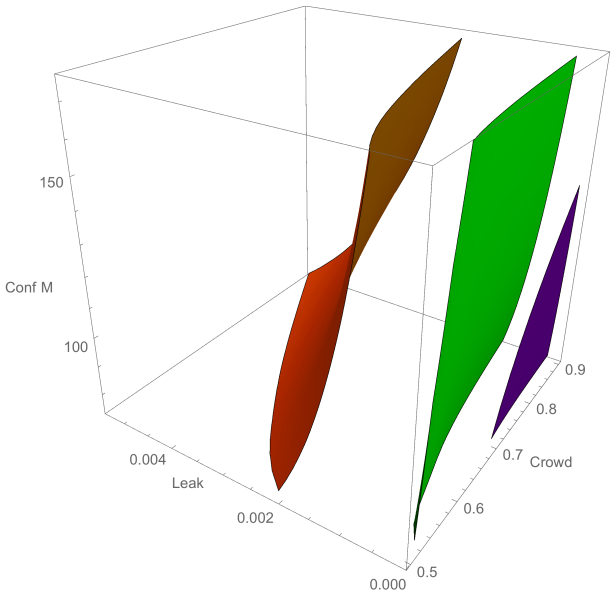

Figure 3.9: **Study of a portion of BRAID's parameter space.** In a variant of the BRAID model, which does not include lexical knowledge, we vary the crowding parameter θ_I (high values indicate low interference), the memory leak parameter θ_P (high values indicate high leak), and the parameter scaling the confusion matrix a (high values indicate slow information input), noted respectively "Crowd", "Leak" and "Conf M" in the plot legend). For each point in this parameter space, we simulate letter recognition for all letters, and for 50 time steps. At the end of these simulations, an average correct recognition score over letters is computed; this plot shows the 99% (in purple), 95% (in green), and 75% (in orange) isosurfaces. Smooth isosurfaces indicate model robustness.

properly summarize the observed statistics of the input-output relation between letter stimuli and recognized letters.

Finally, some parameters remain, which are more difficult to assess from literature. However, they are still easily interpreted. For instance, both lexical and visual short-term memory leak parameters, θ_W and θ_P, influence the dynamics of information loss. Parameter a, which is applied to "degrade" the confusion matrix, affects the dynamics of information gain, as does the crowding interference parameter θ_I. Overall, these can be seen as scaling parameters affecting the quite arbitrary time unit.

We have therefore studied their interaction, experimentally, by a grid search over some combinations of parameters. An illustrative result is shown Figure 3.9, clearly showing smooth variations in this parameter space, without discontinuities or singularities. Our parameter space is thus robust, and parameter values can be set somewhat arbitrarily, with a view to resolve the trade-off between a fine-grained temporal resolution and long computation times for experimental simulations.

3.2.3 Inference in BRAID

The main cognitive task to be solved by the BRAID model is, of course, word recognition. This is modeled by the probabilistic question $P(W^T \mid S^{1:T}_{1:N}\ [\lambda_L{}^{1:T}_{1:N} = 1^{1:T}_{1:N}]\ [\lambda_P{}^{1:T}_{1:N} = 1^{1:T}_{1:N}]\ A^{1:T}\ G^{1:T})$: given gaze position $G^{1:T}$ and attention repartition $A^{1:T}$, given the stimulus image $S^{1:T}_{1:N}$, and assuming that information is allowed to propagate in the whole model architecture ($[\lambda_L{}^{1:T}_{1:N} = 1^{1:T}]$, $[\lambda_P{}^{1:T}_{1:N} = 1^{1:T}]$), what is the probability distribution over words at time step T?

Note that we model a sequence of stimuli $S^{1:T}_{1:N}$, so that priming experiments are easily simulated. For instance, assume that the presented stimulus changes at time t, with $1 < t < T$: $S^{1:t}_{1:N}$ would be the priming word, and $S^{t+1:T}_{1:N}$ would be the target word. Since the probability distribution over words would have not been reset a time t (such a reset would require removing the stimulus and waiting for the word memory decay mechanism to drive the probability distribution over words back to its resting state), priming effects could be experimentally simulated in BRAID.

Bayesian inference to answer the question of word recognition is somewhat too convoluted for the purpose of the present manuscript; we highlight some of its features here. We note the probabilistic question $Q^T = P(W^T \mid S_{1:N}^{1:T}\ [\lambda_L{}_{1:N}^{1:T} = 1_{1:N}^{1:T}]\ [\lambda_P{}_{1:N}^{1:T} = 1_{1:N}^{1:T}]\ A^{1:T}\ G^{1:T})$, and Q^{T-1} the same question at the preceding time step. We obtain:

$$\begin{aligned} Q^T &\propto \sum_{W^{T-1}} \left[Q^{T-1}\ P(W^T \mid W^{T-1})\right] \\ &\quad \prod_{n=1}^{N} \left[\sum_{L_n^T, P_n^T} \left[\begin{array}{l} P(L_n^T \mid W^T)\ P([\lambda_{Ln}{}^T = 1] \mid L_n^T\ P_n^T) \\ P(P_n^T \mid S_{1:N}^{1:T}\ \lambda_P{}_n^{1:T}\ A^{1:T}\ G^{1:T}) \end{array} \right] \right] \qquad (3.13) \\ &\propto \sum_{W^{T-1}} \left[Q^{T-1}\ P(W^T \mid W^{T-1})\right] \\ &\quad \prod_{n=1}^{N} \left\langle P(L_n^T \mid W^T), P(P_n^T \mid S_{1:N}^{1:T}\ \lambda_P{}_n^{1:T}\ A^{1:T}\ G^{1:T}) \right\rangle . \qquad (3.14) \end{aligned}$$

In this derivation, the term $P(P_n^T \mid S_{1:N}^{1:T}\ \lambda_P{}_n^{1:T}\ A^{1:T}\ G^{1:T})$ is the result of the propagation of information from stimulus to letter percepts P_n^T, that is, the whole perception process involving the visual short-term memory and the low-level visual letter recognition models. We do not detail this inference here.

The rest is easily interpreted, with the recursive question Q^{T-1} being multiplied by transition model $P(W^T \mid W^{T-1})$ and by the observation model, as in a classical Hidden Markov Model structure. We remark that here, the observation model is itself structured in a noteworthy manner: the sums yielded by marginalization collapse, thanks to coherence variables, and the result can be interpreted as the dot product between a top-down lexical prediction $P(L_n^T \mid W^T)$ and a bottom-up perceptual process $P(P_n^T \mid S_{1:N}^{1:T}\ \lambda_P{}_n^{1:T}\ A^{1:T}\ G^{1:T})$. In other words, the probabilistic prediction of word form and the probabilistic perception of the stimulus are compared; a high value is returned if they match [3]. An illustrative experimental result of word recognition, for a simulated expert reader, is shown Figure 3.10.

3.2.4 Discussion

We close by mentioning two main directions for future work, related to the two ongoing Ph.D. theses of Thierry Phénix and Svetlana Meyer, respectively concerned with reading acquisition and phonological processes.

In Thierry's work, we first aim to verify that the BRAID model correctly accounts for classical effects of word recognition. We also aim to perform model comparison between variants of the BRAID model, in order to allow studying which components help explain which effect.

For instance, there are parameters that we intentionally left out of the previous presentation of parameter identification in BRAID. Our aim is not to fix them to some calibration value, but, instead, to make them the focus of model comparison experimental studies. They concern eye position g^t of the Dirac probability distribution $P(G^t)$, and position and spread parameters μ_A, σ_A of the Gaussian probability distribution $P(A^t)$ of attention repartition.

[3] The concept of distance measure between probability distributions is mathematically unconstrained, with various measures having different properties and application cases. Here, Bayesian inference and the coherence model with λ variables mathematically result in a dot product form for this distance measure. Since the "norm" of a probability distribution is always unity, this dot product only measures the relative "angle" between the probability distributions. This opens up an intriguing prospect of geometrical interpretation of probabilities and probabilistic calculus. I am not familiar with the corresponding literature, but preliminary research would suggest that it is surprisingly scarce.

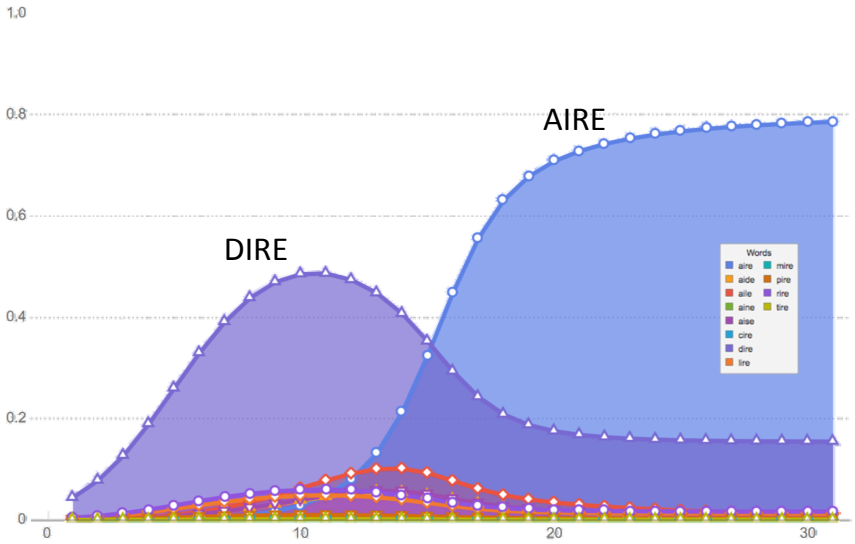

Figure 3.10: **Temporal evolution of word recognition in BRAID.** The x-axis represents 30 simulation time steps. For each time step T, the 12 higher probabilities of the probability distribution over words are plotted. The presented stimulus is the French word "AIRE" (*area*). Because it has low frequency, word recognition first opts towards the "DIRE" (*to say*) word hypothesis, which has higher frequency. This hypothesis is transient, and perceptual information squashes it after around 15 stimulus presentation time steps.

For instance, varying gaze and attention position jointly allows to perform simulations predicting word recognition as a function of viewing position; preliminary results show that optimal viewing positions effects can be reproduced in this manner. We also aim to investigate reading acquisition deficiencies by modulating the visual attention span parameters, e.g., modeling a narrow repartition of attention with a diminished σ_A variance and comparing simulation results to known deficiency patterns of dyslexic readers.

We also expect these parameters to be integrated as components of cognitive control. For instance, we imagine attention repartition strategies to be useful for reading acquisition: when a word is not recognized fast enough, it may be because it is an unknown word, in which case a left-to-right attentional scanning of the word would help identifying portions of the word and commit them to long term lexical memory. Extending the BRAID model in this direction could involve a coherence variable λ_W linking variable W^t and A^t, propagating "soft evidence", that is, the whole probability distribution $P(W^t)$, instead of a single value over space W^t. Then, a probability distribution such as $P(\mu_A \mid [\lambda_W = 1])$ would then allow controlling attention as a function of the distribution over words, to implement strategies adequate in case of slow or fast diminution of entropy of $P(W^t)$. In this manner, we could implement a heuristic control strategy identifying slow convergence of a parallel, global reading procedure, in order to switch to a serial, left-to-right decoding procedure.

Such a heuristic control strategy could also be compared with an optimality based strategy, which would involve computing the optimal attention and gaze parameters so as to maximize word recognition speed. It could be the case that such a strategy would yield a global decoding procedure for well known words, that is, betting that a single eye fixation about the middle of the word, with a large variance of attention reparation, would ensure recognition. Such a strategy might not be efficient for less-known words or words with close lexical neighbors, such that a left-to-right decoding procedure, although slower because of eye movements, would ensure correct

decoding of the stimulus. Implementing such an optimal strategy computation, investigating its result, and comparing with the heuristically defined strategy above could yield insight into reading acquisition procedures.

In Svetlana's project, we aim to extend the BRAID model to include phonological representations. Initial design of this extension revolves around the assumption that the word variable W^t would act as a pivot between lexical and phonological knowledge. Recall that, in BRAID's lexical knowledge sub-model, word identity predicts sequences of letters, with a fusion model $\prod_{n=1}^{N} P(L_n \mid W)$. A phonological knowledge sub-model would feature a similar architecture, in which words would associate with sequences of phonemes $\Phi_{1:K}$, using $\prod_{k=1}^{K} P(\Phi_k \mid W)$. Let us call BRAID-Phi this extension of BRAID.

A crucial question concerns whether BRAID-Phi would require an additional lexical-to-phonological mapping, absent in the design described above, or whether it would implicitly be encoded already. In the same manner that, in BRAID, the knowledge of overall letter frequencies, and position specific letter frequencies can be inferred by marginalizing over W, it could be the case that lexical-to-phonological statistics might be inferred from BRAID-Phi, from the word-to-phonology and the word-to-lexicography models. BRAID-Phi would then implicitly encode lexical-to-phonological knowledge, and would be indistinguishable functionally from a variant where it is explicit. That would question previous models of the literature, in which lexical-to-phonological models are central components, assumed to be necessary to explain the system's behavior.

CHAPTER 4

Bayesian modeling of speech perception and production

This Chapter focuses on a family of models we have developed, in the context of speech perception and production. This family is encompassed by an overall, abstract model, called COSMO (both for "Communication of Objects using Sensori-Motor Operations" and also for listing the five variables of the model, C, O_S, S, M and O_L). COSMO has then been instantiated, in three separate studies: first, concerning the emergence of phonological systems, second, concerning speech perception, and third, concerning speech production. As in the previous Chapter, we focus mainly on model description, omitting discussions about the relevance of these models and justification with respect to the literature.

4.1 Bayesian modeling of communicating agents: COSMO

Biographical note

Collaborators: Pierre Bessière (ISIR, Paris), Jean-Luc Schwartz (GIPSA-Lab, Grenoble)

Supervised students: Clément Moulin-Frier (Master 2 and Ph.D., defended in 2011), Raphaël Laurent (Master 2 and Ph.D., defended in 2014)

Publications: Moulin-Frier et al. (in press), Laurent (2014), Moulin-Frier et al. (2012), Moulin-Frier (2011), Moulin-Frier et al. (2010)

This section adapts material from Moulin-Frier et al. (2012), Moulin-Frier et al. (in press), and a submitted manuscript currently under review.

4.1.1 COSMO model definition

The starting point of our analysis is a conceptual model of speech mediated communication between two cognitive agents (illustrated Figure 4.1, top). It is a straightforward model, where a speaker has the goal to communicate about an object O_S to a listener. "Object" means here a communication object in a broad sense, whatever its nature, either semantic, syllabic, phonemic, etc. The speaker is equipped with a set of representations and control processes acting on a vocal

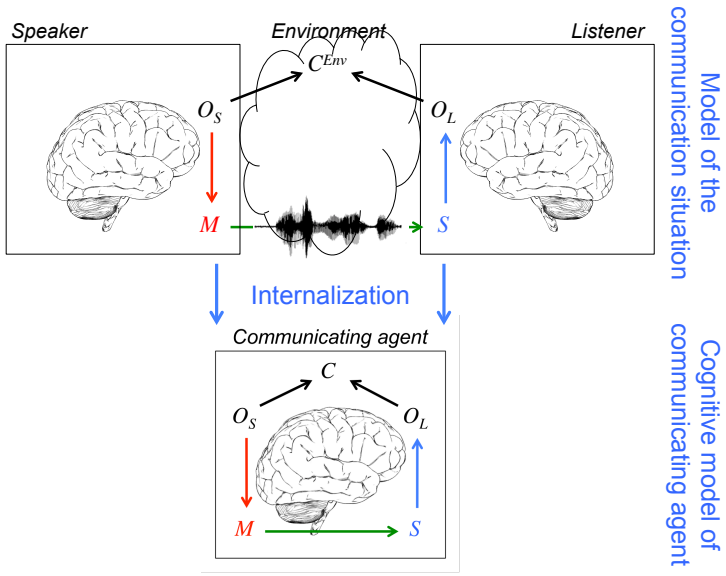

Figure 4.1: **General models of a communication situation and a communicating agent.** COSMO is a communicating cognitive agent model (bottom) resulting from internalizing a communication situation model (top).

tract through different articulators. We globally denote this set as M (for Motor). The speaker's vocal tract produces a sound wave, from which the listener has to infer the communicating object by the means of an ear allowing perceiving the sound, as well as a brain including a set of representations and auditory processing. We globally denote this set S (for Sensory), and call O_L the object inferred by the listener. Finally, the communication success is defined by the condition $O_S = O_L$, noted $C = 1$.

The central hypothesis we make is that a communicating agent, which is potentially both a speaker and a listener, is able to fully internalize the communication situation described previously (Figure 4.1, top) inside an internal model (Figure 4.1, bottom). This "internalization" hypothesis results in the COSMO Bayesian model of a communicating cognitive agent. It is described by a joint probability distribution $P(C\ O_S\ S\ M\ O_L)$, which we decompose according to:

$$P(C\ O_S\ S\ M\ O_L) = P(O_S)P(M \mid O_S)P(S \mid M)P(O_L \mid S)P(C \mid O_S\ O_L)\ . \tag{4.1}$$

This decomposition features, notably, a motor system $P(M \mid O_S)$ able to associate communication objects O_S with motor gestures M; a sensory-motor link $P(S \mid M)$ able to associate motor gestures M with auditory stimuli S, and providing an internal model of the articulatory-to-acoustic transformation; an auditory system $P(O_L \mid S)$ able to associate communication objects O_L with auditory stimuli S; and a communication validation system $P(C \mid O_S\ O_L)$ able to check the communication success condition.

At this stage, the only term of this decomposition which is constrained is $P(C \mid O_S\ O_L)$: C acts as a coherence variable (see Section 2.3.2), so that $P([C = 1] \mid O_S\ O_L) = 1$ if and only if $O_S = O_L$. A final technical point concerns the auditory system, presented here, somewhat counter-intuitively, as $P(O_L \mid S)$: concrete instantiations of the model feature a generative model of the form $P(S \mid O_L)$, which is either included in the COSMO architecture thanks to a sub-model recursive call (see Section 2.3.1), or by a coherence variable (see Section 2.3.2).

4.1.2 Inference in COSMO

An agent provided with a COSMO cognitive architecture possesses a model of the entire communication situation and is thus able to perform both production and perception tasks within

	Production infer $P(M \mid O)$	Perception infer $P(O \mid S)$
Motor theory focus on O_S	$\underbrace{P(M \mid O_S)}_{\text{motor repertoire}}$	$\propto \sum_M \Big(\underbrace{P(M \mid O_S)}_{\text{motor decoder}} \quad \underbrace{P(S \mid M)}_{\text{inverse model}} \Big)$
Auditory theory focus on O_L	$\propto P(M) \sum_S \Big(\underbrace{P(S \mid M)}_{\text{direct model}} \quad \underbrace{P(O_L \mid S)}_{\text{sensory targets}} \Big)$	$\underbrace{P(O_L \mid S)}_{\text{sensory classifier}}$
Perceptuo-motor theory $C{=}True$, i.e. $O_S{=}O_L$	$\propto \underbrace{P(M \mid [O_S{=}O_L])}_{\text{motor production}} \underbrace{\sum_S \Big(P(S \mid M)P(O_L \mid S)\Big)}_{\text{sensory production}}$	$\propto \underbrace{P([O_L{=}O_S] \mid S)}_{\text{sensory perception}} \underbrace{\sum_M \Big(P(M \mid O_S)P(S \mid M)\Big)}_{\text{motor perception}}$

Figure 4.2: **Bayesian questions and inferences for speech production and perception tasks, instantiated within the framework of the motor, auditory and perceptuo-motor theories.** The $\propto$ symbol denotes proportionality, i.e., to correctly obtain probability distributions, the shown expressions have to be normalized.

an integrated perceptuo-motor architecture. Indeed, technically, from the joint probability distribution $P(C\ O_S\ S\ M\ O_L)$, we can apply Bayesian inference to simulate speech perception and production tasks, which appear as probabilistic questions addressed to the joint probability distribution. Perception tasks can be simulated by computing probability distributions over objects, given an input sensory signal, i.e., terms of the form $P(O \mid S)$. Production tasks can be simulated by computing probability distributions over motor gestures, given an object to communicate about, i.e., terms of the form $P(M \mid O)$.

Note that, in these probabilistic questions, we did not yet specify which object variable was O: indeed, we have a dual representation of internal objects, with O_S and O_L. Is O either one, or both? The driving reasoning of the COSMO approach to speech communication is that motor, auditory and perceptuo-motor theories of speech perception and production can be defined in reference to this choice about which object is considered in the probabilistic questions. In other words, auditory theories of speech perception and production amount to choosing O_L, motor theories amount to choosing O_S, and perceptuo-motor theories amount to choosing both, that is to say, considering $C = 1$.

We show Figure 4.2 the Bayesian inference resulting from each question posed to the model of Eq. (4.1). We do not discuss here in detail the content of each inference. We just note a few remarkable results. First, the Bayesian inference corresponding to a motor theory of speech perception can be interpreted as a Bayesian implementation of analysis-by-synthesis, but without an explicit inversion step of the sensorimotor mapping, which still appears as $P(S \mid M)$. Second, the Bayesian inference corresponding to an auditory theory of speech production implements a classic production process, constrained both by prediction of sensory outputs and by the phonetic targets as described in auditory space. Finally, we note that perceptuo-motor theories, both of perception and production, can be interpreted as fusions of motor and auditory theories, as they mathematically amount to products of the corresponding Bayesian equations.

This shows that the COSMO architecture is a theoretical framework rich enough to capture the main trends of the debate between auditory vs motor vs perceptuo-motor theories of speech perception and production. We also argue that it is the minimal framework containing these, and thus allowing their mathematical comparison in a unified setting; indeed, no term can be removed from the joint probability distribution decomposition without removing part of some inference of Figure 4.2.

4.2 Bayesian modeling of language universals: COSMO-Emergence

Biographical note

Collaborators: Pierre Bessière (ISIR, Paris), Jean-Luc Schwartz (GIPSA-Lab, Grenoble)
Supervised student: Clément Moulin-Frier (Master 2 and Ph.D., defended in 2011)
Publications: Moulin-Frier et al. (in press), Moulin-Frier (2011), Moulin-Frier et al. (2011), Moulin-Frier et al. (2010, 2008)
This section adapts material from Moulin-Frier et al. (in press)

Human languages display a number of regularities, called "universals", which result in the fact that all human languages, while they are different from one another, still contain general principles and follow strong statistical trends. This for instance concerns the number and acoustic distribution of vowels and consonants. We have studied these universals with the assumption that they would be the emergent product of an interaction process. This process would induce some commonality in the achieved solutions because of the commonality in the cognitive mechanisms at hand in the involved dynamic mechanisms, and common exterior constraints.

This is a classical framework, where emergence of language systems are explored, experimentally, by multi-agent simulations of "language games". In these multi-agent models, agent populations are made to interact and evolve, leading to the emergence of global properties from local interactions, and these properties are analyzed in relation with those of language universals. We have defined "deictic games" as the central tool of our simulations, in which societies of agents implementing the COSMO model interact in presence of objects that they attempt to designate by the voice. In the context of this manuscript, we refer to this framework as the COSMO-Emergence model.

We have performed three main studies, featuring increasingly complex variants of the COSMO-Emergence model. The first simply involves one-dimensional motor and sensory spaces, and allows to extract general cognitive conditions necessary for the emergence of a speech code. The second involves a more realistic vocal tract simulator, the VLAM (*Variable Linear Articulatory Model*) simulator (Maeda; 1990, Boë; 1999), that we use to investigate vowel emergence; it allows to explore the relative weights of formant dimensions. The third and final study adds control of the jaw, to study the emergence of stop consonants; it allows to explore the effect of cyclical movements on the presence of pharyngeal consonants. A fourth and somewhat preliminary study combined our vowel and stop consonant simulations to verify our system could reproduce the more commonly observed syllabic systems. However, these last results were not solidified with systematic model comparison, so we do not develop them further here (but see (Moulin-Frier et al.; in press)).

In this section, and again, focusing primarily on models and experimental model comparison, we first summarize the three COSMO-Emergence variants, how they were used in deictic games for parameter identification and simulating communication code emergence, and how the experimental results allowed to answer scientific questions of interest.

4.2.1 COSMO-Emergence model definition

One-dimensional COSMO-Emergence variant

We first considered a simple one-dimensional instantiation of the model, which allows extracting general properties of motor, auditory and sensory-motor behaviors. It instantiates the COSMO general model of Eq. (4.1) with the motor and sensory variables, M and S, defined as integer values in the range $\{-10, \ldots, 10\}$. The articulatory-to-acoustic transformation $\mathit{TransMS} : M \rightarrow S$ is a sigmoid function parameterized in a way which allows to change from a purely linear function to a step function. It is defined as:

$$\mathit{TransMS}(m) = S_{max} \left(\frac{\arctan(NL(m - D))}{\arctan(NL\ M_{max})} \right) \tag{4.2}$$

where $S_{max} = M_{max} = 10$ according to M and S range specifications, D, the position of the inflexion point, is 0, and the remaining free parameter NL controls non-linearity. $P(S \mid M)$ is a Gaussian model centered on the value provided by the *TransMS* transfer function, with a constant simulated noise σ_{Env}.

Finally, the motor and auditory prototypes, $P(M \mid O_S)$ and $P(S \mid O_L)$, are defined as Gaussian probability distributions [1], one for each possible object. Recall that $P(S \mid O_L)$ is not included directly as is. Instead, in the COSMO-Emergence variants, it is featured in a sub-model π_{sub}:

$$P(S\ O_L \mid \pi_{sub}) = P(O_L \mid \pi_{sub}) P(S \mid O_L\ \pi_{sub}) \ . \tag{4.3}$$

Sub-model π_{sub} allows computing $P(O_L \mid S\, \pi_{sub})$, which is then included in Eq. (4.1) as parametrical form of $P(O_L \mid S)$. Since all variables are one-dimensional, the bottom part of Figure 4.1 can actually be seen as a graphical representation of this first variant of COSMO-Emergence.

Vocalic COSMO-Emergence variant

The second variant of COSMO-Emergence we defined was based on the VLAM simulator. For the purpose of the present document, we just introduce it as a simulator of the vocal tract. It takes as input a geometrical configuration of the vocal tract, characterized by 7 articulatory variables. Out of these, for simplicity, we only consider 3 as motor variables: $M = TB \wedge TD \wedge LH$, so that we simulate variations of the tongue body position (TB), tongue dorsum (TD) and lips separation height (LH). The four remaining variables are set to a neutral position. The output of VLAM is an acoustic signal, characterized by 3 formant frequencies, that we consider as a sensory variable: $S = F_1 \wedge F_2 \wedge F_3$. Formants are expressed in Barks, a lin-log scale reflecting human frequency perception. With these sensory and motor variable, we define the second variant of COSMO-Emergence as:

$$\begin{aligned} P(C\ O_S\ S\ M\ O_L) &= P(C\ O_S\ F_1\ F_2\ F_3\ TB\ TD\ LH\ O_L) \\ &= P(O_S) P(TB\ TD\ LH \mid O_S) P(F_1\ F_2\ F_3 \mid TB\ TD\ LH) \\ &\quad P(O_L \mid F_1\ F_2\ F_3) P(C \mid O_S\ O_L) \ . \end{aligned} \tag{4.4}$$

The motor prototype term and the acoustic prototypes (in the sub-models) are further simplified as products of three one-dimensional Gaussian distributions:

$$P(TB\ TD\ LH \mid O_S) = P(TB \mid O_S) P(TD \mid O_S) P(LH \mid O_S) \ , \tag{4.5}$$

$$P(F_1\ F_2\ F_3 \mid O_L) = P(F_1 \mid O_L) P(F_2 \mid O_L) P(F_3 \mid O_L) \ . \tag{4.6}$$

[1] Technically, these are discretized and truncated approximations of Gaussian probability distributions. This is also the case wherever this is needed. However, to make the text lighter, we do not precise this every time.

Finally, the term $P(F_1\ F_2\ F_3 \mid TB\ TD\ LH)$ is still a Gaussian model centered on the *TransMS*, but this transfer function is computed off-line as a probabilistic approximation of the transfer function of VLAM. We do not show a graphical representation of this model, as its dependency structure does not correspond to a Bayesian network; thus, such a schema would be hardly useful.

Consonantal COSMO-Emergence variant

In the third and final variant of the COSMO-Emergence model, we added the articulatory variable describing the jaw position J into the motor variable, which becomes four-dimensional: $M = J \wedge TB \wedge TD \wedge LH$. All the rest in unchanged in form, but adapted to this four-dimensional motor space.

4.2.2 Parameter identification and inference in COSMO-Emergence

We have presented three variants of the COSMO-Emergence model for the study of emergent properties of communication interaction. They vary in the dimensionality and realism of the considered motor and sensory spaces. However, in all cases, the COSMO architecture is applied, and the only remaining free parameters are those of Gaussian probability distributions in motor and sensory prototypes. These are learned in emergence simulations, that we call "deictic games".

In such a simulation, we consider a society of N agents interacting in an environment containing Q objects. Initially, the Gaussian prototypes of the form $P(M \mid O_S)$ and $P(S \mid O_L)$ encode lack of knowledge, i.e., with centered means and large standard deviations. A simulation then consists in a series of deictic games.

During a deictic game, two agents meet in presence of a uniformly randomly drawn object o_i. One agent takes the role of speaker, the other of listener, also randomly. The speaker agent chooses a motor gesture m; simulations will vary in the manner m is chosen. Once m is selected, the simulated environment transforms it into an acoustic signal by drawing a stimulus value s according to the articulatory-to-acoustic model, using transfer function *TransMS*(m) and the simulated noise σ_{Env}. We also assume that object o_i is perfectly perceived by both agents. At the end of a deictic game, the speaker and listener agents respectively update parameters of their motor and auditory Gaussian prototypes, according to the observed $< o_i, m >$ pair for the speaker, and to $< o_i, s >$ for the listener.

4.2.3 Model variants and their comparison

Condition on agent inference for communication code emergence

Each of the three variants of the COSMO-Emergence model was used to answer a scientific question of interest, using model comparison. In the first study, using the one-dimensional variant of the model, we investigated the conditions for the emergence of communication codes among societies of agents. We have simulated three types of societies, differing in the strategy for selecting a motor gesture m during a deictic game: in the "motor" society, agents use a motor theory of production, in the "auditory" society, they use an auditory theory of production, and, finally, in a "perceptuo-motor" society, they use a perceptuo-motor theory of production (i.e., they each use a different case of the production column of Figure 4.2). Each simulation concerned societies of $N = 4$ agents and $Q = 4$ objects.

Figure 4.3 shows the state of agent societies at the end of three typical simulations, one for each version of the model. Figure 4.3 (top panel, labeled "A") shows the result at the end of a

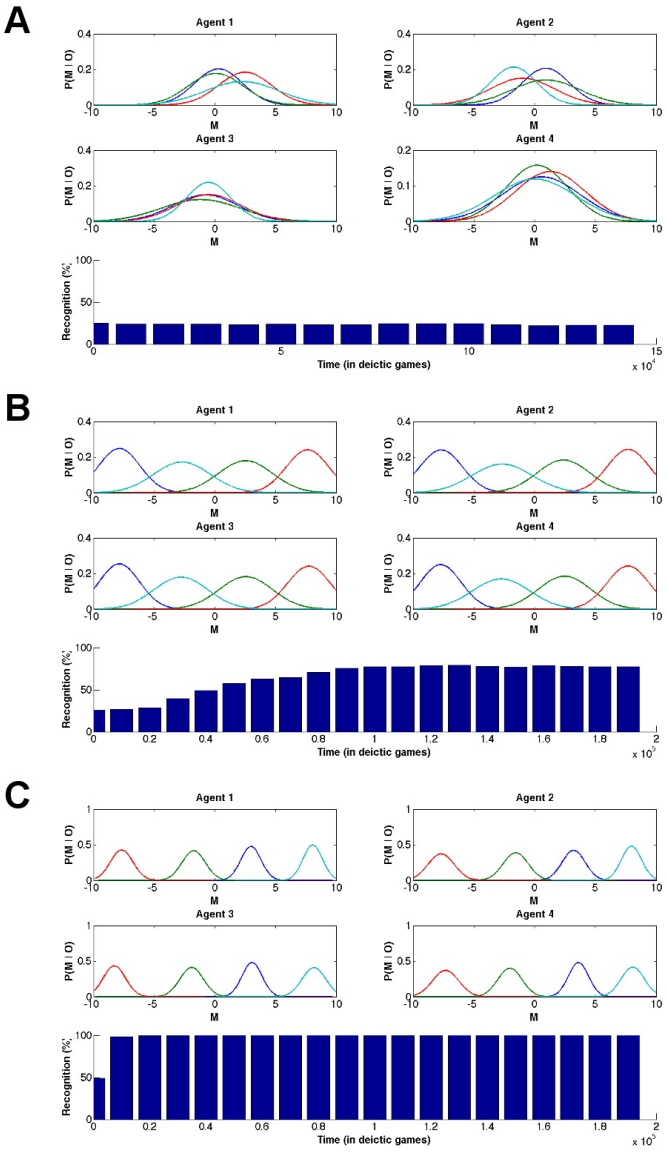

Figure 4.3: **Emergence of communication codes in societies of COSMO-Emergence agents.** Top panel (A), middle panel (B) and botttom panel (C) show a typical simulation of respectively "motor" agents, "sensory" agents and "perceptuo-motor" agents. Each panel shows the motor prototypes $P(M \mid [O_S = o_i])$ for the four objects o_i (each colored curve refers to an object), at the end of a typical simulation, for the four agents (each sub-panel refers to an agent). The bottom plots show the evolution, over simulated time, of communication success rate.

simulation of "motor" agents, i.e., a simulation where agents produce gestures according to the motor production behavior of Figure 4.2). We observe that motor prototypes of agents have evolved relatively randomly during the learning process, and that the recognition rate stays at chance level (around 25%). As agents do not use auditory knowledge, and therefore have no means to link their productions in front of each object with the auditory inputs they receive from the others, there is no chance that such a society would allow the emergence of a structured and efficient speech code to designate objects.

On the contrary, "auditory" and "perceptuo-motor" agent societies converge to usable communication codes (Figure 4.3, middle and bottom panel, respectively). We observe that motor prototypes for each object are well contrasted and that an agreement between agents has emerged. Moreover, the communication success rate increases during simulations, and more rapidly for "perceptuo-motor" societies. Compared to the auditory production behavior, adding the motor sub-system therefore allows reducing the variability of the chosen motor gestures, leading to optimally distinguishable motor prototypes and a rapide increase to 100% recognition rate. This is due to the $P(M \mid [O_S = o_i])$ factor in the perceptuo-motor speech production behavior, which results in anchoring the speech production behavior around selected prototypes.

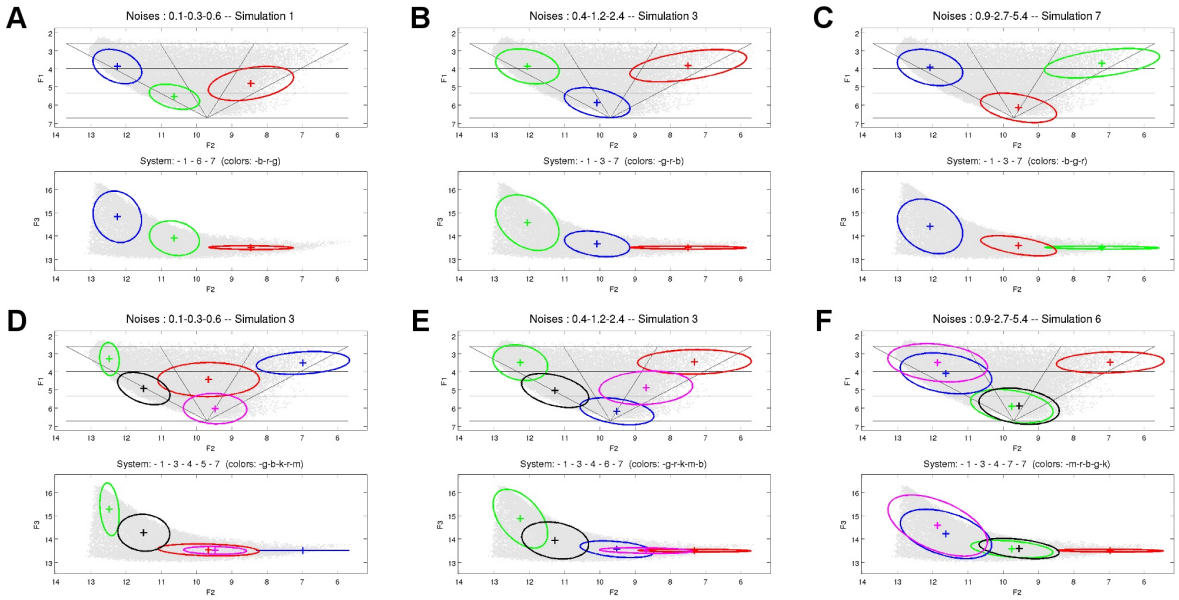

Figure 4.4: **Emergence of vowel systems in societies of COSMO-Emergence agents. Top row:** 3-vowel simulations. **Bottom row:** 5-vowel simulations. **Left, middle, right columns:** Simulations of increasing baseline noise value σ_{F1}. Each panel features two plots, showing the acoustic prototypes at the end of simulations in the $(F2, F1)$ and the $(F2, F3)$ planes, respectively.

Condition on noise parameters for vowel system emergence

In the second study, we used the second COSMO-Emergence variant, that is, with three-dimensional motor and sensory spaces. We studied the emergence of vowel systems, by considering open configurations of the vocal tract in the dictionary of simulated VLAM configurations. We have simulated societies of $N = 2$ "perceptuo-motor" agents, in environments with either $Q = 3$ or $Q = 5$ objects, in order to study 3- and 5-vowel systems. We have varied the environmental noise. In this three-dimensional variant, it is defined by three parameters $\sigma_{F1}, \sigma_{F2}, \sigma_{F3}$. We have both varied their absolute and relative values. That is, we have varied the noise value σ_{F1}, and for each value, compared "1-1-1" noise profiles to "1-3-6" profiles, i.e., simulations where $\sigma_{F1} = \sigma_{F2} = \sigma_{F3}$ or where $\sigma_{F2} = 3\sigma_{F1}$, and $\sigma_{F3} = 6\sigma_{F1}$, in order to test a previously proposed hypothesis stating that successive acoustic dimensions had decreasing cognitive weights.

Some typical results of simulations for "1-3-6" noise profiles are shown, for illustration purposes, Figure 4.4. Our experimental results confirm that the "1-3-6" noise profile favors the emergence of expected vowel systems, i.e., those most common in human languages, contrary to the "1-1-1" noise profile, especially for 5-vowel systems. They also confirm that, when environmental noise is too large, dispersion of acoustic prototypes is not feasible, and communication codes cannot emerge.

The fact that emerging 3-vowel and 5-vowel systems are in line with human language data is not original, but instead reproduces previous studies. However, these simulations enable to specify the range of experimental parameters that lead to realistic predictions of vowel systems.

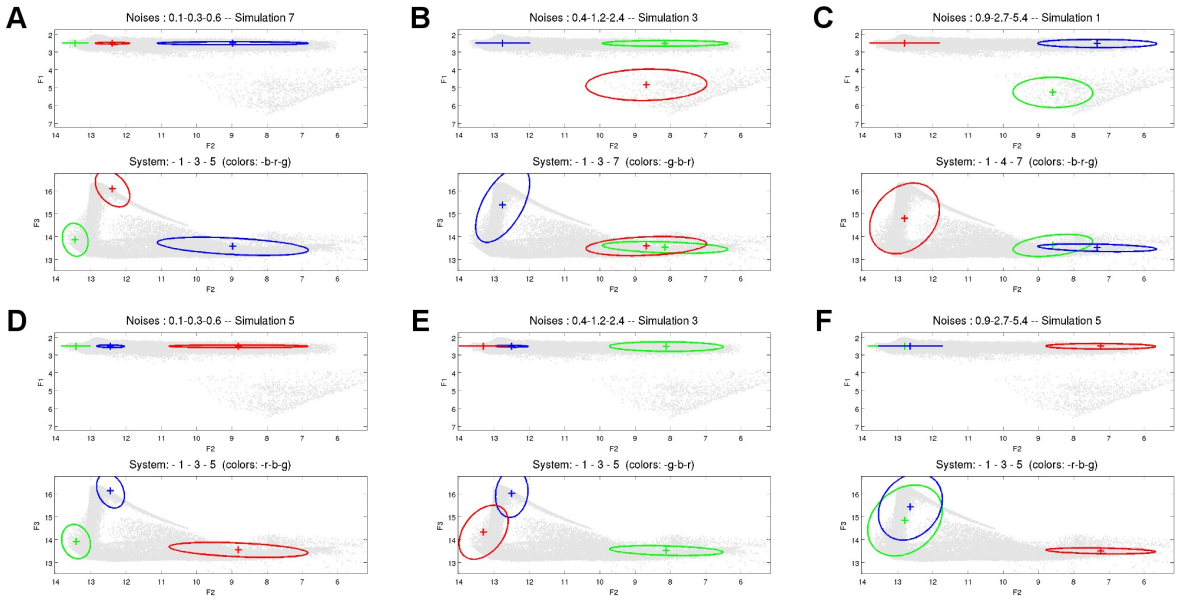

Figure 4.5: **Emergence of stop consonant systems in societies of COSMO-Emergence agents. Top row:** "free jaw" simulations. **Bottom row:** "high jaw" simulations. **Left, middle, right columns:** Simulations of increasing baseline noise value σ_{F1}. Each panel features two plots, showing the acoustic prototypes at the end of simulations in the $(F2, F1)$ and the $(F2, F3)$ planes, respectively.

Condition on jaw cycles for consonant system emergence

In the third and final study we summarize here, we have used the third variant of COSMO-Emergence, where control of the jaw is added as a fourth motor dimension. This allows to study the emergence of stop consonants, by considering almost closed configurations of the vocal tract in the dictionary of simulated VLAM configurations. Starting from the results of previous studies (perceptuo-motor agents, "1-3-6" noise ratios), we have compared two models of stop consonant production, by constraining the motor term $P(J \mid O_S)$. In the "high jaw" condition, we force consonants to be produced in the high jaw portion of the syllabic cycle; this is done by setting a mean set to a high value and low standard deviation, as initial parameters for the $P(J \mid O_S)$ Gaussian probability distribution. On the contrary, in the "free jaw" condition, $P(J \mid O_S)$ is initially unconstrained, with a mean corresponding to a neutral jaw position, and a large initial standard deviation. As previously, we have considered $N = 2$ agents, and $Q = 3$ objects, that is, simulated the emergence of 3-stop consonant systems.

Typical results of simulations are shown, for illustration purposes, Figure 4.5. Our experimental results show that in the "free jaw" condition, /b d g/ systems emerge at a low noise level, but pharyngeal stops tend to appear at medium and high levels, as expected, considering their high $F1$ values which make them excellent candidates for acoustic distinctiveness. In the "high jaw" condition, the /b d g/ consonant system is strongly preferred, pharyngeal consonants being discarded from the simulations by the "high jaw" constraint. In human languages, however, stop consonant systems seldom include pharyngeal consonants. In other words, to adequately reproduce the observed regularity of human languages, we must consider a constraint on the jaw position. This is in line with previous proposals, such as the Frame-Content theory, which gives a fundamental role to the jaw cycle, so that vowels and consonants would respectively correspond to open and closed portions of that cycle.

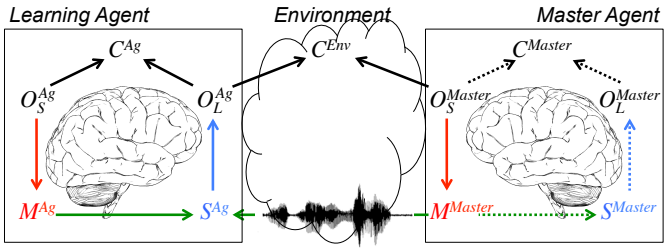

Figure 4.6: **Schema of the supervised learning scenario in COSMO**, where the Master Agent provides the Learning Agent with ⟨object, stimulus⟩ pairs.

4.3 Bayesian modeling of speech perception: COSMO-Perception

Biographical note

Collaborators: Pierre Bessière (ISIR, Paris), Jean-Luc Schwartz (GIPSA-Lab, Grenoble)

Supervised students: Raphaël Laurent (Master 2 and Ph.D., defended in 2014), Adrien Delmas (Master 2), Marie-Lou Barnaud (Ph.D., ongoing), Clément Moulin-Frier (Ph.D., defended in 2011)

Publications: Laurent (2014), Laurent et al. (2013a,b), Moulin-Frier et al. (2012), Laurent et al. (2012)

This section adapts material from Laurent (2014) and a manuscript currently under review

In the previous Section, we dealt with the question of speech communication code emergence, with a focus on constraints on motor production of agents. Perception tasks were not considered, other than to verify whether communication was successful. In this Section, we study perception tasks more closely, but not in the context of language games, focusing instead on on-line perception, that is to say, decoding of speech signals provided by a speaker agent (or master agent, when learning is considered). Our overall goal is to compare purely motor, purely auditory, and perceptuo-motor theories of speech perception (see the perception column of Figure 4.2).

To do so, we have first studied the perception tasks in the abstract COSMO model, i.e., in its general formulation without assumptions about representational dimensionality (Section 4.1). The first result is an indistinguishability theorem. In a nutshell, it demonstrates that, despite having different expressions, auditory and motor theories of speech perception can, in ideal learning conditions, exactly capture the same knowledge, and thus, be experimentally indistinguishable. Identifying such ideal learning conditions allowed to explore a learning algorithm, which we called "learning by accommodation", that falsifies these conditions, and thus makes the theories distinguishable.

We have then experimentally studied learning and perception tasks both in the one-dimensional variant of COSMO that we previously referred to as COSMO-Emergence (see Section 4.2.1), and in a syllabic variant of COSMO, that we referred to as COSMO-S elsewhere (Laurent; 2014). For the purpose of the current manuscript however, we conflate these models, focus on the presentation of the syllabic model, and unify them as COSMO-Perception.

4.3.1 Preamble: indistinguishability theorem

We have first studied how the auditory and motor theories of speech perception (see Figure 4.2, right column) would perform during and after learning, in the abstract COSMO formulation. We have considered a supervised learning scenario, illustrated Figure 4.6, featuring a Learning Agent and Master Agent, each described as a COSMO agent. In order to distinguish their variables, superscripts are added, and variables become O_S^{Ag}, O_S^{Master}, M^{Ag}, M^{Master}, etc.

In the learning scenario, the Learning Agent is provided by the Master Agent with ⟨object, stimulus⟩ pairs obtained as follows. The Master Agent randomly selects an object, selects a motor gesture that corresponds and outputs it, resulting in sensory input S^{Ag}, received by the Learning Agent as stimulus. In an independent communication channel, e.g., a shared attention mechanism, object identity is transferred from the Master to the Learning Agent, via variable C^{Env}.

The received ⟨object, stimulus⟩ pairs allow the learning agent to identify parameters of its sensory classifier $P(S^{Ag} \mid O_L^{Ag})$, following a learning algorithm that, if it is well constructed, will tend towards capturing the observed process, i.e., the productions of the Master Agent. In mathematical terms, this means that the Learning Agent identifies its sensory classifier based on successive data, sampled from a process which approximates:

$$P(O_L^{Ag} \mid S_{Ag}) \approx \sum_M P(M^{Master} \mid O_S^{Master}) P(S^{Ag} \mid M^{Master}) \ . \tag{4.7}$$

We now define the three hypotheses which ensure the indistinguishability of the motor and auditory theories of speech perception: **H1**: the sensory classifier of the Learning Agent is perfectly learned from the Master's productions; **H2**: the motor repertoire of the Learning Agent is perfectly identical to the motor repertoire of the Master; **H3**: the Learning Agent's sensory-motor system perfectly encodes the properties of the transformation performed by the environment.

Each of these hypotheses allows to modify Eq. (4.7): **H1** allows to replace the approximation $\approx$ by an equality $=$; **H2** allows to replace $P(M^{Master} \mid O_S^{Master})$ by $P(M^{Ag} \mid O_S^{Ag})$; **H3** allows to replace $P(S^{Ag} \mid M^{Master})$ by $P(S^{Ag} \mid M^{Ag})$. This yields:

$$P(O_L^{Ag} \mid S_{Ag}) = \sum_M P(M^{Ag} \mid O_S^{Ag}) P(S^{Ag} \mid M^{Ag}) \ . \tag{4.8}$$

The right hand side of Equation (4.8) is the expression of the motor instantiation of the speech perception task, whereas the left hand side is the expression of the perception task instantiated within the framework of the auditory theory (see Figure 4.2).

Therefore, if these three hypotheses are verified, they describe "perfect conditions" for learning, such that the sensory and motor models of perception rely on the same information, make the same predictions, and are thus indistinguishable, whatever the testing conditions might be. Therefore, understanding the potential role and complementarity of the sensory and motor recognition processes requires departing from the perfect conditions defined above.

4.3.2 COSMO-Perception model definition

Deviating from the hypotheses of the indistinguishability theorem can be done in several ways. It can be done structurally, by limiting the expressive power of terms involved in Equation (4.3.1), i.e., by introducing representational assumptions in the COSMO agent. It can also be done algorithmically, for instance by studying the state of the Learning Agent before its asymptotic convergence. We have done both in the COSMO-Perception model, where we considered the learning of Plosive-Vowel syllables.

We have considered 9 syllables combining three stop consonants and three vowels; the domains for variables O_S and O_L are the set containing syllables /ba/, /bi/, /bu/, /ga/, /gi/, /gu/, /da/, /di/, /du/. Since we model a syllable as a vowel state and a consonant state, variable S separates into S_V and S_C, and variable M into M_V and M_C. Apart from that, the COSMO-Perception

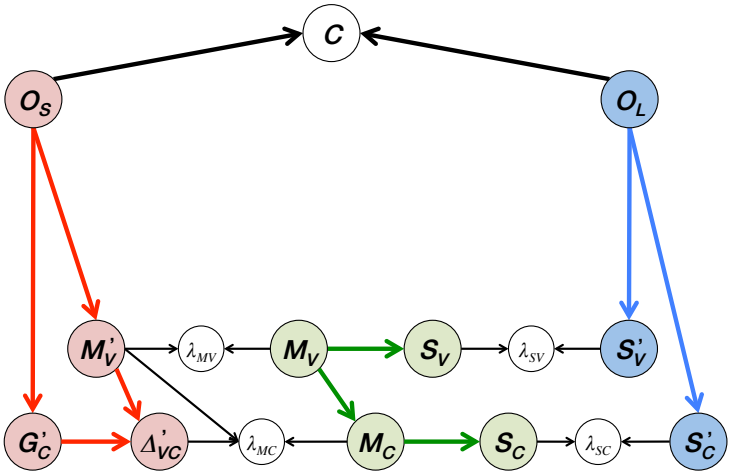

Figure 4.7: **Graphical representation of the structure of the COSMO-Perception model.** In red, the motor system; in green, the sensory-motor system; in blue, the auditory system; in light gray, the coherence variable system connecting the parts of the model.

model shares its global structure with COSMO (see Figure 4.7): (i) the auditory system associates sensory representations with the corresponding phonemic representations O_L (i.e., the syllable labels); (ii) the sensory-motor system associates motor and sensory representations; (iii) the motor system associates motor representations with phonemic representations O_S. These systems are linked together by coherence variables λ, duplicating variables, such as M_V into M_V and M'_V, etc.

The decomposition of the COSMO-Perception joint probability distribution is as follows:

$$\begin{aligned} & P(O_S\ G'_C\ M'_V\ \Delta'_{MC}\ \lambda_{MV}\ \lambda_{MC}\ M_V\ M_C\ S_V\ S_C\ \lambda_{SV}\ \lambda_{SC}\ S'_V\ S'_C\ O_L\ C) \\ = \ & P(O_S)P(M'_V \mid O_S)P(G'_C \mid O_S)P(\Delta'_{MC} \mid M'_V\ G'_C) \\ & P(\lambda_{MV} \mid M'_V\ M_V)P(\lambda_{MC} \mid M'_V\ \Delta'_{MC}\ M_C) \\ & P(M_V)P(S_V \mid M_V)P(M_C \mid M_V)\ P(S_C \mid M_C) \\ & P(\lambda_{SV} \mid S_V\ S'_V)P(\lambda_{SC} \mid S_C\ S'_C) \\ & P(O_L)P(S'_V\ S'_C \mid O_L) \\ & P(C \mid O_S\ O_L) \end{aligned} \tag{4.9}$$

A graphical network representing this decomposition is shown Figure 4.7.

We do not provide here all the technical details for the precise definition of Eq. (4.9); however, we highlight some of the terms. The auditory system describes the knowledge the agent has of the link between phonetic objects O_L and sensory variables S'_V (formants $F1$ and $F2$ for the vowel) and S'_C ($F2$ and $F3$ for the consonant). This is implemented as 4-D Gaussian probability distributions, the mean vectors and covariance matrices of which are estimated during the learning process. The sensory-motor system is composed of two terms predicting the sensory consequences of motor gestures, $P(S_V \mid M_V)$ and $P(S_C \mid M_C)$, which are also learned Gaussian probability distributions, and a fixed, pre-computed term capturing a support for consonants achievable from each vowel, $P(M_C \mid M_V)$.

The motor variables are subsets of VLAM variables: vowels are three-dimensional (tongue body *TB*, tongue dorsum *TD* and lip height *LH*), and consonants five-dimensional (jaw *J*, tongue body *TB*, tongue dorsum *TD*, tongue apex *Apex* and lip height *LH*). The motor system features a simplified coarticulation model, considering a consonant as a perturbation of a vocalic frame. This is mathematically expressed by explicitly introducing a delta variable describing the perturbation superposed to the vowel to obtain a plosive consonant. The vowel production follows a motor repertoire $P(M'_V \mid O_S)$, encoded as a learned Gaussian model. We assume a

set of primitive consonantal gestures corresponding to the choice of the place of articulation for plosives, in the term $P(G'_C \mid O_S)$: combined jaw and lips for bilabials, combined jaw and tongue apex for alveolars and combined jaw and tongue dorsum for velars. The relationship between this "gesture" vocabulary and the syllable stimuli is not provided, it has to be learned. Finally, the term $P(\Delta'_{MC} \mid M'_V \; G'_C)$ is fixed and computes the combination of a chosen vowel gesture with a chosen articulation place, to yield a vocal tract configuration for the consonant.

4.3.3 Parameter identification in COSMO-Perception

We have developed a learning scenario inspired from the known developmental schedule of infant speech. We proceed in three consecutive steps, such that speech perception capabilities precede speech production capabilities, and that undirected speech production capabilities (e.g., babbling) precede object-oriented speech production capabilities: **L1**: learning of the auditory categories; **L2**: learning of the motor-to-auditory mapping; **L3**: learning the motor repertoire.

In all these stages, the Learning Agent receives from the Master Agent syllable acoustic stimuli, that come from a realistic learning database created using VLAM. For stages **L1** and **L3**, the Learning Agent also receives the corresponding object identities.

As previously, in the context of the indistinguishability theorem, learning stage **L1** consists in supervised learning of the auditory model, that is, the parameters of 4-D Gaussian probability distributions $P(S'_V \; S'_C \mid O_L)$ (the blue part of the schema in Figure 4.7); it is done straightforwardly. On the contrary, stages **L2** and **L3** cannot proceed in a fully supervised manner, as motor information cannot be provided to the Learning Agent.

For learning stage **L2**, we apply instead what we call a "learning by accommodation" algorithm, which can be seen as target-oriented imitation. Consider some point during learning: the Learning Agent, having observed a syllable stimulus $\langle s_v, s_c \rangle$, using its current knowledge of the motor-to-sensory mapping $P(S_V \; S_C \mid M_V \; M_C)$ (the green part of the schema in Figure 4.7), can compute the probability distribution over motor gestures $P(M_V \; M_C \mid [S_V = s_v] \; [S_C = s_c])$, and draw a motor gesture $\langle m_v, m_c \rangle$. The Learning Agent then produces this motor gesture, which results in an acoustic stimulus $\langle s'_v, s'_c \rangle$ (using VLAM as an external vocal tract simulator). The Learning Agent then can use $\langle m_v, m_c, s'_v, s'_c \rangle$ as supervised learning data to identify parameters of its motor-to-sensory mapping.

Such a learning algorithm initially performs as random exploration, because the motor-to-sensory mapping $P(S_V \; S_C \mid M_V \; M_C)$ is initially empty of any knowledge allowing to infer correct motor gestures. However, over time, new observations improve the internal model of the motor-to-sensory transform, which in turns improve the motor inversion for imitation. Imitation gradually improves, so that given a stimulus, the probability to draw a relevant motor gesture improves. Therefore, from initial random exploration, the learning by accommodation algorithm transforms into a target-oriented exploration.

For learning stage **L3**, we fix the parameters of the motor-to-sensory mapping, and then proceed in a similar manner, except that object identity is provided along the acoustic stimulus: the recovery of a motor gesture is driven both by the acoustic stimulus and object identity. Given the drawn motor gesture and object identity, parameters of the motor model (the red part of the schema in Figure 4.7) can then be learned in a supervised manner.

4.3.4 Inference in COSMO-Perception

Similarly to what we summarized in Figure 4.2, Bayesian inference within the COSMO-Perception model allows computing conditional probability distributions that implement the purely motor, purely auditory and perceptuo-motor instances of speech perception. Because of the complexity of COSMO-Perception however, we do not detail the corresponding Bayesian inferences here.

They can still be interpreted exactly as previously: auditory perception is expressed as the direct use of the link between auditory representations and the corresponding object labels, motor perception as the combination of the motor repertoire with an internal model allowing to associate motor and sensory representations, and perceptuo-motor perception as the Bayesian fusion of the auditory and motor categorization processes.

4.3.5 Model variants and their comparison

The indistinguishability result of Section 4.3.1 defined a limit at which purely auditory and purely motor theories of speech perception would become identical processes, whatever their difference in mathematical expression. The learning scenario we defined in COSMO-Perception, along with limiting assumptions in the parametric forms of probability distributions, ensure that limit is not reached. We can therefore compare experimentally purely auditory and purely motor speech perception processes, and study how they differ.

To do so, we have both explored the dynamics of the learning of the components of the model, and relative performance of auditory, motor, and perceptuo-motor perception processes. We summarize the main observations here.

First, we have monitored the evolution of entropies of the auditory and motor components of the COSMO-Perception model. We have observed that the entropy of the auditory model converges quickly to a level close to the entropy of the stimuli produced by the master, whereas the entropy of the motor model converges more slowly. This is a replication of results observed in the one-dimensional variant of the model, suggesting a robust property. It is also easily interpreted: the motor model involves exploration of an internal, high-dimensional space, whereas the auditory classification model directly maps stimuli to object values. In other words, the learning task of the auditory classifier is easier than the one of the motor classifier; it is solved and converges quicker.

Second, we have assessed the performance of motor, auditory, and perceptuo-motor perception processes. We tested them on stimuli produced by the Master Agent, adding a variable amount of noise simulated as a Gaussian distributed perturbation on acoustic stimuli of variable variance. We then presented these noisy stimuli to either perception process, and obtain confusion matrices for object recognition. Average values of diagonals of these confusion matrices provides correct recognition scores as a function of stimulus noise; this is shown Figure 4.8.

We observed that, for clean stimuli, the auditory model performs better than the motor one. When noise is added, the motor system performance decreases less rapidly than the auditory one, and it becomes more accurate for noise levels larger than 0.5. The perceptuo-motor model capitalizes on the fusion of the two branches to provide better scores than the separate auditory and motor models, at all noise levels. Reproduction of these results in the one-dimensional variant of the model confirms their robustness; although precise numerical values of performance vary, although the crossing point between performance of motor and auditory perception varies, as a function of simulated non-linearity between motor and sensory spaces, the overall pattern of results holds.

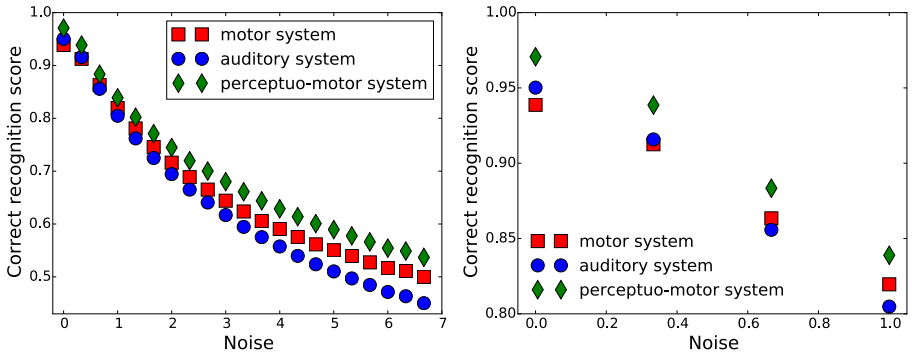

Figure 4.8: **Performance of perception processes for syllables presented at various levels of noise.** Plots display correct recognition scores for the auditory, motor and perceptuo-motor implementations of the perception task in COSMO-Perception. Right plot: a zoom of the left plot for low levels of noise highlights the inversion of performance between the auditory system (better for normal conditions) and the motor system (better for noisy conditions).

We interpret this pattern of results as a "narrow-band" auditory branch vs "wide-band" motor branch property. The auditory system would be able to focus rapidly and precisely on the set of exogenous learning stimuli, leading to a system finely tuned to this learning set. This would provide the auditory system with what we call a "narrow-band" specificity with respect to the learning data.

On the contrary, the motor system would "wander" in the sensory-motor space during its exploration stage, because of the complexity of its learning task. Hence it would evolve more slowly and focus less efficiently on the set of learning stimuli. On the flip side, the exploration stage would enable it to process a wider set of stimuli. This would provide the motor system with what we call a "wide-band" specificity, making it poorer for learned stimuli, but better at generalizing to adverse conditions involving unlearned stimuli.

4.4 Bayesian modeling of speech production: COSMO-Production

Biographical note

Collaborators: Pascal Perrier (GIPSA-Lab, Grenoble), Jean-Luc Schwartz (GIPSA-Lab, Grenoble)

Supervised student: Jean-François Patri (Master 2 and Ph.D., ongoing)

Publications: Patri et al. (in press)

Note: this section adapts material from Patri et al. (in press)

In motor control, the system finds patterns of motor activations to achieve some given motor goal. This is in essence an ill-posed problem, with a multiplicity of possible solutions, since degrees of freedom of articulatory chains often exceed the degrees of freedom of the task. Optimal motor control theories resolve this redundancy problem by using a cost function, and finding the solutions that optimize the value of this cost function. Two situations have to be considered: either there is a single solution to the optimization problem, and an optimal control method always outputs this single solution, or there are several solutions, equivalent with respect to the cost function, and an optimal control method has no principled method for choosing among

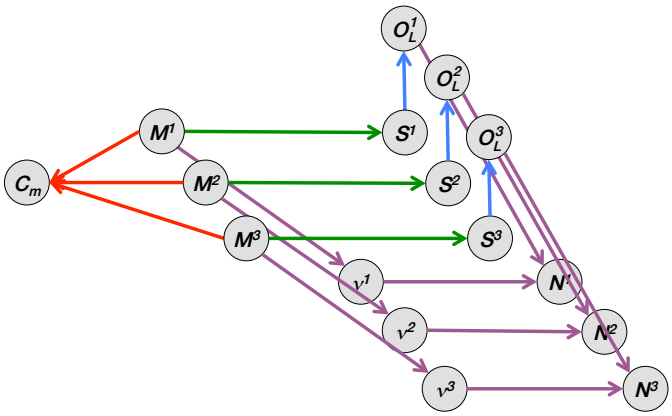

Figure 4.9: **Graphical representation of the structure of the COSMO-Perception model.** In red, the motor system; in green, the sensory-motor system; in blue, the auditory system; in purple, the motor effort system.

them, and instead usually resorts to side-effects of the optimization process, such as outputting the first found solution.

Either way, optimal motor control methods typically generate stereotyped solutions to motor control problems. This sharply contrasts with the observed behaviors of biological agents, where stereotypy and precise reproducibility of movements is the exception, and trial-to-trial variability the norm. Optimal control theories can then be either construed as theories of average patterns of motor control, or it has to be assumed that variability in observed movements are entirely attributed to processes other than trajectory planning, e.g., assuming that movement production is noisy.

In a fully Bayesian model of motor control, however, variability can also be incorporated at the trajectory planning stage in a principled manner, contrary to the optimal control framework. For instance, in speech production, several components of motor planning are commonly assumed to feature redundancy: 1) a particular phonemic goal does not correspond to a unique point in the acoustic domain, since different acoustic signals are perceived as a unique phonemic category, 2) a particular acoustic signal can be produced by different vocal tract configurations, and 3) a particular vocal tract configuration can be attained by different patterns of muscle activation.

By casting optimal control into a fully Bayesian modeling framework, we suggest that both variability and selection of motor control variables in speech production can be obtained in a principled way, from uncertainty at the representational level, and without resorting solely to stochastic noise in the dynamics. We illustrate this approach by presenting a Bayesian reformulation of an optimal control model for speech motor planning of sequence of phonemes, the GEPPETO model (Perrier et al.; 2005).

4.4.1 COSMO-Production model definition

We call the Bayesian rewriting of GEPPETO the COSMO-Production model, for the purpose of the present manuscript. We present it here as a variant of the COSMO general architecture, albeit a partial instantiation of COSMO. Indeed, it lacks both the motor repertoire model $P(M \mid O_S)$ and the communication coherence model $P(C \mid O_S \; O_L)$.

On the other hand, whereas COSMO only considers the acoustic signal resulting of motor commands, COSMO-Production also considers the resulting overall effort, involving variables ν, the total amount of force corresponding to a motor command, and N, a coarse-grained, categorical summary of ν. Finally, whereas COSMO-Perception was applied to CV syllables, COSMO-Production considers the production of sequences of phonemes; we instantiate it in

the case of sequences of 3 phonemes. Variables thus appear annotated by a position index, $i \in \{1, 2, 3\}$; we use the shorthand $X^{1:3} = X^1 \wedge X^2 \wedge X^3$.

The graphical representation of the structure of the COSMO-Perception model is shown Figure 4.9, and the corresponding decomposition of the joint probability distribution is:

$$\begin{aligned} & P(C_m \; M^{1:3} \; S^{1:3} \; O_L^{1:3} \; \nu^{1:3} \; N^{1:3}) \\ = & \prod_i \left[P(M^i)\right] P(C_m \mid M^{1:3}) \prod_i \left[P(S^i \mid M^i) P(O_L^i \mid S^i)\right] \\ & \prod_i \left[P(\nu^i \mid M^i) P(N^i \mid \nu^i \; O_L^i)\right] \; . \end{aligned} \tag{4.10}$$

GEPPETO computes the motor control variables of a bio-mechanical model of the tongue (Payan and Perrier; 1997, Perrier et al.; 2003), based on a finite element structure representing the projection of the tongue on the mid-sagittal plane. Six principal muscles are considered as actuators for shaping the tongue; their activation is specified by λ parameters [2] above which active muscle force is generated. The motor variable of COSMO-Production is thus, as in GEPPETO, $M^i = \lambda_1 \wedge \ldots \wedge \lambda_6$.

The motor model revolves mainly around the term $P(C_m \mid M^{1:3})$: C_m is a binary variable, acting as a switch, being either in the position $C_m = L$ for "Lazy" (corresponding to a minimum effort requirement) or $C_m = H$ for "Hyperactive" (corresponding to its opposite, a "maximum effort" requirement). The "Lazy" mode penalizes sequences of control variables that are far from each other, by attributing them a cost that increases with the perimeter of the triangle that they define in the control space; this is done by:

$$P([C_m = L] \mid M^{1:3}) = e^{-\kappa_M (|M^2 - M^1| + |M^2 - M^3| + |M^3 - M^1|)} \; . \tag{4.11}$$

The κ_M constant tunes the strength of this coarticulation constraint; when κ_M increases, the system is more "lazy" such that it forces the motor commands to be closer together.

As previously in COSMO-Perception, GEPPETO, and thus, COSMO-Production, consider a sensory domain characterized by the first three formants of the acoustic signal, so that $S = F1 \wedge F2 \wedge F3$. Every configuration of the tongue bio-mechanical model generates a deterministically determined acoustic signal. Therefore, for every point in the 6-dimensional control space M there corresponds a unique point in the 3-dimensional acoustic space S. This is learned both in GEPPETO and COSMO-Production by a Radial Basis Function (RBF) network; this network is wrapped as the function $s^*(M)$ of a functional Dirac probability distribution in COSMO-Production: $P(S^i \mid M^i) = \delta_{s^*(M)}(S^i)$.

As GEPPETO only includes an account of the tongue, only phonemes that do not require lip rounding are considered. An additional "no-phoneme" category (denoted by /00/) is further assumed in order to take into account all acoustic configurations that do not fall within any of the modeled phonemic categories, so that $O_L^i = \{/\text{i}/, /\text{e}/, /\varepsilon/, /\text{a}/, /\alpha/, /\text{ɔ}/, /\text{k}/, /00/\}$. GEPPETO assumes that phonemes are characterized as ellipsoid regions of the 3-dimensional acoustic space, the variance of which is multiplied by a κ_S parameter, controlling the required precision in reaching acoustic targets. When κ_S increases, the acoustic tolerance increases, to that reaching acoustic targets is less mandatory; in other words, to increase the force of the acoustic constraint, one has to decrease κ_S (contrary to an increase in κ_M which increases the strength of the motor constraint). In COSMO-Production, acoustic targets are represented, in a sub-model,

[2] Note that λs here are motor dimensions, not coherence variables, contrary to the usual notation in the rest of this manuscript.

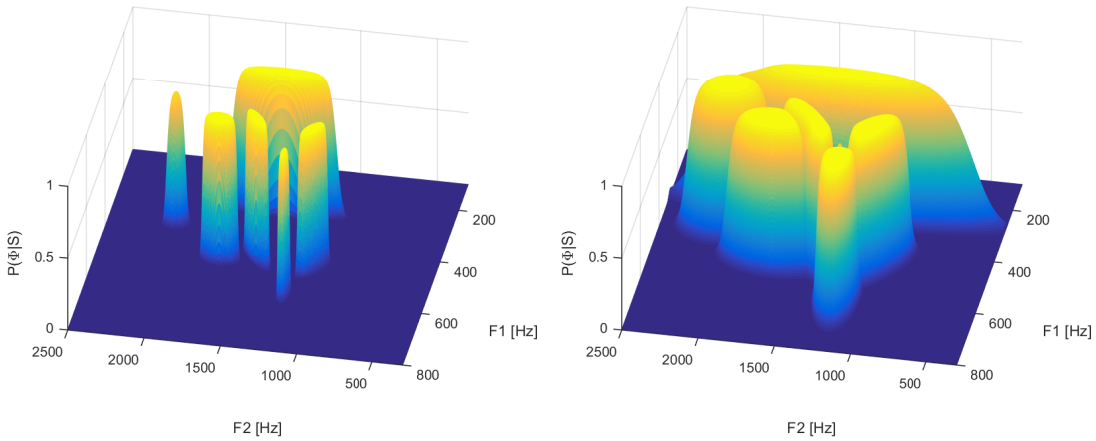

Figure 4.10: **Acoustic categorization in COSMO-Production**, as a function of the κ_S parameter. Plots are the superposition of likelihood functions, for all objects, $P(O_L^i \mid [S^i = s^i])$ (viewed on the $(F1, F2)$ plane, with $F3$ fixed, $F3 = 2,450$ Hz). **Left:** $\kappa_S = 0.3$. **Right:** $\kappa_S = 1$.

by a Gaussian generative model $P(S^i \mid O_L^i)$. This sub-model is then inverted and included in Eq. (4.10) as:

$$P(O_L^i \mid S^i) = \frac{P(S^i \mid O_L^i)P(O_L^i)}{\sum_{O_L^i} P(S^i \mid O_L^i)P(O_L^i)} \ .$$

For illustration, Figure 4.10 presents the resulting likelihood functions $P(O_L^i \mid [S^i = s^i]) = f(s^i)$.

Finally, the motor effort model is similar structurally to the acoustic model; the term $P(\nu^i \mid M^i)$ computes the overall effort for a motor command with a deterministically determined function, approximated by another RBF function $\nu^*(M)$, and the term $P(N^i \mid \nu^i \ O_L^i)$ is a categorization model to define a coarse-grained description of efforts, into three classes, representing "Weak", "Medium" and "Strong" desired level of effort.

4.4.2 Parameter identification in COSMO-Production

In the COSMO-Production, all parameters of the probability distributions of Eq. (4.10) are defined to as to match the parameters of the GEPPETO model. For instance, the acoustic categories, defined as Gaussian models, have parameters copied from GEPPETO, which was itself calibrated by data from the literature, to provide realistic phoneme descriptions.

Overall, the only remaining free parameters are κ_S and κ_M, which respectively modulate the strengths of constraints on acoustic precision and motor coarticulation.

4.4.3 Inference in COSMO-Production

The only inference we consider in COSMO-Production is the one of planning a sequence of motor controls to perform a given phoneme sequence, for a given overall desired force level, and

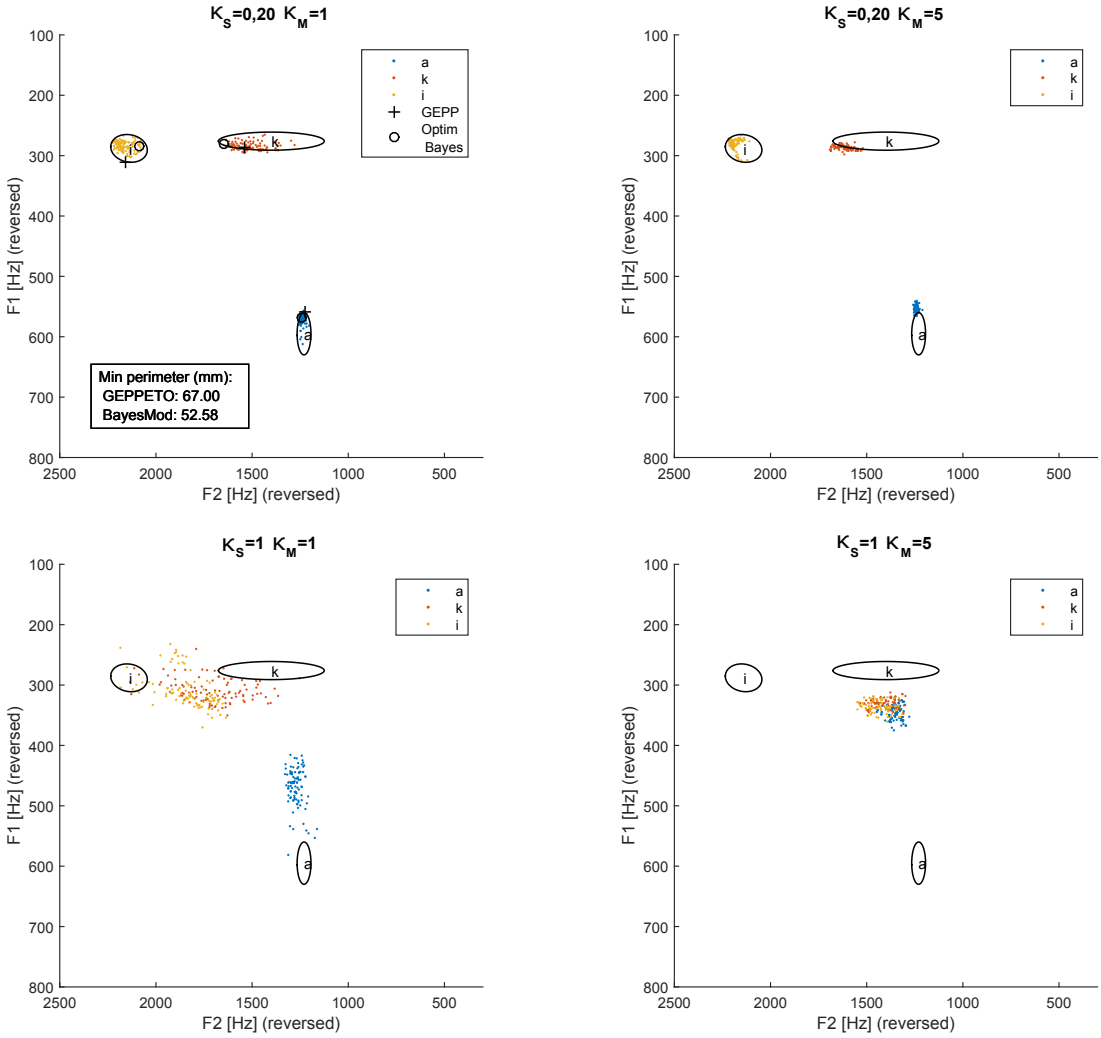

Figure 4.11: **Production of the /aki/ phoneme sequence in COSMO-Production**, as a function of the sensory and motor constraints, respectively piloted by κ_S and κ_M. Dots represent several simulations, illustrating the token-to-token variability of motor planning in COSMO-Production. The top, left plot shows the unique result output by GEPPETO in the same situation, for comparison.

assuming coarticulation effects. The corresponding probabilistic question and inference are:

$$
\begin{aligned}
&P(M^{1:3} \mid O_L^{1:3}\ N^{1:3}\ [C_m = L]) \\
&\quad \propto \sum_{S^{1:3},\nu^{1:3}} P(C_m\ M^{1:3}\ S^{1:3}\ O_L^{1:3}\ \nu^{1:3}\ N^{1:3}) \\
&\quad \propto P([C_m = L] \mid M^{1:3}) \sum_{S^{1:3},\nu^{1:3}} \prod_{i=1:3} \left[P(S^i \mid M^i) P(O_L^i \mid S^i) P(\nu^i \mid M^i) P(N^i \mid \nu^i\ O_L^i)\right] \\
&\quad \propto P([C_m = L] \mid M^{1:3}) \prod_{i=1:3} \left[P(O_L^i \mid s^*(M^i)) P(N^i \mid O_L^i\ \nu^*(M^i))\right] \ . \qquad (4.12)
\end{aligned}
$$

Once this probability distribution is computed, a motor sequence $m^{1:3}$ is drawn at random according to this distribution.

Compared to the COSMO general framework, the inference for production of sequence of phonemes in COSMO-Production can be seen as a realistic instantiation of an auditory theory

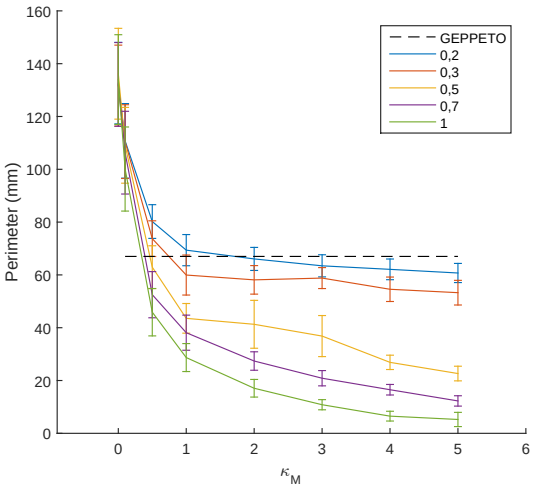

Figure 4.12: **Distances between motor commands of /aki/ sequences in GEPPETO and COSMO-Production**, as a function of the sensory and motor constraints, respectively piloted by κ_S and κ_M. Error bars indicate variability obtained over 100 random samplings. The black horizontal dashed line represents the value obtained with GEPPETO.

of speech production (see Figure 4.2), complemented by a motor effort model. It is however drastically simplified in apparence, because the deterministic functions for the motor-to-sensory and motor-to-force models allow summations over missing variables $\nu^{1:3}$ and $S^{1:3}$ to collapse.

We present results illustrating motor planning of the sequence /aki/, and the balance between acoustic precision and coarticulation. From these results, we evacuate the modeled effort level, for simplicity (i.e., we only consider inferences where the desired effort level is "Medium", throughout the phoneme sequence). Figure 4.11 shows illustrative results for the single phoneme sequence /aki/, and for two possible values for κ_S and κ_M each. Token-to-token variability is clearly seen in successive simulations, as a result of probabilistic modeling, probabilistic inference, and the decision policy made of random sampling on the obtained probability distribution.

4.4.4 Model variants and their comparison

A first result of the comparison between the optimal control model GEPPETO and the Bayesian Algorithmic Model COSMO-Production relates to the form and interpretability of each model. Both models share their assumptions, as we took care to copy those of GEPPETO and use them as is in COSMO-Production. However, they are organized differently.

GEPPETO, being an optimal control model of an articulatory theory of speech production, includes a model of the bio-mechanical plant, relating the motor and sensory dimensions. The rest of hypotheses are cast into the cost function, which thus additively combines the description of acoustic goals and of motor coarticulation constraints. Such an additivity implies a relative scaling of hypotheses, in a difficult to control and interpret mathematical space; it also implies an "averaging" of constraints, which is a combination model among many others; finally, the scalability of such a combination model is unknown.

In contrast, in COSMO-Production, the acoustic goal constraint is expressed in the auditory models $P(S^i \mid O^i_L)$, and the motor coarticulation constraint in the motor model $P(C_m \mid M^{1:3})$. They are expressed separately, and combined by the joint probability distribution and call to the auditory sub-model. We believe this makes these hypotheses more easily managed, and it forces the expression, and thus, interpretation, of the combination model.

In order to further compare the optimal control model GEPPETO and the Bayesian Algorithmic Model COSMO-Production, we have studied their performance, with respect to distances between points of a motor trajectory planned by both models. In COSMO-Production, since

token-to-token variability is an intrinsic property of the model, we present average values over several simulations, and compare them with the unique distance of the stereotyped solution provided by GEPPETO. This is shown Figure 4.12, again for /aki/ sequences.

This illustrates experimentally that GEPPETO appears "contained" as a special case of COSMO-Production. The behaviors of both models closely match when $\kappa_S = 0.2, \kappa_M = 1$; more precisely, the average behavior of COSMO-Production is similar to the stereotyped behavior of GEPPETO. This result can actually be made formal and demonstrated mathematically (Patri et al.; in press), but we do not detail this here.

Finally, consider a slight variant of COSMO-Production, that, instead of drawing motor commands at random according to $P(M^{1:3} \mid O_L^{1:3}\ N^{1:3}\ [C_m = L])$, would select the average value of that distribution, thus selecting a fixed set of motor commands. This variant of COSMO-Production would shed token-to-token variability and become equivalent to, and thus indistinguishable from GEPPETO. This implies that there would be a formal equivalence between an optimal control model and such a Bayesian Algorithmic Model, whatever their conceptual difference. This suggests that optimal control theory then cannot be construed, as is, as a mechanistic theory of motor control. Indeed, there is no plausibility to loss functions and optimization processes as likely neurobiological constructs if models with loss functions and optimization processes can be formally translated into other frameworks with no reference to these concepts. We come back to this idea of framework equivalence and model indistinguishability, and the problem it poses for interpreting models in Cognitive Science, in the next Chapter.

4.5 Discussion

In this Chapter, we have presented a general framework, the COSMO architecture, that casts major trends of theories of speech perception and production into a unified mathematical framework. When instantiated, this architecture allows systematic comparison of these theories, with respect to their ability to account for experimental data. For instance, it appears that agents performing purely motor speech production would not be able to converge to a common communication code. It also appears that it is impossible, in some learning cases, to distinguish between purely auditory and motor theories of speech perception. Furthermore, when these indistinguishability conditions are not met, auditory and motor theories of speech perception appear complementary: a purely auditory process performs better on normal stimuli and worse on adverse stimuli, when compared to a motor process of speech perception. Finally, we showed that a COSMO instantiation of an auditory theory of speech production would allow casting into a different framework, and generalizing, an existing optimal control model of speech production.

We close by mentioning two main directions for future work, related to the two ongoing Ph.D. theses of Marie-Lou Barnaud and Jean-François Patri, respectively concerned with extending COSMO-Perception and COSMO-Production.

In Marie-Lou's work, we first explore how idiosyncratic perception and production processes can be acquired during a realistic speech acquisition learning scenario, and whether they would be correlated. Second, we study how COSMO-Perception could be extended to accommodate both phonetic and syllabic representations of words, and whether they would be experimentally distinguishable.

In Jean-François' work, we explore the nature of speech targets, *i.e.* whether they are acoustic, oro-sensory, or multimodal, and how they could be combined with motor constraints. We are currently designing an extension of COSMO-Production in this direction, with the aim to

assert whether and how, once again, different accounts of speech targets would yield different experimental predictions, and study their functional complementarity.

CHAPTER 5

Bayesian modeling of cognition or modeling Bayesian cognition?

They say that "interesting people talk about ideas, average people talk about other people, boring people talk about themselves" [1]. Unfortunately for the reader, I have so far mostly written about my colleagues and myself; now is the point when I try to "climb a step" of the quote, and write about our work in relation to other people's works. By doing so, I hope to help elucidate ideas and concepts, such as the notion of Bayesian modeling. In other words, we now propose a discussion of our particular approach of Bayesian Algorithmic Modeling, with the aim to placing it in the current panorama of cognitive science.

I begin by an anecdote, that took place while I was a Ph.D. student. I was intrigued by the possibility to apply Bayesian Programming, not as an engineering tool to develop robotic applications, but as a scientific method to investigate natural cognitive systems. At some point however, I realized that I also knew, from a consequence of Cox's theorem (Jaynes; 1988, 2003, van Horn; 2003), that any computable function could be cast as a probabilistic model. This would mean that the Bayesian modeling framework, as a whole, would not be refutable, and thus could not be construed as a scientific theory. Trying to prove that a cognitive system "is" Bayesian would be a dead-end. Instead, Bayesian modeling would merely be a language to describe what cognition does. Our models would be Bayesian, but cognition might not be. We would be doing "Bayesian models of cognition".

I immediately rushed to the office of my Ph.D. advisor, Pierre Bessière, and presented to him my simple, and somewhat frustrating conclusion. To my surprise, he did not agree at all. His reasoning was as follows: if cognitive systems can be elegantly and efficiently described in the Bayesian language, whatever the scale of scrutiny (from neurons to behavior), whatever the observed sub-system (from perception processes, to motor control, language manipulation, abstract reasoning, etc.), then, as a whole, this scientific endeavor would yield credence to the hypothesis that, somehow, the cognitive systems would "use" probabilities, and "be" Bayesian. We would be doing "models of Bayesian cognition".

From my point of view, then, proving the brain to be Bayesian would be impossible, contrary

[1] At least when they do not remember correctly the quote, attributed to Eleanor Roosevelt, who presumably declared "Great minds discuss ideas, average minds discuss events, small minds discuss people".

to Pierre's point of view. It appears impossible for us to be both right simultaneously, and nevertheless, we will try in this Chapter to show how both positions can be reconciled. In order to do so however, we need to elucidate them, and broaden the scope of the discussion, by analyzing what it is exactly that we and other people are doing, when we do Bayesian modeling.

This anecdote is not out-of-the-blue navel-gazing; it turns out to be relevant to a vigorous debate that exploded in recent literature (McClelland et al.; 2010, Griffiths et al.; 2010, Elqayam and Evans; 2011, Jones and Love; 2011, Bowers and Davis; 2012a, Griffiths et al.; 2012a, Bowers and Davis; 2012b, Endress; 2013, Frank; 2013, Marcus and Davis; 2013, Endress; 2014, Hahn; 2014, Zednik and Jäkel; 2014, Goodman et al.; 2015). (Notice that all these references are less than five-year old.)

In a nutshell, one side of this debate groups proponents of Bayesian modeling, and more specifically of Bayesian computational modeling. The other side casts doubt on the interest in proposing such Bayesian models, as they would be theoretically unconstrained and thus, uninformative. Because our approach to Bayesian modeling is somewhat atypical, with Bayesian robot programming as our historical starting point, then it will appear that our stance in this debate is singular, as well. We summarize it now, in broad strokes to be elucidated, hopefully, by the remainder of this Chapter.

In Section 5.1, we recall Marr's analysis of the types of models in cognitive science, the definition of its three levels of computational, algorithmic and implementation level models, and introduce terminology and concepts such as "rationality" and "optimality". We argue that there is a porous frontier between the computational and algorithmic levels, which is detrimental to the clarity of the debate. Therefore, we propose a stricter distinction between the computational and algorithmic levels.

In Section 5.2, we discuss the Bayesian models in cognitive science. We argue that there is an unfortunate confusion between computational and Bayesian modeling. Computational models are focused on rational descriptions of observed behaviors; here, Bayesian modeling offers one mathematical definition of optimality. We show that many Bayesian models, however, are somehow in a "gray area" between computational and algorithmic level models, and seldom presented as such. The debate is thus made all the more difficult.

Section 5.3 and 5.4 return to the debate about the explanatory power of Bayesian models in Cognitive Science. We first argue that, as a whole, the Bayesian modeling framework cannot be construed as a psychological theory, as it is unconstrained, contrary to a collection of algorithmic Bayesian models, which can be suitably constrained to be testable theories. Indeed, computational level models do not allow to explore psychological representations and processes, contrary to algorithmic level models. We thus further argue that a model cannot simultaneously claim Bayesian optimality and be a psychologically constrained model. This would help separate Bayesian-computational models, which are the bulk of the literature, from Bayesian Algorithmic modeling, as we practice it. Since, in Bayesian Algorithmic modeling, rationality is not a defining feature, we then describe how to carefully constrain such models, in order for them to help exploring hypotheses about psychological constructs and processes.

We conclude, in Section 5.5, that a large body of such models would yield credence to the Bayesian brain hypothesis, following an argument of reification from accumulation. Unfortunately, such an argument could increase the plausibility that cognition somehow is Bayesian, but could never settle this question which, ultimately, we show to be beyond the scope of scientific scrutiny.

5.1 Marr's hierarchy of models in cognitive science

Marr's seminal proposition was to identify three levels of analysis of a cognitive systems, the *computational*, *algorithmic and representational* and *implementation* levels (Marr and Poggio; 1976, Poggio; 1981, Marr; 1982). We quote a concise, recent and consensual formulation of their definitions (Peebles and Cooper; 2015, pp. 187–188, italics in the original):

> [...] information-processing systems can be analyzed in terms of the problems that they solve (Marr's *computational* level), the representations and processes by which they solve them (the *algorithmic and representational* level), and the physical instantiation of these representations and processes (the *implementation* level) [...]

The *algorithmic and representational* level is often referred to as, more simply, the *algorithmic* level, and we will follow this terminology in the remainder. The hierarchy itself is sometimes referred to as the "Marr-Poggio hierarchy", or the "Tri-Level Hypothesis"; we will use the simpler terminology with "Marr's hierarchy".

According to Marr, the layers of the hierarchy are at most loosely coupled, and maybe even independent levels of analysis (Sun; 2008). Marr advocated building models by a top-down traversal of these levels, arguing that it is easier to understand a cognitive system starting from its goal than from its extremely complex neurobiological implementation (Marr; 1982, p. 27):

> [...] trying to understand perception by studying only neurons is like trying to understand bird flight by studying only feathers: It just cannot be done. In order to understand bird flight, we have to understand aerodynamics; only then do the structure of feathers and different shapes of birds' wings make sense.

When applying such a distinction to Artificial Intelligence, Marr outlined his preference for methods that solve problems instead of methods that describe mechanisms (Marr; 1977).

Let us note, finally, that the computational level appeared to Marr as under-represented, at least in the domain of the study of visual perception, at the time (during the 70's). This is somewhat at odds with the current situation, where some authors now see a drought of algorithmic models in cognitive science (Peebles and Cooper; 2015, Love; 2015). This particular point deserves attention; we will return to it later.

5.1.1 The three levels of Marr's hierarchy

The easiest level to apprehend, and it appears, the less controversial, is the implementation level. It is clearly delineated from the other two in many aspects, from the wetware nature of the object of study (the brain of a specific cognitive agent, or the average brain of a species), to its scale (from the properties of ion channels to, at most, the connectome architecture (Sporns et al.; 2005)), and to the method of investigation (neuroimaging, intracranial recordings, etc.). This clearly corresponds to the scientific domain of neuroscience, in general.

As an example, consider color perception. An implementation level model would try to mimic the known properties of light receptors of the retina, their spectral sensitivities and transfer functions, the different types of cells in the retina and their projections, etc. This description could be fine-grained, modeling precisely architectures of retinal cells and their response properties, or, abstracting away dimensions such as membrane potential and neurotransmitters, be more coarse-grained, using notions such as center-surround receptive fields.

The algorithmic level is also rather well defined. At this level, assumptions about the representations and the algorithms manipulating them are investigated. This allows to model how some cognitive agent solves a particular task, under known representational and computational constraints.

Let us pursue our example of color perception: a classical theory understands early visual processes as segregating color information into two opponent channels, which are red-green and yellow-blue channels (along with a black-white, luminosity channel). An algorithmic account of color perception would include such a format for representing colors. It would have to describe how retinal inputs map to the opponent-channel space (not necessarily copying to the letter retinal projection architecture, otherwise it would tend towards an implementation level account). It would also consider how opponent channels could be combined, downstream, in order to reconstruct a unified perceptual color space, whether this combination is actually necessary to account for subjective color naming, etc.

Finally, the last level of Marr's hierarchy is the computational level. Unfortunately, this level is less-well defined, and, as a natural consequence, appears controversial and is much discussed in the literature. We therefore introduce three different possible definitions corresponding to three acceptations of this level.

The first of these definitions is to focus exclusively on modeling what it is that a cognitive system is doing; assuming its goal is to solve a well-identified task, this amounts to modeling the task. An extreme view of this level thus concerns mathematical descriptions of *how* to solve a task; in this sense, the cognitive agent is evacuated from the picture. The object of study becomes only the task, such as color perception, or playing chess.

A companion question, in this level, is to model *why* a task is solved. Being extremely wary of scientific questions that begin by "why", I propose to rephrase this, instead, in somewhat more palatable terms, asking how solving a task is beneficial to an organism in an environment. This becomes interesting, for instance, when cast in an evolutionary perspective, as it requires a viable model of fitness and adaptation (J. R. Anderson; 1991, B. L. Anderson; 2015). It allows to infer the cognitive agent's goal.

A second definition of the computational level of Marr's hierarchy considers that it contains models of how cognition should be, as opposed to how it is. This is a normative vs. descriptive distinction [2]. Normative accounts of cognition are interesting in applied cognitive science, providing goals to remediation, education of thinking, or skepticism programs. Beyond this, unfortunately, they do not bring much to pragmatic, theoretical cognitive science, as inferring what cognition is from what it should be is problematic (a.k.a., the "ought-to-is" issue).

The third and final definition of Marr's computational level stills considers what a cognitive system is doing, but makes more precise the observed system; it considers behaviors, defined as input-output relationships. This can serve two purposes. Given some experimentally observed input-output relationship, then a computational-level model of this mapping would either try to identify what it is that this transfer function is computing or, alternatively, would try to reproduce mathematically this transfer function. We believe this last understanding of the computational level of the hierarchy to be the most interesting, and to be closer to the initial definition of Marr (Marr; 1982). It is this sense that we consider in the rest of this discussion.

To pursue and conclude our introductory example of color perception, this would correspond to measuring some mapping between stimuli and perceived colors in subjects, and characterizing

[2] Or, in the more precise terminology of Elqayam and Evans (2011), "prescriptive normativism" vs "empirical normativism".

it mathematically. Doing so would help, for instance, estimating the fidelity of color perception as a function of spectral properties of surfaces, comparing color perception in environments with and without reflections from neighboring objects, comparing color perception in broad daylight and at dusk, etc. All these questions can be answered by considering the experimentally observed input-outputs; to *describe* how the system performs, no knowledge of neuroscience or of the retina is required. To *explain* how the system performs, however, would require mechanistic models.

Whatever the definition of Marr's computational level however, a common denominator is the notion of rationality (Oaksford and Chater; 1998, Chater and Oaksford; 1999, Gershman et al.; 2015) [3]. That is evident in the normative acceptation of the computational level, as it describes a golden-standard for the way to reason and solve a cognitive task. This is also the case in the descriptive acceptation of the computational level; either to verify whether some observed input-output mapping does corresponds to the golden-standard mapping, or, assuming it does, trying to formally define the rationality criterion that corresponds.

Because of the historical separation, in cognitive science, between "high-level cognition" that reasons, and "low-level cognition" that perceives and acts, vocabulary is varied and somewhat troublesome. Rationality being naturally attached to reasoning [4], it is seldom attributed to perception and action; instead, one defines "ideal observer" (Kersten and Yuille; 2003, Clayards et al.; 2008), "ideal navigator" (Stankiewicz et al.; 2005), "ideal listener" (Sonderegger and Yu; 2010), "ideal adapter" models (Kleinschmidt and Jaeger; 2015), "ideal actor" models (Braun et al.; 2010, Genewein et al.; 2015), etc, with a shared focus on the golden-standard for observing, navigating, adapting speech perception, motor control, etc.

When mathematically defined, this rationality concept becomes an optimality criterion. In general, when the considered framework is probability calculus, the usual terminology refers to "statistical optimality". A domain in which this approach is well established, along with a specific terminology, is optimal motor control. Whether it is expressed in probabilistic terms, such as in Bayesian decision theory and minimum variance models, or not, the crux of optimal motor control is the cost or reward function to be optimized.

5.1.2 Delineation between Marr's computational and algorithmic levels

Whereas the identification of the implementation level and its frontier with the algorithmic level appear straightforward, it is unfortunately not so between the computational and algorithmic levels. Indeed, their delineation appears unclear and somewhat porous. This means that some models appear to stand halfway between the computational and algorithmic models, or, alternatively, they appear to be somehow both computational and algorithmic.

We discuss two issues that, unfortunately, can make identifying whether a given model is computational or algorithmic complicated. The first concerns the considered output, the other concerns the problem of simplifying assumptions.

To explain the first issue, we go back to the original writings of Marr. Indeed, Marr considers the problem of mathematically describing an input-output mapping, but was not very specific concerning constraints on the output of this mapping. Recall that Marr was concerned with human visual processing. In Marr's exposition, it is clearly stated that identifying the input is

[3] During redaction of this Chapter, I came upon this very apropos quote by Oscar Wilde: "Man is a rational animal who always loses his temper when he is called upon to act in accordance with the dictates of reason."

[4] "Rational" and "reason" share their etymological origin, along with "ratio"; reasoning and computing are close cousins, which probably explains why the analogy of cognitive systems as information processing and computing devices is firmly rooted.

Table 5.1: Anderson's methods for developing a rational cognitive theory (reprinted from J. R. Anderson (1991)).

1. Specify precisely the goals of the cognitive system
2. Develop a formal model of the environment to which the system is adapted
3. Make the minimal assumptions about computational limitations
4. Derive the optimal behavioral function given 1–3 above
5. Examine the empirical literature to see whether the predictions of the behavioral function are confirmed
6. Repeat, iteratively refining the theory

straightforward, as it simply is the retina. However, identifying the output of visual processing is not so easy (Marr; 1982, p. 31); it appears that it is not a behavioral output, but some representation, internal to the overall chain of visual processes. This is a slippery slope towards stating hypotheses about representations, which could be argued to belong instead to the algorithmic level.

It is possible to select some input and some output spaces in order to ensure that no internal representation is considered; that is when the input is sensory, and when the output is behavioral (or motor). That was the constraint on the behaviorist program, that would only consider observable events, only measure physical variables external to the cognitive system. In that sense, Marr's suggestion to consider, as the output of the perception process, some internal, perceptive representation, is problematic: this assumes the existence of such a representation. Can a computational model of this form still be considered computational, or does it become "a bit" algorithmic and representational?

To explain the second issue, concerning simplifying assumptions in computational models, we consider Anderson's well-known and widely cited rational analysis (J. R. Anderson; 1991), which we recall Table 5.1. Note that it contains an iterative loop, that gradually refines a model by adding assumptions, until it conforms to experimental data. Consider two situations. Imagine first developing some cognitive model, using only a single pass through this iterative loop: constraining assumptions will be few, maybe even non-existent, in the resulting model. Imagine instead that, to obtain a satisfactory model, a hundred passes through the loop are required: it would mean that the original "rational" model would have needed much constraining to adequately fit data. Is the resulting model still optimal? Is it still computational? In that extreme case, we would argue it is not. Instead, it is now filled with simplifying assumptions, hypotheses about representations, etc, which make the output of Anderson's method an algorithmic model.

If the start of Anderson's method is a computational model, and if its end point is an algorithmic model, is there a clear delineation point where the model under study suddenly stops being computational and becomes algorithmic? We see no way this change could be considered an identifiable event; instead, it appears that a model becomes gradually less computational, and more algorithmic, as assumptions are added.

Some assumptions make a model take large steps toward the algorithmic level, some only incrementally make it less computational. Mathematically quantifying this deviation from optimality surely is feasible. Concerning probabilistic models, for instance, mutual information or cross-entropy could be viable measures for this task; whether this already was explored in the

literature, in this context of Marr's classification of models, is unknown to us.

This raises the question of the status of the concept of "rationality" in highly constrained model: whether it is still a useful concept in this case is an open issue. A solution would be to set the "event", the departure from optimality, at zero: a model would either be totally unconstrained and computational (and rational, optimal), or it would involve simplifying assumptions about representations and processes and be algorithmic (without useful reference to concept such as rationality and optimality). Such a strict delineation appears contrary to Anderson's proposal, and is further discussed now.

5.1.3 Alternatives to Marr's hierarchy

Our proposal: a strict reading of Marr's hierarchy

In our view, we would prefer Marr's hierarchy to include some sort of exclusivity principle: we propose that a given model should not be simultaneously computational and algorithmic. We believe this can be achieved with a strict reading of Marr's hierarchy.

We would argue that computational models should stick with optimal descriptions of input-output mappings, preferably set between sensor and motor responses. This would make such models surface descriptions only, and not mechanistic descriptions: they would help inferring the task that is solved by this input-output mapping ("study the goal of the system and why it does what it does", in the words of Poggio (1981)), or, in reverse, assuming the task known, they would restrict their focus on the notion of optimality. They would also be closer to the original aims of behaviorism. The question they would address is whether an observed behavior is, or not, optimal, for a precisely given definition of optimality. Alternatively, assuming optimality, an observed behavior could help precise the possible optimality criteria of the system. None of these research programs would be able to explore representations and processes.

Instead, to do so, one would have to resort to algorithmic models, where questions about representations and processes are valid. The methodology then would be based not on optimality notions, but on model adequacy to data, on model parsimony, etc., that is to say, the standard fare of model evaluation in the scientific method. Here, we would strive to make all assumptions in a model explicit, interpretable, and possibly inspired from either the implementation or the computational levels.

This would make Marr's hierarchy segregate models in a strict manner, but a very classical one. In a way, it could be seen as formalizing the usual approaches to cognition. For instance, in this strict view, the computational level would be distinct from the algorithmic level in a way that is similar to how behaviorism is separated from cognitive psychology. These two approaches to psychology are themselves, in effect, separated from neuroscience, which considers mostly the implementation level.

Also, and, to make a point probably less rehashed, we find that such a strict distinction between the computational and algorithmic levels would mirror approaches to Artificial Intelligence. Indeed, AI is either construed as the quest to create a perfect cognitive agent, or a human-like (or animal-like) cognitive agent (Russell and Norvig; 1995). In both Artificial Intelligence and Cognitive Science, that is, independently of whether artificial or natural cognition is considered, this would distinguish between "agents that are" and "agents that could be", delineating in effect complementary, and not opposing, approaches.

Another analogy with computer science, and more precisely computer programming and computer programming languages, can be proposed. In a sense, since computational level models,

in a strict reading of Marr's hierarchy, are only concerned with the description of input-output relationships, they can be thought of as "formal specifications" (Love; 2015). On the contrary, algorithmic and implementation level models are mechanistic, but at different granularities. In programming terms, an algorithmic model would correspond to a program written in a high-level language, whereas an implementation level model would correspond to a program written in machine language.

Other proposals: amending or extending Marr's hierarchy

In our analysis, we concluded there was a porous delineation between the computational and algorithmic levels of Marr's hierarchy. On similar grounds, Griffiths et al. (2015) have proposed to amend Marr's hierarchy, adding an intermediary level. They consider ressource limitations that constrain the inference process; rational models constrained in this manner would fall in what they call the "rational-process" level.

In this proposed layer, and, anticipating somewhat our forthcoming discussion of probabilistic models of cognition, are models that are inspired from modern computer implementation of approximate Bayesian inference. Computer implementation and cognitive agents are both constrained by limited computation time; scientists have devised efficient approximation tools for probabilistic inference, such as Markov Chain Monte Carlo techniques (MCMC), particle filters, and so on. The gist of these methods is, instead of fully representing probability distributions, which can be cumbersome, they approximate them using samples. An advantage is the notion of any-time computation, that is to say, approximations are coarse when time constraints are drastic, and any additional time can be devoted to refining the initial approximations. It might be the case that probability distributions are efficiently represented using similar tricks in natural cognitive agents (Shi and Griffiths; 2009, Sanborn et al.; 2010, Shi et al.; 2010, Griffiths et al.; 2012b).

Another approach that would also yield models at the proposed rational-process level is the resource-rational analysis: a rational agent could take into account its own processing power limits, its bounds on computation time, to allocate its "computation" resources, possibly in an efficient, and maybe even rational manner (Vul et al.; 2014).

For others, the fact that computational and algorithmic models sometimes overlap is interpreted as calling for a more thorough exploration of multi-level analyses (Cooper and Peebles; 2015), to better separate which model components contribute to each level.

Finally, we note that Marr's hierarchy, on which our analysis was centered because it is arguably the most influential, is of course not the only taxonomy of models of cognitive science that was proposed in the literature. Classical alternatives include Marr's earlier distinction of Type 1 and Type 2 theories (Marr; 1977), Chomsky's distinction of competence and performance, Pylyshyn's four-level hierarchy, Newell's three-level hierarchy, etc. Historical references, introductory material, and a discussion of how they relate to Marr's hierarchy is provided by J. R. Anderson (1990, Ch. 1).

We note another proposal (Sun et al.; 2005, Sun; 2008), that basically consists in extending Marr's hierarchy "from above": because Marr's hierarchy is only concerned with models of a single cognitive agent, all other agents formally belong to the "environment" of this agent, and should be modeled as external constraints of the considered agent. But many cognitive agents definitely have a social dimension to them, which motivates considering a fourth level, the "sociological" level. In this level, inter-agent processes are modeled, such as collective behaviors, cultural processes, etc.

We outline an intriguing potential extension of this idea. A particular case of possible interaction between agents is when of these is a statistician observing another agent, as in experimental psychology. More generally, consider a scientist. A scientist, hopefully, is a cognitive agent. It has sensory inputs in the form of experimental data, it has behavioral output when it selects the next experiment to perform, it certainly has internal representations of the observed phenomenon, in the form of classes of models or theoretical frameworks.

The process of scientific enquiry could then be formalized, maybe as part of the "sociological" level of Sun's hierarchy Sun et al. (2005) concerning the modeling of the cognitive scientist, or maybe as a specific level altogether to also capture the cognition of the physicist. Mathematical models at this level would encompass all of statistics, including experimental design and Bayesian statistics [5], when the statisticians' priors are taken into account. We provide some entry points into the recent flurry of literature and debates on the application of Bayesian statistical methods to experimental psychology (Lee; 2008, Andrews and Baguley; 2013, Gelman and Shalizi; 2013, Kruschke; 2015).

A striking feature of mathematical models at this level is that they are almost exclusively cast in the probabilistic language. It is an exciting goal to consider mathematical continuity between modeling cognition and modeling cognition of the agent that observes cognition; thus we find ourselves naturally drawn towards Bayesian statistics, Bayesian computational modeling, Bayesian algorithmic modeling and Bayesian neuroscience. This continuity already has been discussed (Kording; 2014, Kruschke; 2011), and makes sense historically, as Bayes' theorem originally comes from normative considerations about updating knowledge in the face of new evidence (Fienberg; 2006).

5.2 Bayesian models in Marr's hierarchy

Providing an exhaustive list of Bayesian models in cognitive science would be a daunting challenge, well beyond the scope of the current manuscript. It would be even more so if we expanded the vocabulary, to consider any kind of probabilistic model. Indeed, most of psychophysics, for instance, would then have to be referenced. The reader interested in reviews and tutorial introductions will find mostly recent entry points elsewhere (Kersten et al.; 2004, Chater and Manning; 2006, Wolpert; 2007, Griffiths et al.; 2008, Chater et al.; 2010, Colas et al.; 2010, Jacobs and Kruschke; 2010, Holyoak and Cheng; 2011, Jones and Love; 2011, Perfors et al.; 2011, Ma; 2012, Norris; 2013, Hahn; 2014, Gopnik and Bonawitz; 2015, Vincent; 2015a,b).

Instead of a thorough review, we will aim for sampling the current variety of Bayesian modeling approaches in order to correctly expose the central point of contention in the current debate about the explanatory power of Bayesian models in cognitive science, that we introduced above.

This debate being concerned with the relevance of the Bayesian approach to the understanding of psychological constructs, it is mostly focused on the computational and algorithmic level models. Therefore, we purposefully set aside Bayesian accounts at the implementation level of Marr's hierarchy. Nevertheless, we still provide a few entry points to the probabilistic neuroscience literature, sometimes referred to as the "Bayesian brain hypothesis" (Friston; 2010a,

[5] A note here about Bayesian statistics. The working title of this *habilitation* was "Bayesian comparison of Bayesian modeling of Bayesian cognition", which was nifty. Unfortunately, it would have required a thorough treatment of at least Bayesian model comparison, and possibly Bayesian statistics. The first is a subject I am interested in (and in which I have dabbled, see Section A.6), but was ultimately considered out of the scope of the manuscript. The second is a subject I am curious about, but would require first becoming versed in classical, frequentist statistics to really understand.

2012). We note that neural populations might implement probability distributions and the product rule (see recent reviews (Pouget et al.; 2013, Ma and Jazayeri; 2014)), lateral inhibitions might implement the normalization rule, itself useful for implementing the marginalization rule (Beck et al.; 2011), and that diffusion-drift to threshold in neural populations might implement bayesian evidence accumulation and a decision strategy (see a recent introduction (Drugowitsch and Pouget; 2012)).

We also acknowledge the now famous free-energy principle, whereby the architecture of cortical macro-columns might compute probability distributions about predicted sensory events (Friston et al.; 2006, Friston and Kiebel; 2009), and that hypothesizes an explanation of this local structure as, in effect, minimizing the free-energy of the system [6].

5.2.1 Bayesian modeling at the computational level

For many authors, most, if not all, of Bayesian models in cognitive science reside at the computational level of Marr's hierarchy. Indeed, many equate Bayesian modeling with rational or optimal modeling, in a probabilistic setting (Griffiths et al.; 2008, Oaksford and Chater; 2009, Jacobs and Kruschke; 2010, Eberhardt and Danks; 2011, Griffiths et al.; 2012b, 2015).

A few handy examples are computational Bayesian models of speech perception (Norris and McQueen; 2008, Clayards et al.; 2008, Sonderegger and Yu; 2010, Kleinschmidt and Jaeger; 2015), a rational model of speech perception (Feldman et al.; 2009), models of visual perception (Szeliski; 1990, Weiss et al.; 2002, Kersten et al.; 2004), such as color perception (Brainard and Freeman; 1997, Schrater and Kersten; 2000), models of motor control (Harris and Wolpert; 1998, Jordan and Wolpert; 1999, Wolpert and Ghahramani; 2000, Todorov; 2004), along with many other references in this Chapter.

A note concerning terminology is needed here. Of course, ours differs, as we use the term "Bayesian" to refer to the subjectivist "Bayesianism" approach to probability, as opposed to the frequentist definition of probability. Other, more technical meanings of the term "Bayesian" concern the use of informative prior probability distributions, as opposed to maximum likelihood estimation, or the fact of treating model parameters as random variables. These meanings are well-established and widespread in the physics and statistical literature (Fienberg; 2006). All these various definitions appear unfortunately legitimate, well-anchored historically, albeit in different domains, and without obvious alternatives. Enhancing the general awareness of the ambiguity of the term would go a long way towards clarifying the status of "Bayesian" modeling in the scientific community.

Bayesian models at the computational level aim at describing input-output mappings in probabilistic terms. Consider the classical example where a perception process is studied, from a stimulus space S to a percept space P. In a probabilistic setting, the most general description of the relation between S and P is the joint probability distribution $P(S\,P)$. This can be decomposed using the product rule into either $P(S \mid P)P(P)$ or into $P(P \mid S)P(S)$. The three expressions above are mathematically equal, and thus, they are formally indistinguishable.

Following the classical theory of perception, that considers the process of perception as the inversion of a generative model (Marroquin et al.; 1987, Bertero et al.; 1988, Knill and Richards; 1996, Kersten and Yuille; 2003, Kersten et al.; 2004, Yuille and Kersten; 2006), Bayesian models

[6] An uncommon and laudable effort to bridge the gap between implementation and computational level accounts; in our view, it unfortunately misses the algorithmic component, i.e., it does not bring much towards understanding psychological level processes and representations.

usually focus on the first decomposition:

$$P(S\ P) = P(S \mid P)P(P)\ . \tag{5.1}$$

Indeed, with this decomposition, a perception process can be modeled by the question $P(P \mid S)$, and computed from Eq. (5.1) by Bayesian inference:

$$P(P \mid S) = \frac{P(S\ P)}{P(S)} = \frac{P(S \mid P)P(P)}{\sum_P P(S \mid P)P(P)} \propto P(S \mid P)P(P)\ . \tag{5.2}$$

In these equations, $P(P)$ is called the *prior distribution* over percepts. $P(S \mid P)$ is the generative model that predicts stimuli, given an assumed percept. When considered as a function of percepts P, it is called a *likelihood function* [7] $f(P) = P(S \mid P)$. $P(S)$ in Eq. (5.2) is called the *evidence*, and most often dismissed and considered as a normalization constant. Finally, $P(P \mid S)$ is the *posterior distribution* over percepts.

Considering a decomposition of the joint as a product of a prior and a likelihood function is just a notation trick. Because of the mathematical equality in Bayes' theorem, both are different expressions of the same mathematical object. As Fiorillo (2012) puts it:

> [Bayes' theorem] does not specify any transformation of information, it only expresses the same information in alternate forms. In this regard, [Bayes' theorem] is fully analogous to a statement of the form $12 = 3 \times 4 = 24/2$, etc.

Or, paraphrasing: "A product of prior and likelihood is no more a model of the joint distribution than 3×4 is a model of 12". In that sense, a computational model of this form is still general and can still describe any observed input-output mapping.

However, and somewhat oddly, because they express the joint probability distribution in this manner, computational models of this form are called Bayesian: they "combine" the prior and likelihood in the statistical optimal manner, that is to say, by application of Bayes' theorem. We emphasize once more that this "combination" is not a transformation, not a process; it is a representational choice.

Even though the prior and likelihood objects here stem from an arbitrary representational choice concerning the joint probability distribution, some have considered them as hypothetical constructs, and tried to reify them. In simpler terms, the question is whether priors and likelihoods are "somewhere in the brain", and connected to something that implements Bayes' theorem. This question begins to be investigated from a neuroanatomical perspective (Vilares et al.; 2012), but another manner is through the interpretability of the prior and likelihood. For instance, an optimal Bayesian model can be based on a prior and likelihood that are very easily interpreted, and motivated from considerations external to the current modeled phenomenon.

This is a classical manner to justify prior distributions. Examples abound where they are explained by an ecological argument, whereby capturing statistics of the agent's natural environment would provide an evolutionary advantage, compared to ignoring these statistics (Geisler and Diehl; 2003, Geisler; 2008). This has been proposed in models of visual perception of scenes, with the prior that light comes from above (Ramachandran; 1988, Mamassian and Goutcher;

[7] Note that this terminology predates the modern definition of probabilities, and was justified in the early XXth century, when likelihoods and probabilities were construed as different mathematical concepts (Fienberg; 2006). We technically do not find much interest in considering a probability distribution as a function of its right-hand side; the interesting feature of probability distributions is the normalization rule, which concerns the left-hand side. We expect the term "likelihood" to either disappear, or gradually shift towards mistakenly describing terms of the form $P(\textit{data} \mid \textit{parameters})$.

2001, Kersten et al.; 2004, Morgenstern et al.; 2011) and is stationary (Mamassian et al.; 1998), that the viewpoint is situated above the scene (Mamassian et al.; 2003), or that contours usually follow precise statistics in natural scenes (Geisler et al.; 2001); in models of object perception, with the prior that geometrical shapes are regular (Kersten and Yuille; 2003) and convex (Mamassian and Landy; 1998); in models of facial perception, with the prior that faces are convex (Hill and Johnston; 2007); in models of movement perception, with priors that objects' translation or rotation speeds are low (Weiss et al.; 2002, Stocker and Simoncelli; 2006, Colas et al.; 2007); in models of body proprioception, with priors that body rotation speeds are low (Laurens and Droulez; 2007); etc.

When subjects use "inadequate" priors, or when experimenters assume priors which do not correspond to the ones used by the subjects, it may result in faulty analyses; some have rejected the Bayesian approach on these grounds, which is probably hasty (Trimmer et al.; 2011).

5.2.2 Bayesian modeling at the algorithmic level

Instead of decomposing the joint probability distribution into a product of a prior $P(P)$ and a generative model $P(S \mid P)$, consider now the alternate decomposition, $P(P \mid S)P(S)$. It also features terms that are easily interpreted: $P(P \mid S)$ is a direct model of the perception process, and $P(S)$ is a prior distribution over sensations, that captures regularities of the sensor readings. It would be easy to assume that the system is able to capture the relation between sensations and perceptions with this model (and, because of the mathematical equality highlighted above, without further assumptions, it would also be indistinguishable from the previous model). Therefore, a computational model does not really bring credence to the "perception as inversion" theory, since at this stage of the analysis, it is totally equivalent to a "perception as decoding" theory.

We showed that all four terms $P(P)$, $P(S)$, $P(P \mid S)$ and $P(S \mid P)$ could easily be interpreted in the context of a model of a perception process. However, in Bayesian modeling, all four cannot be defined independently. For instance, defining $P(P)$ and $P(S \mid P)$ automatically constrains the other two.

When no simplifying assumption is made, no representational constraint limits the power of expression of either $P(P)$ or $P(S \mid P)$. They are general probability distributions, and can capture any statistical regularity that might be in the experimental data. In that case, choosing $P(P)$ and $P(S \mid P)$ or $P(S)$ and $P(P \mid S)$ is not an interesting issue, and Bayesian inference allows to pass from one another, without loss of information.

Another case in which any statistical regularity can still be captured in a joint probability distribution $P(S\ P)$ is when intermediary variables are added into the model. Consider for instance $P(S\ H\ P)$; whatever the hidden variable, and except some pathological cases, it is still possible to retrieve an arbitrary $P(S\ P)$ from $P(S\ H\ P)$, by marginalizing over H. This is what we illustrated in COSMO-Perception, in our indistinguishability theorem. We showed that, in some ideal learning conditions, a motor theory of speech perception $P(O_S\ M\ S)$ would yield a perceptual classifier $P(O_S \mid S) \propto \sum_M P(O_S\ M\ S)$ indistinguishable from a direct auditory classifier $P(O_L \mid S)$.

On the contrary, when simplifying assumptions are made, for instance, concerning the parametrical form of probability distributions, the generality of the model breaks down. Such assumptions limit the power of expression of probabilistic terms, which then can only approximate the statistical regularities in experimental data.

Consider prior probability distributions $P(P)$ and $P(S)$. It is usually argued that it makes

sense, for the system, to encode statistical frequencies of the percepts into a prior distribution $P(P)$; it is equally likely that the system would be sensitive to the statistical frequencies of the stimuli, that would be encoded into $P(S)$. For instance, in an activation based implementation of these distributions, correctly encoding the most frequent stimuli and perceptions as resting states speeds up their identification. Treating frequent input faster is a good strategy towards minimizing overall computation time, which certainly has an adaptative advantage; it is also a well-known trick in computer science, e.g., Huffman coding.

Consider now compact representations of such prior distributions, such a Gaussian probability distributions. It would be impossible for a vanilla Bayesian model, that is to say, a straight decomposition of the joint probability distribution $P(S\,P)$ to feature both Gaussian priors $P(P)$ and $P(S)$ along with either the generative model $P(S \mid P)$ or the decoding model $P(P \mid S)$. Therefore, all models that are justified because they feature a representation of the prior probability distribution over percepts, motivating it on an evolutionary advantage it would provide, should also justify why they discard capturing the prior over stimuli.

A Bayesian model can be written to feature approximated representations of both prior probability distributions and one, or both of the transfer models; using Bayesian programming and coherence variables, e.g.:

$$P(S\,\Lambda_S\,S'\,P'\,\Lambda_P\,P) = P(S)P(\Lambda_S \mid S\,S')P(S' \mid P')P(\Lambda_P \mid P\,P')P(P)\;, \tag{5.3}$$

$$\begin{aligned} P(S\,\Lambda_S\,S'\,P'\,S''\,P''\,\Lambda_P\,P) &= P(S)P(\Lambda_S \mid S\,S'\,S'')P(S' \mid P')P(P'' \mid S'') \\ &\quad P(\Lambda_P \mid P\,P'\,P'')P(P)\;. \end{aligned} \tag{5.4}$$

To the best of our knowledge, such models of perception do not exist in the literature, in this probabilistic form, although they could correspond to the pairings of forward and inverse models suggested in the motor control domain (Wolpert and Kawato; 1998, Imamizu et al.; 2003). Also, it would be difficult to construe them as computational models, as they clearly include assumptions which make them deviate from an optimal description of the joint probability distribution $P(S\,P)$. Such models would be at the algorithmic level of Marr's hierarchy.

We have discussed how a simple Gaussian assumption on a prior probability distribution already limits the power of expression of a probabilistic model. We have shown that making the model more complex, by duplicating variables, also makes the model deviate from optimality. This is also the case for all other possible simplifying assumptions, such as using approximate inference instead of exact inference, using conditional independence assumptions, either directly or through the decomposition of the model into independent sub-models, etc. In our strict reading of Marr's hierarchy, any such deviation from optimal description of the joint probability distribution would imply that the model becomes algorithmic. This would mean that much of the models, claimed in the literature to be computational models, would instead be construed as algorithmic in our view.

Whatever their classification, whether they are considered computational despite a few constraining assumptions, as they are mostly presented by their authors, or algorithmic as in our strict interpretation of Marr's hierarchy, many models lie in this "gray area". A notable example is the classical sensor fusion model, most well-known for its application to visuo-haptic sensor fusion for height estimation (Ernst and Banks; 2002). This model claims statistical optimality in the sense of variance-minimization, but features a crucial conditional independence hypothesis in its naïve Bayesian sensor fusion model, along with, usually, Gaussian and uniform parametrical forms that respectively constrain the sensor prediction probability distributions and prior probability distribution. This model has been applied to a wide variety of sensory fusion cases,

either intra-sensory cue fusion or inter-sensory fusion, either bi-modal or tri-modal, etc (van Beers et al.; 1996, Knill; 1998, Jacobs; 1999, Anastasio et al.; 2000, Hillis et al.; 2002, Jacobs; 2002, Zupan et al.; 2002, Weiss et al.; 2002, Battaglia et al.; 2003, Gepshtein and Banks; 2003, Oruç et al.; 2003, Alais and Burr; 2004, Banks; 2004, Hillis et al.; 2004, Drewing and Ernst; 2006, Jürgens and Becker; 2006, Wozny et al.; 2008, Nardini et al.; 2012). It has also been applied to the sensorimotor case, as is (Körding and Wolpert; 2004), or in more sophisticated formulations involving Kalman Filters (Wolpert et al.; 1995, van der Kooij et al.; 1999). In most cases, this model is presented by putting forward its statistical optimality as a defining feature.

The more recent "causal inference" model of multi-sensory perception (Körding and Tenenbaum; 2006, Körding et al.; 2007, Sato et al.; 2007, Beierholm et al.; 2008, 2009, Shams and Beierholm; 2010) is also noteworthy. It is an extension of the "optimal" sensor fusion above, with a hierarchical layer articulated by an internal, latent variable. This variable represents the number of physical sources from which stimulations originated. Two sub-models perform perception under the assumption that there is either one or two sources, and their "optimal" combination is provided by marginalization over the internal variable. Causal inference models clearly appear to stand halfway between the two levels of computational and algorithmic models: on the one hand, they assume an internal representation of the number of causes of the observed stimulations and corresponding sub-models, on the other hand, Bayesian inference is used to predict optimal combination of the sub-models.

Such "sensory" causal inference models can also be seen as the application to the sensorimotor domain of previous "high-level cognitive" causal inference models. These models are concerned with how a cognitive agent can capture categorical and structural relations between external objects, and usually are treated in optimality and rationality terms (Kemp and Tenenbaum; 2008, Holyoak; 2008, Holyoak and Cheng; 2011, Tenenbaum et al.; 2011).

Another example is Norris' Bayesian Reader model, in the domain of visual word recognition (Norris; 2013). It was initially described as an optimality motivated model (Norris; 2006, Norris and Kinoshita; 2008, Norris; 2009), but was later refined, by adding substantial hypotheses to the model (Norris and Kinoshita; 2012), e.g., about letter position encoding and algorithms for letter sequence comparisons. Norris acknowledges the gradual slide into the algorithmic level (Norris and Kinoshita; 2012, p. 520):

> In terms of Marr's [...] analysis of levels of description and explanation, our theory can be thought of as occupying a space between the computational and algorithmic levels. Our claim that evidence is accumulated gradually by a stimulus-sampling process is a specifically psychological assumption and therefore is much more concrete than Marr's computational level. However, the model is more abstract than Marr's algorithmic level, as we have no theoretical or empirical basis for speculating about the exact algorithms employed.

Our contribution in the domain of visual word recognition, the BRAID model which is part of ongoing research, is a stark contrast to this approach. The BRAID model is clearly an algorithmic model, inspired by other models of the literature, constrained by neurobiological and behavioral evidence.

Other examples include the models at the rational-process level, encompassing rational models performing inference with limited computation and time resources, which we have already discussed above, but also the ideal adapter model of speech perception Kleinschmidt and Jaeger (2015), that, right from the start, assumes a Gaussian generative model, the incrementally con-

strained ideal observer models of rule learning (Frank and Tenenbaum; 2011, 2014), the many probabilistic models of sentence parsing, requiring internal representations of syntactic entities and syntactic structures (for a review, see Chater and Manning (2006)), and surely many others.

All this analysis advocates in favor of our use of the terminology, contrary to the currently dominant acceptation of the term "Bayesian", in which it corresponds to an "optimal" combination of probability distributions according to Bayes theorem. It is true that for a strictly computational model expressed in the probabilistic framework, Bayes' theorem formalizes statistically optimal probability calculus, so that being Bayesian or being optimal can be interchangeably used.

However, consider now an algorithmic model expressed in the probabilistic framework: the core of such models, what makes them interesting, meaningful, and even refutable, is more likely to be their many assumptions, than the fact that Bayes' theorem is used to perform inference in them. In that sense, even though knowledge manipulation by Bayes' theorem remains "optimal" in an algorithmic model, it becomes a very weak sense of optimality. It is even the case that carefully crafted probabilistic models can "bypass" the constraints of Bayes' theorem, e.g., using coherence variables to mix several probability distributions over a single variable (see Section 2.3.2), or implementing a linear summation model instead of the more natural product of distributions advocated by Bayes' theorem, etc. This is because of the high flexibility of general purpose Bayesian modeling.

Furthermore, in the "gray area" between computational and algorithmic models of cognition, it is troublesome to identify what explains the behavior of a probabilistic model that includes some constraining assumptions; is it still near-optimal and thus the fact that it revolves around Bayes' theorem is still interesting, or is it mainly driven by its simplifying assumptions? We conclude that equating "Bayesian" with "optimal" is not a tenable proposition, except for purely computational models. In our view, whether a model is computational or algorithmic depends on its deviation from optimality; whether it is expressed in probabilistic terms or not is orthogonal and irrelevant. We use the term "Bayesian" not to refer to optimality, but to refer to the fact that, in line with subjective Bayesianism, we model cognitive states of knowledge and knowledge manipulation, in a probabilistic framework. We therefore defend the use of "Bayesian algorithmic" cognitive models to describe models we contribute.

5.3 Is Bayesian modeling a psychological theory?

We now return to the original issue, pondering whether Bayesian modeling as a whole could be a psychological theory. We have already shown that computational modeling was problematic in this regard, as it could be argued that it is not even cognitive modeling at all. In other words, it is not able to delve into psychological constructs, either representational or algorithmic. We have also shown that some Bayesian models, presented in the literature as computational, should in fact be construed as algorithmic, as they (rightly) feature assumptions about psychological processes, along with (wrongly so, in our opinion) optimality claims. We have argued that such models could be considered as algorithmic, despite their optimality claims. Concerning the models we contributed (see Chapters 3 and 4), we have argued that they were algorithmic, too, and did not resort to optimality claims. We rephrase the debate, casting it into a more general formulation: is Bayesian modeling, whatever its form, a scientific theory?

Our initial answer remains: because of the expressive power of Bayesian modeling, no, it cannot be a psychological theory, and not even a scientific theory. Indeed, as any computable

function can be expressed in a Bayesian model, merely being Bayesian is not a specific, refutable, proposition.

This is not a unique, or even surprising, situation. There are many mathematical frameworks which are too general, allowing to express whatever input-output mapping we throw at them, and thus not considered as scientific propositions. For instance, to start from a close neighbor of Bayesian modeling, consider connectionist modeling: the well-known universal approximation theorems of Artificial Neural Networks demonstrate their ability to infinitely precisely approximate whatever mathematical function, given enough neurons at the hidden layer (Cybenko; 1989, Hornik et al.; 1989, Hornik; 1991). This is also the case for all mathematical frameworks that consist in function approximation in some functional basis of interest, such as the Fourier Transform [8], wavelet transforms, probabilistic mixture models, principal component analysis and dimensionality reduction methods in general, etc.

A similar property plagues optimal control theory and Bayesian decision theory, in which complete class theorems (Robert; 2007, Daunizeau et al.; 2010, Friston et al.; 2012) imply that, for any observed input-output mapping, there exists a pair of a prior distribution and cost function that capture this mapping in an optimal manner. In other words, without commitment to additional constraints, such as regularization forms on the prior or the cost function, optimal control is an under constrained theory [9].

Ways to avoid this issue include dismissing the optimality concept as a useful descriptive theory (Loeb; 2012, Patri et al.; in press), or, instead, focusing on the neurobiological plausibility and interpretability of the considered loss function, exploring for instance the plausibility that the central nervous system would measure and compute acceleration, energy expenditure, torque, jerk, or end-point position variance (Flash and Hogan; 1985, Uno et al.; 1989, Viviani and Flash; 1995, Harris and Wolpert; 1998, Kawato; 1999, van Beers et al.; 2002).

The same problem has been discussed, more generally, about the concept of rationality itself (Eberhardt and Danks; 2011, Elqayam and Evans; 2011, Marcus and Davis; 2013). Indeed, one of the elements of the probabilistic cognitive revolution during the 90's was to replace the notion of logical rationality by probabilistic rationality, in a variety of high-level cognitive reasoning tasks (Oaksford and Chater; 2001). A well-known objection to probabilistic rationality concerns the asymmetries in the subjective evaluation of probabilities in subjects (Kahneman and Tversky; 1979, Tversky and Kahneman; 1981). In response, some have argued (Dehaene; 2012) that probabilistic computations could be correct and rational, but followed by a faulty decision process, or that the brain would only approximate Bayesian-optimal computations, with large deviations from optimality for extremely low or high probability values, or even that probabilistic optimality would only apply for a subset of cognitives processes (for instance, low-level cognition would perform perception and motor tasks optimally, contrary to high-level, language based cognition). Recall that statistical optimality is itself an ill-defined mathematical concept, with various incarnations, from minimization of residual error (Brainard and Freeman; 1997), minimization of

[8] Here is a personal anecdote. I was taught in high school a probably gross summary of the "theory of business cycles in economy", which claims that stock market indexes and GDPs could be understood as being the superposition of business cycles of varying time-scales. Later, as an undergraduate, I of course learned about the Fourier Transform. To this day, I wonder if my high-school teacher was at fault, whether economists really consider the "theory of business cycles" as explanatory or merely descriptive, or if they are even aware of the mathematical sleight of hand. I prefer to believe that economy is a serious science with rigorous mathematics and sound epistemological foundations, and that I must simply not know enough about the theory to understand it.

[9] According to my, possibly faulty, memory, I believe that Daniel Wolpert once said, in a talk, something along the lines of: "It is true that the behavior you observe is optimal, there is no question about it. However, it may be the case that you have no idea about the cost function."

variance (Ernst and Banks; 2002), selection of an optimal action policy in the sense that it maximizes expected future reward, as in Bayesian decision theory and optimal motor control, or even just applying Bayes' theorem to perform inference (Frank and Tenenbaum; 2011).

In our view, this situation can be summarized as being a "moving the goalpost" fallacy. In other words, rationality is a massively under constrained concept, always able to adapt to new experimental evidence. Therefore, as a framework, we conclude it is not scientifically refutable. Recall that this is also true of Bayesian modeling; however, we find that Bayesian modeling offers a formally sound mathematical language for knowledge representation and manipulation, which rationality, as a concept, does not.

So far, we have considered the refutability of whole frameworks; we now consider, inside frameworks, broad classes of models. Here again, poor constraining is detrimental, as it leads to indistinguishability between models. For instance, we found one such case in the COSMO model, where we demonstrated that implementations of auditory and motor theories of speech perception could, in optimal learning conditions, be indistinguishable. Such conditions lead to a scientific impasse: whatever the experimental data, if the conditions for the indistinguishability theorem are met, there is no way to discriminate between a perception process with or without involvement of motor knowledge.

Another terminology refers, more broadly, to "mimicry theorems" (Chater; 2005, Chater and Oaksford; 2013): for instance, between pictorial (mental images) vs propositional (verbal) representations of visual memory (J. R. Anderson; 1978), between similarity-based vs model-based categorization (Hahn and Chater; 1998), between exemplar based models and probabilistic sampling based algorithms (Shi et al.; 2010). McClelland (2013) describes a Bayesian model that is formally equivalent to his multinomial interactive activation (MIA) connectionist model. Cooper and Peebles (2015) describe how a Bayesian model, an association model, and a hypothesis testing model could all account for the same behavioral data in a medical diagnostic task; however, these three models have very different interpretations (Cooper et al.; 2003). In motor control modeling, it is also known that performance in nominal situations can be described adequately by a variety of mechanisms, which only differ in the manner they predict reactions to perturbations (Loeb et al.; 1999). Many more examples surely exist.

As a summary, the Bayesian modeling framework as a whole, along with any poorly constrained Bayesian models, are of course unconstrained. In the above discussion, the Bayesian nature of models is irrelevant, and an unfortunate collateral damage. It is my belief that much of the current debate about the explanatory power of Bayesian modeling is made complex because it conflates separate issues: whether computational level (optimal, rational) modeling is useful for cognitive modeling or not, and whether probabilistic modeling is useful for cognitive modeling or not.

5.4 Can Bayesian models be psychological theories?

If Bayesian modeling, as a framework, is not a scientific theory, does it mean that Bayesian models cannot be, either? That is not the case.

First of all, any Bayesian model, if it is constrained enough, can certainly be refuted from experimental evidence. Moreover, the classical tools to evaluate the quality of a model (Jacobs and Grainger; 1994, Myung and Pitt; 1997, Pitt et al.; 2002, Myung and Pitt; 2004) also certainly apply to Bayesian models: goodness of fit, parsimony (both in terms of the number of parameters but also of functional form), generalizability, faithfulness (the fact that what drives the model

is not a spurious effect of implementation), falsifiability (or refutability), interpretability and explanatory adequacy (the model features constructs that make sense, both internally, and with respect to the rest of scientific findings). We can observe that these criteria appear ordered, gradually going from criteria that are well formalized and mathematically explored, to criteria that are qualitative considerations of the scientific practitioner. We note, as an aside, that nothing prevents further mathematical study of criteria down this list; an exciting prospect is the Bayesian comparison of classes of models, in relation with experimental design, to quantify model faithfulness and refutability (see Section A.6).

5.4.1 Constraining a Bayesian Algorithmic model

We recapitulate the steps that can be taken, or that we have followed in our examples, in order to ensure that a Bayesian Algorithmic model is properly constrained. We follow the Bayesian Programming methodology as a guideline here, and first consider model definition.

At this stage, our aim is to define a model of knowledge representation, and thus we are interested in previous evidence about possible psychological representations, from the experimental psychology literature, evidence about coarse-grained neurological pathways, from the neuroscience literature, and even well-established previous mathematical models, from the computational cognitive science literature. This usually provides inspiration about plausible representational spaces, suggesting the domains of probabilistic variables, and even plausible relations and independences between variables, suggesting dependency graphs for the decomposition of the joint probability distribution.

The next step is to mathematically define each term of the probability distribution. Here, the focus is on the balance between expressiveness and model parsimony. Indeed, if left unconstrained, a probabilistic term can theoretically capture any statistical regularity in its domain; however, this expressiveness requires free parameters. Unconstrained non-parametric models and learning, in this regard, are to be manipulated with extreme caution. In our practice, we prefer to limit each probability term with a parametrical form; we usually choose them so as to be consistent with other models of the literature, have plausible flexibility, and remain mathematically tractable, while not necessarily providing closed-form solutions to inference problems. For instance, Gaussian generative models for phonetic categorization models are prevalent in the speech perception modeling literature (Clayards et al.; 2008, Norris and McQueen; 2008, Sonderegger and Yu; 2010, Kleinschmidt and Jaeger; 2015); we have followed this assumption in the variants of the COSMO model.

The third step to define the probabilistic model itself is to set values for its free parameters. This is of course a delicate step. Ideally, some parameters have a physical interpretation, and experimental evidence is available, from independent research, that provides plausible values. For instance, this could be the case in memory retention portions of the BRAID model (see Section 3.2.2). Other methods include learning values from experimental database, setting large variance values to as to poorly inform the model, or exploring the parameter space formally to study its structure, to verify its robustness, to verify whether there are mathematically particular points in the parameter space, etc.

Once all these steps are complete, the probabilistic model is mathematically defined. We then define probabilistic questions asked to the model, in order to simulate cognitive tasks. We must emphasize a particularity of our approach to Bayesian modeling, inherited from Bayesian Programming: once the probabilistic question is defined, we have almost no further modeler's

degrees of freedom in the rest of the process. In other words, once we have defined the knowledge representation model, knowledge manipulation is entirely piloted by Bayesian inference. The answer to the probabilistic question only requires correctly applying probabilistic computation rules, that is to say, the sum and product rules. In this purely automatic manner, we obtain a symbolic expression of the answer to the probabilistic question. In our methodology, we do not model processes directly, we model knowledge that yields processes.

The few remaining choices we can make, at this stage, only concern implementation and possible approximation of the answer to the probabilistic question. Sophisticated methods for Bayesian inference abound, and several have the desirable property to first provide, as approximations, the main statistical trend of the non-approximated answer. For instance, sampling methods, unless pathologically trapped in extreme local minima, will tend to first provide good approximations of the main moments of probability distributions, and further refine subsequent moments. It is a desirable property of a cognitive model to be parsimonious enough not to rely on high-order moments (unless the phenomenon under scrutiny is, itself, a minute effect).

5.4.2 Comparing Bayesian Algorithmic models

At any step of the above process, there usually are one or a couple of modeling choices that stick out. For instance, literature appears inconsistent, debates oppose proponents of one choice or another; or, some previous model screams for implementation of one of its mathematical alternatives. This yields a small number of alternative models; in our practice, we have used such alternative models to explore and answer scientific questions of interest.

Here again, the usual tools for model comparison are available; formal study of the structure of mathematical equations, indistinguishability theorems between the explored variants, experimental simulations and comparison of the results to some known effect, ability to capture experimental evidence, etc. Each well-constrained Bayesian model can be refuted from experimental evidence, discounted when compared to a more able model, etc.

Moreover, since our models are expressed as probability distributions, they could also be compared formally using the tools of Bayesian model comparison and Bayesian distinguishability of models; that would yield analyses in a mathematically unified framework. Unfortunately, so far in our practice, the usual tools of qualitative mathematical comparison have sufficed.

Careful comparison of properly constrained Bayesian models is the means to answer scientific questions about psychological constructs and processes, and thus, build psychological theories. In other words, whereas Bayesian modeling as a whole is certainly not a scientific theory, each well-constrained Bayesian Algorithmic model is certainly a scientific proposition.

5.4.3 Accumulation of Bayesian Algorithmic models

Assume now that we know how to properly constrain Bayesian models, and properly evaluate them as viable models of psychological theories. They help investigate and progress on the scientific exploration of psychological constructs and processes. Now, imagine that the scientific community amasses quite a lot of such good Bayesian models. Would it say anything about the nature of knowledge representation and knowledge manipulation in humans? Would it somehow yield credence to the Bayesian brain hypothesis?

We believe that yes, with an argument from reification by accumulation. We have already demonstrated that no single Bayesian model could definitively answer the question of whether the brain is Bayesian, as they are theoretically not constrained enough, making the question not

Table 5.2: Desiderata that a formal system must satisfy to be demonstrated as equivalent to probability calculus, according to Cox's theorem. Wording adapted from Jaynes (2003).

(I)	Degrees of plausibility are represented by real numbers.
(II)	Qualitative correspondence with common sense: infinitesimal increase in knowledge is measured by an infinitesimal increase in degree of plausibility.
(III)	Consistency
(IIIa)	If a conclusion can be reasoned out in more that one way, then every possible way must lead to the same result.
(IIIb)	The robot always takes into account all of the evidence it has relevant to a question. It does not arbitrarily ignore some of the information, basing its conclusions only on what remains. In other words, the robot is completely nonideological.
(IIIc)	The robot always represents equivalent states of knowledge by equivalent plausibility assignments. That is, if in two problems the robot's state of knowledge is the same (except perhaps for the labeling of propositions), then it must assign the same plausibilities in both.

refutable. However, if many parsimonious and well-constrained Bayesian models account for a wide variety of phenomena related to human cognition, from perception to action, from low-level to high-level cognition, across sensory modalities and amodal processes, then, it might be argued that Bayesian models capture something of cognition. It might, at the least, suggest that the language of probabilities might be the most useful language for describing what the brain is doing; at most, it might suggest that the brain might very well be, "Bayesian".

If that is the case, what are the ingredients of this "language of probabilities", as a language for knowledge representation and knowledge manipulation? What is the common denominator of a large number of parsimonious Bayesian models?

Recall that Cox's theorem provides two manners to understand probability calculus. One is the output of Cox's theorem, that is to say, the notion of the probability of logical propositions, and the rules to manipulate them, that is to say, the sum and product rules (see Section 2.1.2). Another is to consider the input of Cox's theorem. Indeed, Cox's theorem considers the space of all possible formal systems for knowledge representation and manipulation. Out of these, Cox considers some that satisfy a few desiderata, that is to say, desirable properties that a formal system should have. These desiderata are recalled Table 5.2. Cox theorem demonstrates that all formal systems consistent with these desiderata are mathematically equivalent to probability calculus.

The first, and easiest of these desiderata to analyze, concerns the way in which degrees of plausibilities about propositions are represented. Indeed, one of the starting assumptions of Cox's theorem is to represent plausibilities using a single quantity, and more precisely a real number. It follows that plausibilities about any two propositions can be compared, which is called the universal comparability of plausibilities (van Horn; 2003).

Alternatives of course exist. For instance, it could be the case that cognition is based on poorer representations, directly encoding (deterministic) values of properties of interest, and not probability distributions over value domains. This would not make Bayesian modeling technically at fault. Indeed, Bayesian modeling has logical modeling as a special case: a Bayesian model

entirely composed of Dirac delta probability distributions is still a Bayesian model. However, one can mathematically explore any Bayesian model and verify whether it is exactly or close to being a logical model. The scientific community, as a whole, could certainly realize after a while that most Bayesian models in the literature were actually of very small variance, and could safely fall back to logical modeling. However, this is not the direction taken, as, in many domains, the cognitive system seems to be sensitive to and integrate non-zero variance.

It could also be the case that cognition uses richer representations than probabilities, such as the two-dimensional models of belief-function theory, also known as Dempster-Shafer theory (Barnett; 1981, Yager; 1987, Smets; 1991) and possibility theory (Dubois and Prade; 1989). We aggregate these theories under the term "multi-dimensional representational" theories. Such theories are sometimes construed as generalizations of Bayesian modeling, but it is unsure whether they are strict generalization. In other words, it is unclear to us whether there exists some model that can only be described in these theories, or if a Bayesian model can always be "enriched" to capture anything a multi-dimensional representational model can. If that were the case, then Bayesian modeling would not be disproven mathematically, but, over time, multi-dimensional representational theories would prevail on parsimony accounts.

A final point we discuss concerns the strict separation and sequencing, in Bayesian Algorithmic modeling as we practice it, between knowledge modeling and process modeling. This implies that, when we consider two different probabilistic questions asked to the same Bayesian model, they are constrained, as they both come from the same knowledge. In a way, this broad assumption could be refuted over time, if, over and again, we found that the central nervous system would use inconsistent pieces of knowledge in different neurological pathways.

It could also be the case that knowledge encoding and manipulation are more intricate than our methodology suggests. For instance, it could be the case that asking a question systematically affects and modifies the model, so that the order of questions becomes important. This is explored in the Quantum Probability (QP) theory of cognitive modeling (Pothos and Busemeyer; 2013, 2014), as opposed to the classical probabilistic (CP) theory, that we have used throughout this document. QP follows the notation for probability calculus used in quantum physics; accessing information in a probability distribution uses a projection operator, which modifies the initial probability distribution. Such projections are of course not commutative. Proponents of QP argue that order-dependence, which is natural in their notation, is also a pervasive feature of human cognition.

As previously, between Bayesian modeling and multi-dimensional representational theories, it is unclear how CP and QP are articulated. Whether one is a generalization, or a strict-generalization of the other, or whether they are equivalent model classes is, to the best of our knowledge, an open question. We believe, but have no demonstration, that both are actually equivalent, as it seems that a carefully crafted Bayesian model expressed in CP could very well be made to be order-dependent as well. If that were the case, as previously, QP and CP could not be opposed on mathematical grounds, but one could very well prevail on parsimony grounds, over time.

5.5 Summary and position

In this Chapter, we have proposed a panorama of Bayesian cognitive modeling, contrasting the main trends in the literature and our approach. We have recalled Marr's hierarchy of models in cognitive science and discussed the notions of rationality and optimality characteristic of com-

putational level models. We have argued that Bayes-optimality, found in many computational models expressed in the probabilistic framework, was somewhat problematic, especially when applied to models including representational or algorithmic assumptions. We have found that such models would better sit at the algorithmic level, where, shedding their optimality claims, they would instead be able to be considered as refutable models of psychological constructs and processes.

We have recalled the ongoing debate about the explanatory power of Bayesian modeling in cognitive science. We have argued for an original position based on Cox's theorem, which yields a theoretical impasse because it demonstrates that Bayesian modeling, in general, has too much power of expression. In other words, we have shown that Bayesian modeling, as a whole, could not be construed as a psychological theory as it could not even be a scientific theory. Defining *a* Bayesian model of some cognitive phenomenon certainly does not allow to conclude that cognition is Bayesian. In other words, and in a nutshell: it is true that unconstrained Bayesian models are unconstrained; not because they are Bayesian, but because they are unconstrained. It is bad practice to just show one Bayesian model, even when it is the first in a domain; just because it is probabilistic, it does not bring anything to the scientific table. It is also bad practice to interpret a computational-level model as claiming anything about representations and processes; and also bad practice to interpret a de facto algorithmic model as computational, even when its simplifying assumptions, that make it deviate from optimality, are hidden in supplementary material.

Instead, we have proposed an argument from reification. Each Bayesian model, when properly constrained and evaluated using the usual scientific tools, can commit and be a refutable proposition about a psychological phenomenon. Accumulation of such Bayesian models, over time, might provide incremental evidence towards the probabilistic nature, if not of cognition, at least of the most parsimonious language to describe cognition. If the framework is not a scientific theory, a large collection of demonstrably good Bayesian models would constitute a scientific proposition, from reification.

We have outlined our contribution in this regard, that is to say, Bayesian Algorithmic cognitive modeling [10]. In this approach, we apply Bayesian Programming so as to define Bayesian models at the algorithmic level, without reference to optimality. But Bayesian modeling, in this manner, is too flexible because of the power of expression of the probabilistic framework. In other words, and, as usual, experimental validation of complex models with large number of parameters is not easy. This means that Bayesian Algorithmic models must be carefully crafted, calibrated and experimentally validated. To do so, we apply the standard tools: we anchor our assumptions in neuroscientific and behavioral evidence, we strive towards parsimony, we carefully consider free-parameter dimensionality, etc. This is the standard fare of scientific enquiry. Comparison of Bayesian Algorithmic models provides a rigorous method for scientific modeling of psychological representations and processes.

However, does our analysis provide a definite answer to the original question? Recall that we summarized it as the question of whether we were doing Bayesian models of cognition, or models of Bayesian cognition. We have refined this question, provided an initial answer concerning the framework as a whole, contrasted this answer with one concerning each Bayesian model, and

[10] Pondering terminology and acronyms, I considered for a moment adding a word to oppose the usual "optimality" based approach and our "descriptive" approach. However, I found that Bayesian Algorithmic Descriptive modeling would be construed as a BAD approach to cognitive modeling, and would leave unfairly biased memory traces in people's mind. I was, unfortunately, unable to find a synonym of "modeling" that would start with an *n*; that would have yielded a nice BACoN acronym, which might have been more favorably appraised.

their accumulation. However, even if cognitive science were built entirely and exclusively on Bayesian models, would it mean that cognition is Bayesian?

Our final answer requires to realize that this last question is merely an instance of the eternal epistemological debate between instrumentalism and realism. In a nutshell, the instrumentalist stance would argue in favor of "Bayesian modeling of cognition": cognition is whatever it is, but to understand it, we use Bayesian models. In contrast, the realist stance would argue in favor of "modeling of Bayesian cognition": because our tools to understand cognition are all Bayesian, it makes sense to assume that the cognitive system is also Bayesian.

Note that this position is contrary to the one of Colombo and Seriès (2012), where a realist position requires mechanistic models (algorithmic or implementation), whereas an instrumentalist position is satisfied with computational accounts. We argue that the mechanistic/computational and realist/instrumentalist dimensions are orthogonal. As a demonstration, one can find a clearly instrumentalist neuroscientist (Fiorillo; 2012). We argued with caution that probabilities are only descriptive tools; Fiorillo (2012) almost denies that the central nervous system would "perform computation", casting doubt on the brain-as-computer analogy. His argument is that in physics, no one would argue that a falling rock "computes" gravitational forces (for a counter-argument, see Edelman (2008)). This purely instrumentalist view in the context of Bayesian modeling is also found in other domains; for instance, Jaynes (1957) analyzes statistical mechanics, not as a physical theory, but as a description of the physicist's ability to predict micro-states from macro-measurements (but see the counterargument by Friston (2010b)).

In a way, the distinction between the frequentist and subjective approaches to probabilities mirrors the instrumentalist vs realist debate. In the frequentist view, probabilities are properties ascribed to objects, independently of the observer. This is an ontological definition, clearly in line with a realist view. On the contrary, in the subjective view, probabilities are properties of the observer. This is an epistemic definition, clearly in line with an instrumentalist view. A remarkable coincidence is that the frequentist and subjective approaches to probabilities, whatever their axiomatics, almost perfectly match technically. Therefore, it is quite possible to be a practitioner of probabilistic modeling without ever wondering whether they represent frequencies or states of knowledge. Also, it is quite possible to develop Bayesian models in cognitive science without ever pondering whether the brain is Bayesian; this gives hope for rapid progress in cognitive science.

The instrumentalist vs realist question is an age old debate, and, ultimately, an epistemological stance that one has to take. We would further argue that it is actually a metaphysical question, since we can imagine no scientific experience that would discriminate both propositions. Even more importantly, whatever your stance in this matter, from a purely frequentist-idealist stance to a purely subjectivist-instrumentalist stance, the mathematics are almost the same, the main difference being one of interpretation. In other words, it has almost no effect on the day-to-day proceedings of the scientific endeavor, which strives towards theories and models from a technical standpoint.

An interesting position is then to be agnostic. On the surface it may appear as a cop-out; instead, I believe it is the only scientifically viable stance, as matters that are not decidable by experimental observations should be outside our scope. In a way, this is the ultimate indistinguishability result I propose. Finally, I also believe the agnostic position to be a wise one, casting doubt on the purpose of debates that are demonstrably pointless; moreover, the agnostic position surely helps lower blood pressure, which is a definite bonus.

CHAPTER 6

Conclusion

6.1 Main contribution

In this *habilitation*, we have outlined Bayesian Algorithmic Modeling, a methodology for defining structured and hierarchical probabilistic models of cognitive agents. This methodology directly inherits from Bayesian Programming, a probabilistic language previously widely applied for the programmation of artificial cognitive agents. Its particular feature is that, as a programming language, it falls into the class of declarative languages, and can be seen as a "probabilistic Prolog". Therefore, as a language for cognitive models, it allows expressing models of knowledge, which automatically yield models of processes by Bayesian inference. In other words, cognitive processes are not modeled directly and independently from one another: they are assumed to coherently refer to the same knowledge that the system would have acquired.

We have then described two domains in which we have applied Bayesian Algorithmic Modeling: reading and writing on the one hand, and speech perception and production on the other hand. With five Ph.D. students, we have defined 5 main models in these domains. The first model was BAP, a model of cursive letter recognition and production, involving a simulated perception sub-model. The second model was BRAID, a model of word recognition with an explicit attention repartition model and involving interference and dynamic effects, paving the way towards modeling reading acquisition and phonological processes. The last three models have been declinations of COSMO, a general architecture for communicating agents. The third model was COSMO-Emergence, a model for the study of the emergence of phonological systems. The fourth mode was COSMO-Perception, a model for the comparison of motor and auditory theories of speech perception, applied to the case of syllable perception. The fifth and final model was COSMO-Production, a Bayesian reformulation of an existing optimal control model for the production of sequences of phonemes.

Finally, we have discussed the place of Bayesian Algorithmic Modeling in the current panorama of Bayesian modeling in cognitive science. We have argued that, because it originates from Bayesian Programming, our approach is quite singular in the domain. Whereas most Bayesian cognitive models claim to derive from some optimality principle, we have argued that such a principle should better be restricted to computational models, in the sense of Marr's hierarchy. Unfortunately, a strict reading of computational modeling contradicts the goal of investigating

psychological processes and representations. We have concluded that the common conflation of Bayesian modeling and probabilistically optimal modeling was an unfortunate feature of current literature in cognitive modeling, making the terminology of the field difficult to apprehend.

Instead of Bayes-optimal computational models, we have argued for Bayesian models at the algorithmic level of Marr's hierarchy, without reference to optimality, but featuring instead hypotheses about internal representations and assumptions about the properties of cognitive processes. Because Bayesian Algorithmic Modeling is general purpose, it allows to express arbitrarily complex probabilistic models; because probabilities can be construed as an extension of logic, the power of expression of the whole formalism is too large, and models are not individually theoretically constrained. Providing a single Bayesian model does not allow to conclude directly that the cognitive system, somehow, would be Bayesian.

We have thus advocated model comparison, either formal or experimental, to properly constrain models and their parameters. Such model comparison features various considerations, such as the distinguishability of models, their relative adequacy with respect to experimental data, their parsimony, their ability to account for previous results, their interpretability, etc. This is the usual fare of scientific modeling.

6.2 Main perspectives

In the course of this *habilitation*, we have outlined technical and short term perspectives opened by current work on the BRAID model and on COSMO variants, as they were introduced. Such perspectives are not recalled here, and will be the object of ongoing Ph.D. projects. Instead, we wish to highlight three other perspectives.

The first concerns model comparison. Indeed, the Bayesian framework is of course not only a language for expressing models, it also extends to their formal comparison. This mathematical continuity is a definite advantage, when comparing Bayesian modeling with other mathematical frameworks. In the work we presented here, unfortunately, we did not have the opportunity to apply Bayesian model comparison to some crucial experimental data (but see the project described in Annex A.3). Instead, we have performed model comparison either on qualitative grounds, or on mathematical indistinguishability grounds. Hopefully, once our models are settled as viable accounts of cognitive processes in the literature, their predictions will yield crucial experiments and they will be amenable to formal model comparison.

The flip side of model comparison is experimental design. Once again, we did not have yet the opportunity to apply Bayesian formal tools here. This means we did not have yet the chance to pursue our theoretical work on Bayesian distinguishability of models, a measure of the distance between models we have defined, and which can be included into an adaptive experimental design procedure (see Annex A.6). Future work will aim at applying model comparison and adaptive experimental design based on Bayesian distinguishability to Bayesian Algorithmic Models we have designed.

The second perspective we wish to highlight results from the uniformity of the mathematical framework between the various models we have defined. As we have illustrated time and again in this manuscript, the models we consider are structured, featuring several components, articulated either by coherence variables, by the product rule, or by recursive sub-model calls, into the joint probability distribution of an overarching model.

We have thus defined the BAP model as featuring visual perception of letters, motor planning of movement sequences to trace letters, and effector control to produce such sequences. We

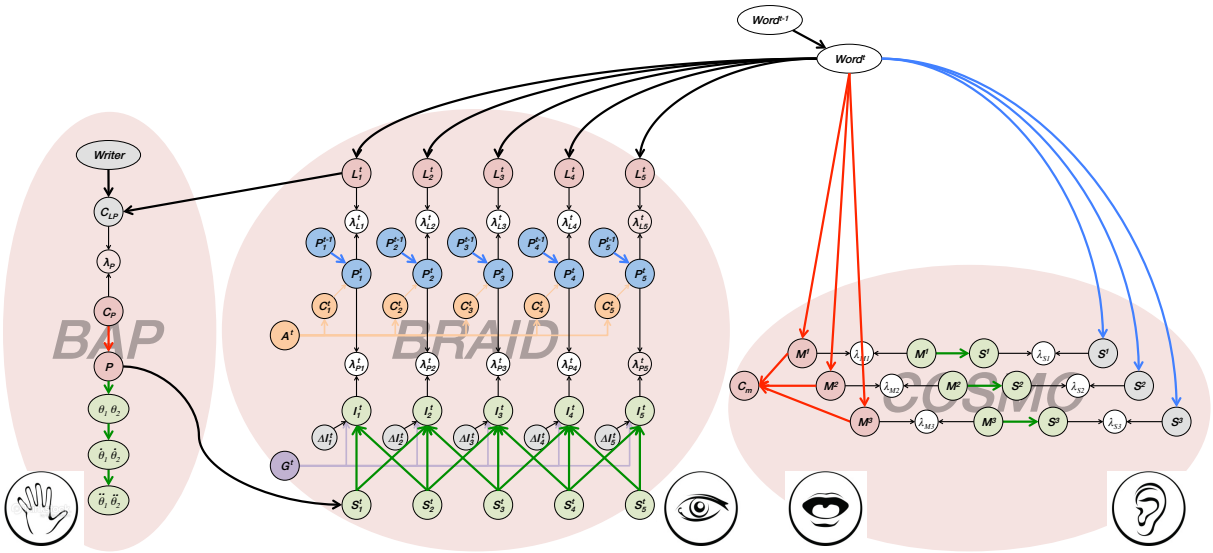

Figure 6.1: **Graphical representation of the structure of a model integrating BAP, BRAID and COSMO components.** This model would include an eye and an ear as sensors, for visual processing (of letter images) and auditory processing (of speech sounds), and a hand and a mouth as actuators, for drawing and writing (letters) and articulation (of speech movements).

remark that letters are also a component of the BRAID model, where sequences of letters are processed visually. However, the BRAID model does not feature any motor process (other than modifying gaze position). We thus imagine a combination between BRAID and BAP, where letters of words read by BRAID could be input to the motor process of BAP.

We have also defined the COSMO model for speech perception and production. It heavily features speech "object" representations, which can either refer to phonemes, syllables, or even words. We also remark that words are also a component of the BRAID model. We thus imagine a combination between BRAID and COSMO, where words read by BRAID could be read out loud, or where words heard by COSMO could evoke visual images of the way they are written. Finally, in a BAP-BRAID-COSMO combination (illustrated Figure 6.1), words heard by COSMO could be written down by BAP. One could even imagine extending this model to include other visual processing, such as shape and color processing.

Such a model would include visual and auditory sensory processes, along with speech articulation and hand trajectory motor processes. It could be a non ad-hoc and unique opportunity to explore cross-modal effects, seldom treated in the literature, such as the Bouba-Kiki effect (often hypothesized to result from interaction between visual shape and speech articulation processes), or the Stroop effect (often hypothesized to result from interaction between visual color and lexical decoding processes).

The third and final perspective we wish to highlight, in conclusion, is Bayesian Algorithmic Modeling in itself. Indeed, this *habilitation* helped us evaluate its current place in the panorama of the literature of cognitive science, and more generally, the place of algorithmic cognitive models: both are completely marginalized by Bayes-optimal methods. We are thus forced to humbly realize that establishing these as viable and valid research methods, in the cognitive science community, will be no small endeavor. We hope the present contribution is a positive step in this direction.

Bibliography

David Alais and David Burr. The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, 14:257–262, February 2004. *cited page 78*

Thomas Anastasio, Paul Patton, and Kamel Belkacem-Boussaid. Using bayes' rule to model multisensory enhancement in the superior colliculus. *Neural Computation*, 12:1165–1187, 2000. *cited page 78*

Barton L. Anderson. Can computational goals inform theories of vision? *Topics in Cognitive Science*, 7: 274–286, 2015. *cited page 68*

John R. Anderson. Arguments concerning representations from mental imagery. *Psychological Review*, 85(4):249–277, 1978. *cited page 81*

John R. Anderson. *The adaptive character of thought*. Erlbaum, Hillsdale, NJ, 1990. *cited page 72*

John R. Anderson. Is human cognition adaptive? *Behavioral and Brain Sciences*, 14:471–517, 1991. *2 citations pages 68, 70*

Mark Andrews and Thom Baguley. Prior approval: The growth of Bayesian methods in psychology. *British Journal of Mathematical and Statistical Psychology*, 66:1–7, 2013. *cited page 73*

Martin S. Banks. Neuroscience: What you see and hear is what you get. *Current Biology*, 14(6):R236–R238, 2004. *cited page 78*

Marie-Lou Barnaud, Nicolas Morgado, Richard Palluel-Germain, Julien Diard, and Anne Spalanzani. Proxemics models for human-aware navigation in robotics: Grounding interaction and personal space models in experimental data from psychology. In *Proceedings of the 3rd IROS'2014 workshop "Assistance and Service Robotics in a Human Environment"*, 2014. *cited page 114*

Jeffrey A. Barnett. Computational methods for a mathematical theory of evidence. In *Proceedings of the Seventh International Conference on Arti- ficial Intelligence*, pages 868–875, 1981. *cited page 85*

Peter W. Battaglia, Robert A. Jacobs, and Richard N. Aslin. Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A*, 20(7):1391–1397, 2003. *cited page 78*

Jeffrey M. Beck, Peter E. Latham, and Alexandre Pouget. Marginalization in neural circuits with divisive normalization. *The Journal of Neuroscience*, 31(43):15310–15319, 2011. *cited page 74*

Ulrik R. Beierholm, Konrad P. Körding, Ladan Shams, and Wei Ji Ma. Comparing Bayesian models for multisensory cue combination without mandatory integration. In J.C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 81–88, Cambridge, MA, 2008. MIT Press. *cited page 78*

Ulrik R. Beierholm, Steven R. Quartz, and Ladan Shams. Bayesian priors are encoded independently from likelihoods in human multisensory perception. *Journal of Vision*, 9(5):1–9, 2009. *cited page 78*

Mario Bertero, Tomaso A Poggio, and Vincent Torre. Ill-posed problems in early vision. *Proceedings of the IEEE*, 76(8):869–889, 1988. *cited page 74*

Pierre Bessière. Method for determining a value given to different parameters of a system. WO Patent WO/2004/013,714, 2004. *cited page 18*

Pierre Bessière, Eric Dedieu, Olivier Lebeltel, Emmanuel Mazer, and Kamel Mekhnacha. Interprétation vs. description I : Proposition pour une théorie probabiliste des systèmes cognitifs sensori-moteurs. *Intellectica*, 26-27:257–311, 1998a. *cited page 5*

Pierre Bessière, Eric Dedieu, Olivier Lebeltel, Emmanuel Mazer, and Kamel Mekhnacha. Interprétation vs. description II : Fondements mathématiques. *Intellectica*, 26-27:313–336, 1998b. *cited page 5*

Pierre Bessière, Juan-Manuel Ahuactzin, Olivier Aycard, David Bellot, Francis Colas, Christophe Coué, Julien Diard, Ruben Garcia, Carla Koike, Olivier Lebeltel, Ronan LeHy, Olivier Malrait, Emmanuel Mazer, Kamel Mekhnacha, Cédric Pradalier, and Anne Spalanzani. Survey: Probabilistic methodology and techniques for artefact conception and development. Technical Report RR-4730, INRIA Rhône-Alpes, Montbonnot, France, 2003. *cited page 16*

Pierre Bessière, Christian Laugier, and Roland Siegwart, editors. *Probabilistic Reasoning and Decision Making in Sensory-Motor Systems*, volume 46 of *Springer Tracts in Advanced Robotics*. Springer, Berlin, 2008. *3 citations pages 6, 15, 16*

Pierre Bessière, Emmanuel Mazer, Juan Manuel Ahuactzin, and Kamel Mekhnacha. *Bayesian Programming*. CRC Press, Boca Raton, Florida, 2013. *3 citations pages 6, 11, 20*

Louis-Jean Boë. Modelling the growth of the vocal tract vowel spaces of newly-born infants and adults: consequences for ontogenesis and phylogenesis. In John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville, and Ashlee C. Bailey, editors, *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS 99)*, pages 2501–2504, 1999. *cited page 46*

Jeffrey S. Bowers and Colin J. Davis. Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, 138(3):389–414, 2012a. *cited page 66*

Jeffrey S. Bowers and Colin J. Davis. Is that what Bayesians believe? reply to Griffiths, Chater, Norris, and Pouget (2012). *Psychological Bulletin*, 138(3):423–426, 2012b. *cited page 66*

D. H. Brainard and W. T. Freeman. Bayesian color constancy. *Journal of the Optical Society of America A*, 14(7):1393–1411, 1997. *2 citations pages 74, 80*

Daniel A. Braun, Stephan Waldert, Ad Aertsen, Daniel M. Wolpert, and Carsten Mehring. Structure learning in a sensorimotor association task. *PLoS ONE*, 5(1):e8973, 2010. *cited page 69*

Myriam Chanceaux, Vincent Rynik, Jean Lorenceau, and Julien Diard. Writer recognition in cursive eye writing: a Bayesian model. In P. Bello, M. Guarini, M. McShane, and B. Scassellati, editors, *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, pages 2014–2019, Austin, TX, 2014. Cognitive Science Society. *cited page 116*

Nick Chater. A minimum description length principle for perception. In Mark A. Pitt and In Jae Myung, editors, *Advances in minimum description length: Theory and applications*. MIT Press, Cambridge, MA, 2005. *2 citations pages 8, 81*

Nick Chater and Christopher D. Manning. Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences, special issue: Probabilistic models of cognition*, 10(7):335–344, 2006. *2 citations pages 73, 79*

Nick Chater and Mike Oaksford. Ten years of the rational analysis of cognition. *Trends in Cognitive Sciences*, 3(2):57–65, 1999. *cited page 69*

Nick Chater and Mike Oaksford. Programs as causal models: Speculations on mental programs and mental representation. *Cognitive Processes*, 37:1171–1191, 2013. *2 citations pages 8, 81*

Nick Chater, Mike Oaksford, Ulrike Hahn, and Evan Heit. Bayesian models of cognition. *WIREs Cognitive Science*, 1(6), 2010. *cited page 73*

Meghan Clayards, Michael K. Tanenhaus, Richard N. Aslin, and Robert A. Jacobs. Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, 108:804–809, 2008. *3 citations pages 69, 74, 82*

Francis Colas, Jacques Droulez, Mark Wexler, and Pierre Bessière. A unified probabilistic model of the perception of three-dimensional structure from optic flow. *Biological Cybernetics*, 97:461–477, 2007. *2 citations pages 7, 76*

Francis Colas, Fabien Flacher, Thomas Tanner, Pierre Bessière, and Benoît Girard. Bayesian models of eye movement selection with retinotopic maps. *Biological Cybernetics*, 100:203–214, 2009. *cited page 7*

Francis Colas, Julien Diard, and Pierre Bessière. Common Bayesian models for common cognitive issues. *Acta Biotheoretica*, 58(2-3):191–216, 2010. *5 citations pages 5, 7, 11, 20, 73*

Matteo Colombo and Peggy Seriès. Bayes in the brain–on bayesian modelling in neuroscience. *British Journal of the Philosophy of Science*, 63:697–723, 2012. *cited page 87*

G. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42:393–405, 1990. *cited page 17*

Richard P. Cooper and David Peebles. Beyond single-level accounts: The role of cognitive architectures in cognitive scientific explanation. *Topics in Cognitive Science*, 7:243–258, 2015. *2 citations pages 72, 81*

Richard P. Cooper, Peter Yule, and John Fox. Cue selection and category learning: A systematic comparison of three theories. *Cognitive Science Quarterly*, 3(2):143–182, 2003. *cited page 81*

Christophe Coué, Cédric Pradalier, Christian Laugier, Thierry Fraichard, and Pierre Bessière. Bayesian occupancy filtering for multitarget tracking: an automotive application. *The International Journal of Robotics Research*, 25(1):19–30, 2006. *cited page 7*

George Cybenko. Approximation by superpostion of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4):303–314, 1989. *cited page 80*

Jean Daunizeau, Hanneke E. M. den Ouden, Matthias Pessiglione, Stefan J. Kiebel, Klaas E. Stephan, and Karl J. Friston. Observing the observer (I): Meta-Bayesian models of learning and decision-making. *PLoS one*, 5(12):e15554, 2010. *cited page 80*

Luc De Raedt and Angelika Kimmig. Probabilistic (logic) programming concepts. *Machine Learning*, 100:5–47, 2015. *cited page 19*

Stanislas Dehaene. Cours 2011–2012 du Collège de France : Le cerveau statisticien : la révolution bayésienne en sciences cognitives. Lecture in French, available in video form at http://www.college-de-france.fr/site/stanislas-dehaene/, 2012. *cited page 80*

Julien Diard. *La carte bayésienne – Un modèle probabiliste hiérarchique pour la navigation en robotique mobile*. Thèse de doctorat, Institut National Polytechnique de Grenoble, Grenoble, France, Janvier 2003. *4 citations pages 11, 21, 22, 111*

Julien Diard. Bayesian model comparison and distinguishability. In *Proceedings of the International Conference on Cognitive Modeling (ICCM 09)*, pages 204–209, 2009. *cited page 117*

Julien Diard and Pierre Bessière. Bayesian maps: probabilistic and hierarchical models for mobile robot navigation. In Pierre Bessière, Christian Laugier, and Roland Siegwart, editors, *Probabilistic Reasoning and Decision Making in Sensory-Motor Systems*, volume 46 of *Springer Tracts in Advanced Robotics*, pages 153–175. Springer-Verlag, Berlin Heidelberg, 2008. *cited page 111*

Julien Diard and Olivier Lebeltel. Bayesian learning experiments with a Khepera robot. In Löffler, Mondada, and Rückert, editors, *Experiments with the Mini-Robot Khepera : Proceedings of the 1st International Khepera Workshop*, pages 129–138, Germany, 1999. HNI-Verlagsschriftenreihe, band 64. *cited page 6*

Julien Diard and Olivier Lebeltel. Bayesian programming and hierarchical learning in robotics. In Meyer, Berthoz, Floreano, Roitblat, and Wilson, editors, *SAB2000 Proceedings Supplement Book*, pages 140–149, Honolulu, USA, 11–16 2000. International Society for Adaptive Behavior. *cited page 6*

Julien Diard, Pierre Bessière, and Emmanuel Mazer. Hierarchies of probabilistic models of space for mobile robots: the Bayesian map and the abstraction operator. In *Proceedings of Reasoning with Uncertainty in Robotics (IJCAI'03 Workshop)*, Acapulco, Mexico, August 2003a. *cited page 111*

Julien Diard, Pierre Bessière, and Emmanuel Mazer. A survey of probabilistic models, using the Bayesian Programming methodology as a unifying framework. In *The Second International Conference on Computational Intelligence, Robotics and Autonomous Systems (CIRAS 2003)*, Singapore, December 2003b. *2 citations pages 21, 22*

Julien Diard, Pierre Bessière, and Emmanuel Mazer. Hierarchies of probabilistic models of navigation: the Bayesian map and the abstraction operator. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA04)*, pages 3837–3842, New Orleans, LA, USA, 2004a. *2 citations pages 7, 111*

Julien Diard, Pierre Bessière, and Emmanuel Mazer. A theoretical comparison of probabilistic and biomimetic models of mobile robot navigation. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA04)*, pages 933–938, New Orleans, LA, USA, 2004b. *2 citations pages 7, 111*

Julien Diard, Pierre Bessière, and Emmanuel Mazer. Merging probabilistic models of navigation: the Bayesian map and the superposition operator. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS05)*, pages 668–673, Edmonton, Canada, 2005. *2 citations pages 7, 111*

Julien Diard, Panagiota Panagiotaki, and Alain Berthoz. Biomimetic Bayesian models of navigation: How are environment geometry-based and landmark-based strategies articulated in humans? In *Proceedings of the International Conference on Cognitive Modeling (ICCM 09)*, pages 210–215, 2009. *cited page 112*

Julien Diard, Estelle Gilet, Éva Simonin, and Pierre Bessière. Incremental learning of Bayesian sensorimotor models: from low-level behaviours to large-scale structure of the environment. *Connection Science*, 22(4):291–312, 2010a. *2 citations pages 7, 111*

Julien Diard, Muriel Lobier, and Sylviane Valdois. Bayesian modeling of human performance in a visual processing training software. In Vincent Duffy, editor, *Advances in Applied Digital Human Modeling*, Advances in Human Factors and Ergonomics Series, pages 467–476, Boca Raton, Florida, 2010b. CRC Press, Taylor & Francis Group. *cited page 116*

Julien Diard, Pierre Bessière, and Alain Berthoz. Spatial memory of paths using circular probability distributions: Theoretical properties, navigation strategies and orientation cue combination. *Spatial Cognition & Computation: An Interdisciplinary Journal*, 13(3):219–257, 2013a. *2 citations pages 15, 112*

Julien Diard, Vincent Rynik, and Jean Lorenceau. A Bayesian computational model for online character recognition and disability assessment during cursive eye writing. *Frontiers in Psychology*, 4:843, 2013b. *2 citations pages 15, 116*

Knut Drewing and Marc O. Ernst. Integration of force and position cues for shape perception through active touch. *Brain research*, 1078:92–100, 2006. *cited page 78*

Jan Drugowitsch and Alexandre Pouget. Probabilistic vs. non-probabilistic approaches to the neurobiology of perceptual decision-making. *Current Opinion in Neurobiology*, 22:963–969, 2012. *cited page 74*

Didier Dubois and Henri Prade. Fuzzy sets, probability and measurement. *European Journal of Operational Research*, 40:135–154, 1989. *cited page 85*

Frederick Eberhardt and David Danks. Confirmation in the cognitive sciences: The problematic case of Bayesian models. *Minds & Machines*, 21:389–410, 2011. *2 citations pages 74, 80*

Shimon Edelman. On the nature of minds, or: Truth and consequences. *Journal of Experimental & Theoretical Artificial Intelligence*, 20(3):181–196, 2008. *cited page 87*

Michael G Efran and James A Cheyne. Shared space: The co-operative control of spatial areas by two interacting individuals. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 5(3):201–210, 1973. *cited page 115*

Mark Eilers and Claus Möbus. Learning of a Bayesian autonomous driver mixture-of-behaviors (BAD MoB) model. In Vincent Duffy, editor, *Advances in Applied Digital Human Modeling*, Advances in Human Factors and Ergonomics Series, pages 436–445, Boca Raton, Florida, 2010. CRC Press, Taylor & Francis Group. *cited page 7*

Mark Eilers and Claus Möbus. Learning the human longitudinal control behavior with a modular hierarchical Bayesian mixture-of-behaviors model. In *Intelligent Vehicles Symposium (IV), 2011 IEEE*, pages 540–545. IEEE, 2011. *cited page 7*

Shira Elqayam and Jonathan St. B. T. Evans. Subtracting "ought" from "is": Descriptivism versus normativism in the study of human thinking. *Behavioral and Brain Sciences*, 34:233–290, 2011. *3 citations pages 66, 68, 80*

Ansgar D. Endress. Bayesian learning and the psychology of rule induction. *Cognition*, 127:159–176, 2013. *cited page 66*

Ansgar D. Endress. How are Bayesian models really used? Reply to Frank (2013). *Cognition*, 130:81–84, 2014. *cited page 66*

Marc O. Ernst and Martin S. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–33, 2002. *2 citations pages 77, 81*

Naomi H. Feldman, Thomas L. Griffiths, and James L. Morgan. The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review*, 116(4): 752–782, 2009. *cited page 74*

João Filipe Ferreira, Pierre Bessière, Kamel Mekhnacha, Jorge Lobo, Jorge Dias, and Christian Laugier. Bayesian models for multimodal perception of 3D structure and motion. In *International Conference on Cognitive Systems (CogSys 2008)*, pages 103–108, 2008. *cited page 6*

João Filipe Ferreira, Jorge Lobo, and Jorge Dias. Bayesian real-time perception algorithms on GPU. *Journal of Real-Time Image Processing*, 6(3):171–186, 2011. *cited page 6*

João Filipe Ferreira, Miguel Castelo-Branco, and Jorge Dias. A hierarchical Bayesian framework for multimodal active perception. *Adaptive Behavior*, 20(3):172–190, 2012a. *cited page 6*

João Filipe Ferreira, Christiana Tsiourti, and Jorge Dias. Learning emergent behaviours for a hierarchical bayesian framework for active robotic perception. *Cognitive processing*, 13(1):155–159, 2012b. *cited page 6*

João Filipe Ferreira, Jorge Lobo, Pierre Bessière, Miguel Castelo-Branco, and Jorge Dias. A Bayesian framework for active artificial perception. *Cybernetics, IEEE Transactions on*, 43(2):699–711, 2013. *cited page 6*

João Filipe Ferreira and Jorge Dias. *Probabilistic Approaches for Robotic Perception*. Springer Tracts in Advanced Robotics. Springer, Berlin, 2014. *cited page 6*

Stephen E. Fienberg. When did Bayesian inference become "Bayesian"? *Bayesian Analysis*, 1(1):1–40, 2006. *3 citations pages 73, 74, 75*

Christopher D. Fiorillo. Beyond Bayes: On the need for a unified and Jaynesian definition of probability and information within neuroscience. *Information*, 3:175–203, 2012. *3 citations pages 5, 75, 87*

Tamar Flash and Neville Hogan. The coordination of arm movements: an experimentally confirmed model. *Journal of Neuroscience*, 5(7):1688–1703, 1985. *cited page 80*

Michael C. Frank. Throwing out the Bayesian baby with the optimal bathwater: Response to Endress (2013). *Cognition*, 128:417–423, 2013. *cited page 66*

Michael C. Frank and Joshua B. Tenenbaum. Three ideal observer models for rule learning in simple languages. *Cognition*, 120:360–371, 2011. *2 citations pages 79, 81*

Michael C. Frank and Joshua B. Tenenbaum. Corrigendum to "Three ideal observer models for rule learning in simple languages" [Cognition 120 (3) (2011) 360–371]. *Cognition*, 132:501, 2014. *cited page 79*

Nir Friedman, Lise Getoor, Daphne Koller, and Avi Pfeffer. Learning probabilistic relational models. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI-97)*, volume 99, pages 1300–1309, 1999. *cited page 23*

Karl Friston. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11: 127–138, 2010a. *cited page 73*

Karl Friston. Is the free-energy principle neurocentric? *Nature Reviews Neuroscience*, 11(8):605, 2010b. *cited page 87*

Karl Friston. The history of the future of the Bayesian brain. *NeuroImage*, 62:1230–1233, 2012. *cited page 74*

Karl Friston and Stefan Kiebel. Predictive coding under the free-energy principle. *Phil. Trans. R. Soc. B*, 364:1211–1221, 2009. *cited page 74*

Karl Friston, James Kilner, and Lee Harrison. A free energy principle for the brain. *Journal of Physiology – Paris*, 100:70–87, 2006. *cited page 74*

Karl J. Friston, Rick A. Adams, and Read Montague. What is value – accumulated reward or evidence? *Frontiers in Neurorobotics*, 6(11):1–25, 2012. *cited page 80*

Wilson S. Geisler. Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology*, 59:167–92, 2008. *cited page 75*

Wilson S. Geisler and Randy L. Diehl. A bayesian approach to the evolution of perceptual and cognitive systems. *Cognitive Science*, 27:379–402, 2003. *cited page 75*

W.S. Geisler, J.S. Perry, B.J. Super, and D.P. Gallogly. Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, 41:711–724, 2001. *cited page 76*

Andrew Gelman and Cosma Rohilla Shalizi. Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, 66:8–38, 2013. *cited page 73*

Tim Genewein, Eduard Hez, Zeynab Razzaghpanah, and Daniel A. Braun. Structure learning in bayesian sensorimotor integration. *PLoS Computational Biology*, 11(8):e1004369, 2015. *cited page 69*

Sergei Gepshtein and Martin S. Banks. Viewing geometry determines how vision and haptics combine in size perception. *Current Biology*, 13:483–488, 2003. *cited page 78*

Samuel J. Gershman, Eric J. Horvitz, and Joshua B. Tenenbaum. Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, 349(6245):273–278, 2015. *cited page 69*

Estelle Gilet. *Modélisation Bayésienne d'une boucle perception-action : application à la lecture et à l'écriture*. Thèse de doctorat, Université Joseph Fourier, Grenoble, France, Octobre 2009. *cited page 25*

Estelle Gilet, Julien Diard, and Pierre Bessière. Bayesian modelling of sensorimotor systems: application to handwriting. In *NeuroComp 08*, pages 203–206, 2008a. *cited page 25*

Estelle Gilet, Julien Diard, and Pierre Bessière. Bayesian modelling of a sensorimotor loop: application to handwriting. In *Proceedings of the Neural Information Processing Systems (NIPS 2008) workshop "Principled theoretical frameworks for the perception-action cycle"*, 2008b. *cited page 25*

Estelle Gilet, Julien Diard, Richard Palluel-Germain, and Pierre Bessière. Bayesian action-perception loop modeling: Application to trajectory generation and recognition using internal motor simulation. In Ali Mohammad-Djafari, Jean-Francois Bercher, and Pierre Bessière, editors, *Proceedings of the Thirtieth International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (Maxent 2010)*, pages 59–66, Melville, New-York, USA, 2010. American Institute of Physics Conference Proceedings. *cited page 25*

Estelle Gilet, Julien Diard, and Pierre Bessière. Bayesian action-perception computational model: Interaction of production and recognition of cursive letters. *PLoS ONE*, 6(6):e20387, 2011. *3 citations pages 11, 20, 25*

Noah D. Goodman, Vikash K. Mansinghka, Daniel M. Roy, Keith Bonawitz, and Joshua B. Tenenbaum. Church: a language for generative models. In *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence*, volume 22, page 23, 2008. *cited page 6*

Noah D. Goodman, Michael C. Frank, Thomas L. Griffiths, Joshua B. Tenenbaum, Peter W. Battaglia, and Jessica Hamrick. Relevant and robust. a response to Marcus and Davis. *Psychological Science*, 226:539–541, 2015. *cited page 66*

Alison Gopnik and Elizabeth Bonawitz. Bayesian models of child development. *WIREs Cognitive Science*, 6(2):75–86, 2015. *cited page 73*

Andrew D. Gordon, Thomas A. Henzinger, Aditya V. Nori, and Sriram K. Rajamani. Probabilistic programming. In *Proceedings of the 36th International Conference on Software Engineering (ICSE 2014, Future of Software Engineering track)*, pages 167–181, New York, NY, USA, 2014. ACM. *cited page 6*

Thomas L. Griffiths, Charles Kemp, and Joshua B. Tenenbaum. Bayesian models of cognition. In Ron Sun, editor, *The Cambridge Handbook of Computational Psychology*, pages 59–100. Cambridge University Press, New York, USA, 2008. *2 citations pages 73, 74*

Thomas L. Griffiths, Nick Chater, Charles Kemp, Amy Perfors, and Joshua B. Tenenbaum. Probabilistic models of cognition: exploring representations and inductive biases. *Trends in Cognitive Sciences*, 14: 357–364, 2010. *cited page 66*

Thomas L. Griffiths, Nick Chater, Dennis Norris, and Alexandre Pouget. How the Bayesians got their beliefs (and what those beliefs actually are): Comment on Bowers and Davis (2012). *Psychological Bulletin*, 138(3):415–422, 2012a. *cited page 66*

Thomas L. Griffiths, Edward Vul, and Adam N. Sanborn. Bridging levels of analysis for probabilistic models of cognition. *Current Directions in Psychological Science*, 21(4):263–268, 2012b. *2 citations pages 72, 74*

Thomas L. Griffiths, Falk Lieder, and Noah D. Goodman. Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, 7:217–229, 2015. *2 citations pages 72, 74*

Ulrike Hahn. The Bayesian boom: good thing or bad? *Frontiers in Psychology*, 5:765, 2014. *2 citations pages 66, 73*

Ulrike Hahn and Nick Chater. Similarity and rules: distinct? exhaustive? empirically distinguishable? *Cognition*, 65:197–230, 1998. *cited page 81*

Edward Twitchell Hall. *The hidden dimension.* Anchor Books, New York, USA, 1966. *cited page 115*

Christopher M. Harris and Daniel M. Wolpert. Signal-dependent noise determines motor planning. *Nature*, 394:780–784, 1998. *2 citations pages 74, 80*

Leslie A Hayduk. The shape of personal space: An experimental investigation. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 13(1):87–93, 1981. *cited page 115*

Dirk Helbing and Peter Molnar. Social force model for pedestrian dynamics. *Physical Review*, 51:4282–4286, 1995. *cited page 115*

Harold Hill and Alan Johnston. The hollow-face illusion: Object-specific knowledge, general assumptions or properties of the stimulus? *Perception*, 36:199–223, 2007. *cited page 76*

James M. Hillis, Simon J. Watt, Michael S. Landy, and Martin S. Banks. Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, 4(12):967–992, 2004. *cited page 78*

J.M. Hillis, M.O. Ernst, M.S. Banks, and M.S. Landy. Combining sensory information: Mandatory fusion within, but not between, senses. *Science*, 298(5598):1627–1630, 2002. *cited page 78*

Keith J. Holyoak. Induction as model selection. *Proceedings of the National Academy of Sciences*, 105 (31):10637–10638, 2008. *cited page 78*

Keith J. Holyoak and Patricia W. Cheng. Causal learning and inference as a rational process: The new synthesis. *Annual Review of Psychology*, 62:135–163, 2011. *2 citations pages 73, 78*

Kurt Hornik. Approximation capabilities of muitilayer feedforward networks. *Neural Networks*, 4:251–257, 1991. *cited page 80*

Kurt Hornik, Maxwell Stinchcombe, and Halbert White. ?multilayer feedforward networks are universal approximators. *Neural Networks*, 2:359–366, 1989. *cited page 80*

Hiroshi Imamizu, Tomoe Kuroda, Satoru Miyauchi, Toshinori Yoshioka, and Mitsuo Kawato. Modular organization of internal models of tools in the human cerebellum. *Proceedings of the National Academy of Sciences (PNAS)*, 100(9):5461–5466, 2003. *cited page 77*

Arthur M. Jacobs and Jonathan Grainger. Models of visual word recognition – sampling the state of the art. *Journal of Experimental Psychology: Human Perception and Performance*, 20(6):1311–1334, 1994. *cited page 81*

Robert A. Jacobs. Optimal integration of texture and motion cues to depth. *Vision Research*, 39(21): 3621–3629, 1999. *cited page 78*

Robert A. Jacobs. What determines visual cue reliability? *Trends in Cognitive Sciences*, 6(8):345–350, 2002. *cited page 78*

Robert A. Jacobs and John K. Kruschke. Bayesian learning theory applied to human cognition. *WIREs Cognitive Science*, 2010. *2 citations pages 73, 74*

Karin H. James and Isabel Gauthier. When writing impairs reading: letter perception's susceptibility to motor interference. *Journal of Experimental Psychology: General*, 138(3):416–431, 2009. *cited page 34*

Edwin T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106(4):620–630, 1957. *cited page 87*

Edwin T. Jaynes. How does the brain do plausible reasoning? In G. J. Erickson and C. R. Smith, editors, *Maximum-Entropy and Bayesian Methods in Science and Engineering 1*, pages 1–24, Dordrecht, 1988. Kluwer Academic Publishers. *2 citations pages 6, 65*

Edwin T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge, UK, June 2003. Edited by G. Larry Bretthorst. *4 citations pages 5, 6, 65, 84*

Matt Jones and Brad Love. Bayesian fundamentalism or enlightenment? on the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, 34:169–231, 2011. *2 citations pages 66, 73*

Michael I. Jordan and Daniel M. Wolpert. Computational motor control. In Michael S. Gazzaniga, editor, *The Cognitive Neurosciences*, pages 601–620. MIT Press, Cambridge, MA, 1999. *cited page 74*

Reinhart Jürgens and Wolfgang Becker. Perception of angular displacement without landmarks: evidence for Bayesian fusion of vestibular, optokinetic, podokinesthetic, and cognitive information. *Experimental Brain Research*, 174:528–543, 2006. *cited page 78*

Daniel Kahneman and Amos Tversky. Prospect theory: an analysis of decision under risk. *Econometrica*, 47(2):263–292, 1979. *cited page 80*

Mitsuo Kawato. Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, 9:718–727, 1999. *cited page 80*

C. Kemp and J.B. Tenenbaum. The discovery of structural form. *Proc. Natl. Acad. Sci. USA*, 105(31): 10687–10692, 2008. *cited page 78*

D. Kersten, P. Mamassian, and A. Yuille. Object perception as Bayesian inference. *Annual Review of Psychology*, 55:271–304, 2004. *3 citations pages 73, 74, 76*

Daniel Kersten and Alan Yuille. Bayesian models of object perception. *Current Opinion in Neurobiology*, 13:150–158, 2003. *3 citations pages 69, 74, 76*

Kristian Kersting and Luc De Raedt. Bayesian logic programming: Theory and tool. In Lise Getoor and Ben Taskar, editors, *Introduction to Statistical Relational Learning*, pages 291–321. MIT Press, Cambridge, MA, 2007. *cited page 23*

Dave F. Kleinschmidt and T. Florian Jaeger. Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, 122(2):148–203, 2015. *4 citations pages 69, 74, 78, 82*

David Knill and Whitman Richards. *Perception as Bayesian Inference*. Cambridge University Press, Cambridge, UK, 1996. *cited page 74*

David C. Knill. Ideal observer perturbation analysis reveals human strategies for inferring surface orientation from texture. *Vision Research*, 38:2635–2656, 1998. *cited page 78*

Carla Cavalcante Koike, Pierre Bessière, and Emmanuel Mazer. Bayesian approach to action selection and attention focusing. In Pierre Bessière, Christian Laugier, and Roland Siegwart, editors, *Probabilistic Reasoning and Decision Making in Sensory-Motor Systems*, volume 46 of *Springer Tracts in Advanced Robotics*, pages 177–201. Springer-Verlag, Berlin Heidelberg, 2008. *cited page 6*

Daphne Koller and Avi Pfeffer. Learning probabilities for noisy first-order rules. In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence (IJCAI-97)*, pages 1316–1323, Nagoya, Japan, 1997a. *cited page 23*

Daphne Koller and Avi Pfeffer. Object-oriented Bayesian networks. In *Proceedings of the Thirteenth conference on Uncertainty in Artificial Intelligence*, pages 302–313. Morgan Kaufmann Publishers Inc., 1997b. *cited page 23*

Konrad P. Körding and Joshua B. Tenenbaum. Causal inference in sensorimotor integration. In Bernhard Schölkopf, John Platt, and Thomas Hofmann, editors, *Advances in Neural Information Processing Systems 19 (NIPS)*, pages 737–744, Cambridge, MA, 2006. MIT Press. *cited page 78*

Konrad P. Körding and Daniel M. Wolpert. Bayesian integration in sensorimotor learning. *Nature*, 427: 244–247, 2004. *cited page 78*

Konrad P. Körding, Ulrik Beierholm, Wei Ji Ma, Steven Quartz, Joshua B. Tenenbaum, and Ladan Shams. Causal inference in multisensory perception. *PLoS one*, 2(9):e943, 2007. *cited page 78*

Konrad Paul Kording. Bayesian statistics: relevant for the brain? *Current Opinion in Neurobiology*, 25: 130–133, 2014. *cited page 73*

John K. Kruschke. Doing bayesian data analysis. Personal blog post, dated Wednesday, October 26, 2011 [Accessed February 23, 2015], 2011. *cited page 73*

John K. Kruschke. *Doing Bayesian Data Analysis – A tutorial with R, JAGS, and Stan.* Academic Press, Waltham, MA, 2nd edition, 2015. *cited page 73*

Johan Kwisthout, Todd Wareham, and Iris van Rooij. Bayesian intractability is not an ailment that approximation can cure. *Cognitive Science*, 35:779–784, 2011. *cited page 17*

Jean Laurens and Jacques Droulez. Bayesian processing of vestibular information. *Biological Cybernetics*, 96:389–404, 2007. *2 citations pages 7, 76*

Raphaël Laurent. *COSMO : un modèle bayésien des interactions sensori-motrices dans la perception de la parole.* Thèse de doctorat, Université de Grenoble, Octobre 2014. *2 citations pages 43, 52*

Raphaël Laurent, Jean-Luc Schwartz, Pierre Bessière, and Julien Diard. COSMO, un modèle bayésien de la communication parlée : application à la perception des syllabes (COSMO, a Bayesian model of speech communication, applied to syllable perception) [in french]. In *Actes de la conférence conjointe JEP-TALN-RECITAL 2012, volume 1: JEP*, pages 305–312, Grenoble, France, June 2012. ATALA/AFCP. *cited page 52*

Raphaël Laurent, Clément Moulin-Frier, Pierre Bessière, Jean-Luc Schwartz, and Julien Diard. Integrate, yes, but *what* and *how*? A computational approach of sensorimotor fusion in speech perception. *Behavioral and Brain Sciences (commentary)*, 36(4):364–365, 2013a. *cited page 52*

Raphaël Laurent, Jean-Luc Schwartz, Pierre Bessière, and Julien Diard. A computational model of perceptuo-motor processing in speech perception: learning to imitate and categorize synthetic CV syllables. In *Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013)*, pages 2797–2801, Lyon, France, 2013b. *cited page 52*

Ronan Le Hy, Anthony Arrigoni, Pierre Bessière, and Olivier Lebeltel. Teaching Bayesian behaviours to video game characters. *Robotics and Autonomous Systems*, 47:177–185, 2004. *cited page 7*

Olivier Lebeltel. *Programmation Bayésienne des Robots*. Ph.D. thesis, Institut National Polytechnique de Grenoble, Grenoble, France, Septembre 1999. *cited page 6*

Olivier Lebeltel, Julien Diard, Pierre Bessière, and Emmanuel Mazer. A Bayesian framework for robotic programming. In Ali Mohammad-Djafari, editor, *Twentieth International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (Maxent 2000)*, pages 625–637, Melville, New-York, USA, 2000. American Institute of Physics Conference Proceedings. *cited page 6*

Olivier Lebeltel, Pierre Bessière, Julien Diard, and Emmanuel Mazer. Bayesian robot programming. *Autonomous Robots*, 16(1):49–79, 2004. *2 citations pages 6, 11*

Michael D. Lee. Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin & Review*, 15(1):1–15, 2008. *cited page 73*

Jan Charles Lenk and Claus Möbus. Modeling lateral and longitudinal control of human drivers with multiple linear regression models. In Vincent Duffy, editor, *Advances in Applied Digital Human Modeling*, Advances in Human Factors and Ergonomics Series, pages 446–456, Boca Raton, Florida, 2010. CRC Press, Taylor & Francis Group. *cited page 7*

Jorge Lobo, João Filipe Ferreira, and Jorge Dias. Robotic implementation of biological Bayesian models towards visuo-inertial image stabilization and gaze control. In *Robotics and Biomimetics, 2008. ROBIO 2008. IEEE International Conference on*, pages 443–448. IEEE, 2009. *cited page 6*

Gerald E. Loeb. Optimal isn't good enough. *Biological Cybernetics*, 106:757–765, 2012. *cited page 80*

Gerald E. Loeb, Ian E. Brown, and Ernest J. Cheng. A hierarchical foundation for models of sensorimotor control. *Experimental Brain Research*, 126:1–18, 1999. *cited page 81*

M. Longcamp, T. Tanskanen, and R. Hari. The imprint of action: Motor cortex involvement in visual perception of handwritten letters. *NeuroImage*, 33:681–688, 2006. *cited page 33*

Marieke Longcamp, Jean-Luc Anton, Muriel Roth, and Jean-Luc Velay. Visual presentation of single letters activates a premotor area involved in writing. *NeuroImage*, 19:1492–1500, 2003. *cited page 33*

Jean Lorenceau. Cursive writing with smooth pursuit eye movements. *Current Biology*, 22:1506–1509, 2012. *cited page 116*

Bradley C. Love. The algorithmic level is the bridge between computation and brain. *Topics in Cognitive Science*, 7:230–242, 2015. *2 citations pages 67, 72*

Wei Ji Ma. Organizing probabilistic models of perception. *Trends in Cognitive Sciences*, 16(10):511–518, 2012. *cited page 73*

Wei Ji Ma and Mehrdad Jazayeri. Neural coding of uncertainty and probability. *Annual Review of Neuroscience*, 37:205–20, 2014. *cited page 74*

Shinji Maeda. Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In W. J. Hardcastle and A. Marchal, editors, *Speech production and speech modelling*, pages 131–149. Kluwer Academic Publishers, 1990. *cited page 46*

Pascal Mamassian and Ross Goutcher. Prior knowledge on the illumination position. *Cognition*, 81: B1–B9, 2001. *cited page 75*

Pascal Mamassian and Michael S. Landy. Observer biases in the 3D interpretation of line drawings. *Vision Research*, 38:2817–2832, 1998. *cited page 76*

Pascal Mamassian, David C. Knill, and Daniel Kersten. The perception of cast shadows. *Trends in Cognitive Sciences*, 2(8):288–295, 1998. *cited page 76*

Pascal Mamassian, Michael Landy, and Laurence T. Maloney. Bayesian modelling of visual perception. In R. P. N. Rao, B. A. Olshausen, and M. S. Lewicki, editors, *Probabilistic Models of the Brain: Perception and Neural Function*, pages 13–36. MIT Press, Cambridge, MA, 2003. *cited page 76*

Gary F. Marcus and Ernest Davis. How robust are probabilistic models of higher-level cognition? *Psychological Science*, 24(12):2351–2360, 2013. *2 citations pages 66, 80*

David Marr. Artificial intelligence – a personal view. *Artificial Intelligence*, 9(1):37–48, 1977. *2 citations pages 67, 72*

David Marr. *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*. W.H. Freeman and Company, New York, USA, 1982. *4 citations pages 5, 67, 68, 70*

David Marr and Tomaso Poggio. From understanding computation to understanding neural circuitry. A.I. Memo 357, MIT-AIL, 1976. *cited page 67*

Jose Marroquin, Sanjoy Mitter, and Tomaso Poggio. Probabilistic solution of ill-posed problems in computational vision. *Journal of the American Statistical Association*, 82(397):76–89, 1987. *cited page 74*

James L. McClelland. Integrating probabilistic models of perception and interactive neural networks: a historical and tutorial review. *Frontiers in Psychology*, 4:503, 2013. *cited page 81*

James L. McClelland, Matthew M. Botvinick, David C. Noelle, David C. Plaut, Timothy T. Rogers, Mark S. Seidenberg, and Linda B. Smith. Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences*, 14:348–356, 2010. *cited page 66*

Kamel Mekhnacha. *Méthodes Probabilistes Bayésiennes pour la prise en compte des incertitudes géométriques : Application à la CAO-Robotique*. PhD thesis, Institut National Polytechnique de Grenoble, 1999. *cited page 7*

Kamel Mekhnacha, Emmanuel Mazer, and Pierre Bessiere. The design and implementation of a Bayesian CAD modeler for robotic applications. *Advanced Robotics*, 15(1):45–69, 2001. *cited page 7*

Kamel Mekhnacha, Juan-Manuel Ahuactzin, Pierre Bessière, Emmanuel Mazer, and Linda Smail. Exact and approximate inference in ProBT. *Revue d'Intelligence Artificielle*, 21(3):295–331, 2007. *cited page 18*

Brian Milch and Stuart Russell. First-order probabilistic languages: Into the unknown. In *Inductive Logic Programming*, pages 10–24. Springer, 2007. *2 citations pages 22, 23*

Claus Möbus and Mark Eilers. Further steps towards driver modeling according to the Bayesian Programming approach. In *Digital Human Modeling (HCI 2009)*, Lecture Notes in Computer Science (LNCS) and Lecture Notes in Artificial Intelligence (LNAI), pages 413–422. Springer, 2009. *cited page 7*

Claus Möbus and Mark Eilers. Integrating anticipatory competence into a Bayesian driver model. In Springer, editor, *HMAT 2010 (Human Modeling in Assisted Transportation) Workshop*, Heidelberg, 2010a. *cited page 7*

Claus Möbus and Mark Eilers. Mixture-of-behaviors and levels-of-expertise in a Bayesian autonomous driver model. In Vincent Duffy, editor, *Advances in Applied Digital Human Modeling*, Advances in Human Factors and Ergonomics Series, pages 425–435, Boca Raton, Florida, 2010b. CRC Press, Taylor & Francis Group. *cited page 7*

Claus Möbus, Mark Eilers, Hilke Garbe, and Malte Zilinski. Probabilistic and empirical grounded modeling of agents in (partial) cooperative traffic scenarios. In *Digital Human Modeling (HCI 2009)*, Lecture Notes in Computer Science (LNCS) and Lecture Notes in Artificial Intelligence (LNAI), pages 423–432. Springer, 2009. *cited page 7*

Yaniv Morgenstern, Richard F. Murray, and Laurence R. Harris. The human visual system's assumption that light comes from above is weak. *Proceedings of the National Academy of Sciences*, 108(30):12551–12553, 2011. *cited page 76*

Clément Moulin-Frier. *Rôle des relations perception-action dans la communication parlée et l'émergence des systèmes phonologiques : étude, modélisation computationnelle et simulations.* Thèse de doctorat, Université de Grenoble, June 2011. *2 citations pages 43, 46*

Clément Moulin-Frier, Jean-Luc Schwartz, Julien Diard, and Pierre Bessière. Emergence du langage par jeux déictiques dans une société d'agents sensori-moteurs en interaction. In *Journées d'Etude de la Parole*, 2008. *cited page 46*

Clément Moulin-Frier, Jean-Luc Schwartz, Julien Diard, and Pierre Bessière. A unified theoretical Bayesian model of speech communication. In Vincent Duffy, editor, *Advances in Applied Digital Human Modeling*, Advances in Human Factors and Ergonomics Series, pages 457–466, Boca Raton, Florida, 2010. CRC Press, Taylor & Francis Group. *2 citations pages 43, 46*

Clément Moulin-Frier, Jean-Luc Schwartz, Julien Diard, and Pierre Bessière. Emergence of articulatory-acoustic systems from deictic interaction games in a "vocalize to localize" framework. In Anne Vilain, Jean-Luc Schwartz, Christian Abry, and Jacques Vauclair, editors, *Primate communication and human language: Vocalisations, gestures, imitation and deixis in humans and non-humans*, Advances in Interaction Studies, pages 193–220. John Benjamins Pub. Co., Amsterdam / Philadelphia, PA, 2011. *cited page 46*

Clément Moulin-Frier, Raphaël Laurent, Pierre Bessière, Jean-Luc Schwartz, and Julien Diard. Adverse conditions improve distinguishability of auditory, motor and perceptuo-motor theories of speech perception: an exploratory Bayesian modeling study. *Language and Cognitive Processes*, 27(7–8):1240–1263, 2012. *2 citations pages 43, 52*

Clément Moulin-Frier, Julien Diard, Jean-Luc Schwartz, and Pierre Bessière. COSMO ("Communicating about Objects using Sensory-Motor Operations"): a Bayesian modeling framework for studying speech communication and the emergence of phonological systems. *Journal of Phonetics (special issue "On the cognitive nature of speech sound systems")*, in press. *2 citations pages 43, 46*

Kevin Murphy. *Dynamic Bayesian Networks: Representation, Inference and Learning.* Ph. D. thesis, University of California, Berkeley, Berkeley, CA, July 2002. *2 citations pages 22, 23*

In Jae Myung and Mark A Pitt. Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review*, 4(1):79–95, 1997. *cited page 81*

In Jae Myung and Mark A. Pitt. Model comparison methods. *Methods in Enzymology*, 383:351–366, 2004. *cited page 81*

Marko Nardini, Katarina Begus, and Denis Mareschal. Multisensory uncertainty reduction for hand localization in children and adults. *Journal of Experimental Psychology: Human Perception and Performance*, 39(3):773–787, 2012. *cited page 78*

Dennis Norris. The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, 113(2):327–357, 2006. *cited page 78*

Dennis Norris. Putting it all together: A unified account of word recognition and reaction-time distributions. *Psychological Review*, 116(1):207–219, 2009. *cited page 78*

Dennis Norris. Models of visual word recognition. *Trends in Cognitive Sciences*, 17(10):517–524, 2013. *2 citations pages 73, 78*

Dennis Norris and Sachiko Kinoshita. Perception as evidence accumulation and Bayesian inference: Insights from masked priming. *Journal of Experimental Psychology: General*, 137(3):434–455, 2008. *cited page 78*

Dennis Norris and Sachiko Kinoshita. Reading through a noisy channel: Why there's nothing special about the perception of orthography. *Psychological Review*, 119(3):517–545, 2012. *cited page 78*

Dennis Norris and James M. McQueen. Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115(2):357–395, 2008. *2 citations pages 74, 82*

Mike Oaksford and Nick Chater. An introduction to rational models of cognition. In Mike Oaksford and Nick Chater, editors, *Rational models of cognition*. Oxford University Press, Oxford, 1998. *cited page 69*

Mike Oaksford and Nick Chater. The probabilistic approach to human reasoning. *Trends in Cognitive Sciences*, 5(8):349–357, 2001. *cited page 80*

Mike Oaksford and Nick Chater. Précis of Bayesian rationality: The probabilistic approach to human reasoning. *Behavioral and Brain Sciences*, 32:69–120, 2009. *cited page 74*

Ipek Oruç, Laurence T. Maloney, and Michael S. Landy. Weighted linear cue combination with possibly correlated error. *Vision Research*, 43:2451–2468, 2003. *cited page 78*

Jean-François Patri, Julien Diard, and Pascal Perrier. Optimal speech motor control and token-to-token variability: A Bayesian modeling approach. *Biological Cybernetics*, in press. *3 citations pages 57, 63, 80*

Yohan Payan and Pascal Perrier. Synthesis of V–V sequences with a 2D biomechanical tongue model controlled by the equilibrium point hypothesis. *Speech communication*, 22(2):185–205, 1997. *cited page 59*

David Peebles and Richard P. Cooper. Thirty years after Marr's *Vision*: Levels of analysis in cognitive science. *Topics in Cognitive Science*, 7:197–190, 2015. *cited page 67*

Amy Perfors, Joshua B. Tenenbaum, Thomas L. Griffiths, and Fei Xu. A tutorial introduction to bayesian models of cognitive development. *Cognition*, 120:302–321, 2011. *cited page 73*

Pascal Perrier, Yohan Payan, Majid Zandipour, and Joseph Perkell. Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study. *The Journal of the Acoustical Society of America*, 114(3):1582–1599, 2003. *cited page 59*

Pascal Perrier, Liang Ma, and Yohan Payan. Modeling the production of vcv sequences via the inversion of a biomechanical model of the tongue. In *Proceedings of InterSpeech 2005*, pages 1041–1044, Lisbon, Portugal, 2005. *cited page 58*

William A. Phillips. Self-organized complexity and coherent infomax from the viewpoint of jaynes's probability theory. *Information*, 3:1–15, 2012. *cited page 12*

Mark A. Pitt, In Jae Myung, and Shaobo Zhang. Toward a method of selecting among computational models of cognition. *Psychological Review*, 109(3):472–491, 2002. *cited page 81*

Tomaso Poggio. Marr's computational approach to vision. *Trends in Neurosciences*, 4(10):258–262, 1981. *2 citations pages 67, 71*

Emmanuel M. Pothos and Jerome R. Busemeyer. Can quantum probability provide a new direction for cognitive modeling? *Behavioral and Brain Sciences*, 36(3):255–274, 2013. *cited page 85*

Emmanuel M. Pothos and Jerome R. Busemeyer. In search for a standard of rationality. *Frontiers in Psychology*, 5:49, 2014. *cited page 85*

Alexandre Pouget, Jeffrey M Beck, Wei Ji Ma, and Peter E Latham. Probabilistic brains: knowns and unknowns. *Nature Neuroscience*, 16(9):1170–1178, 2013. *cited page 74*

Cédric Pradalier, Francis Colas, and Pierre Bessière. Expressing Bayesian fusion as a product of distributions: Applications in robotics. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS03)*, volume 2, pages 1851–1856, 2003. *cited page 20*

Cédric Pradalier, Jorge Hermosillo, Carla Koike, Christophe Braillon, Pierre Bessière, and Christian Laugier. The *CyCab*: a car-like robot navigating autonomously and safely among pedestrians. *Robotics and Autonomous Systems*, 50(1):51–68, 2005. *cited page 7*

Vilayanur S. Ramachandran. Perceiving shape from shading. *Scientific American*, 1988. *cited page 75*

Guy Ramel and Roland Siegwart. Probabilistic contextual situation analysis. In Pierre Bessière, Christian Laugier, and Roland Siegwart, editors, *Probabilistic Reasoning and Decision Making in Sensory-Motor Systems*, volume 46 of *Springer Tracts in Advanced Robotics*, pages 129–151. Springer-Verlag, Berlin Heidelberg, 2008. *cited page 7*

Christian P. Robert. *The Bayesian Choice – From Decision-Theoretic Foundations to Computational Implementation*. Springer, New York, NY, USA, 2nd edition edition, 2007. *cited page 80*

Sam Roweis and Zoubin Ghahramani. A unifying review of linear gaussian models. *Neural Computation*, 11(2):305–345, February 1999. *cited page 22*

Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall Series in Artificial Intelligence, Englewood Cliffs, New Jersey, 1995. *2 citations pages 29, 71*

Adam N. Sanborn, Thomas L. Griffiths, and Daniel J. Navarro. Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review*, 117(4):1144–1167, 2010. *cited page 72*

Yoshiyuki Sato, Taro Toyoizumi, and Kazuyuki Aihara. Bayesian inference explains perception of unity and ventriloquism aftereffect: Identification of common sources of audiovisual stimuli. *Neural Computation*, 19(12):3335–3355, 2007. *cited page 78*

Paul Schrater and Daniel Kersten. Vision, psychophysics and Bayes. In R. Rao, B. Olshausen, and M. Lewicki, editors, *Probabilistic Models of the Brain*, chapter 2. MIT Press, 2000. *cited page 74*

Jihane Serkhane, Jean-Luc Schwartz, and Pierre Bessière. Building a talking baby robot: A contribution to the study of speech acquisition and evolution. *Interaction Studies*, 6(2):253–286, 2005. *cited page 7*

Ladan Shams and Ulrik R. Beierholm. Causal inference in perception. *Trends in Cognitive Sciences*, 14: 425–432, 2010. *cited page 78*

Lei Shi and Thomas L. Griffiths. Neural implementation of hierarchical bayesian inference by importance sampling. In Y. Bengio, D. Schuurmans, J.D. Lafferty, C.K.I. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems 22 (NIPS 2009)*, pages 1669–1677. Curran Associates, Inc., 2009. *cited page 72*

Lei Shi, Thomas L. Griffiths, Naomi H. Feldman, and Adam N. Sanborn. Exemplar models as a mechanism for performing Bayesian inference. *Psychonomic Bulletin & Review*, 17(4):443–464, 2010. *2 citations pages 72, 81*

Éva Simonin, Julien Diard, and Pierre Bessière. Learning Bayesian models of sensorimotor interaction: from random exploration toward the discovery of new behaviors. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS05)*, pages 1226–1231, 2005. *2 citations pages 6, 111*

Philippe Smets. The transferable belief model and other interpretations of Dempster-Shafer's model. In P. P. Bonissone, M. Henrion, L.N. Kanal, and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, volume 6, pages 375–383. Elsevier Science Publishers, 1991. *cited page 85*

Morgan Sonderegger and Alan Yu. A rational account of perceptual compensation for coarticulation. In *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society (CogSci10)*, pages 375–380, 2010. *3 citations pages 69, 74, 82*

Olaf Sporns, Giulio Tononi, and Rolf Kötter. The human connectome: A structural description of the human brain. *PLoS Computational Biology*, 1(4):e42, 2005. *cited page 67*

Brian J. Stankiewicz, Gordon E. Legge, J. Stephen Mansfield, and Erik J. Schlicht. Lost in virtual space: Studies in human and ideal spatial navigation. *Journal of Experimental Psychology: Human Perception and Performance (submitted to)*, 32(3):688–704, 2005. *cited page 69*

Alan A. Stocker and Eero P. Simoncelli. Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, 9(4):578–585, 2006. *cited page 76*

Ron Sun. Introduction to computational cognitive modeling. In Ron Sun, editor, *The Cambridge Handbook of Computational Psychology*, pages 3–19. Cambridge University Press, New York, USA, 2008. *2 citations pages 67, 72*

Ron Sun, L. Andrew Coward, and Michael J. Zenzen. On levels of cognitive modeling. *Philosophical Psychology*, 18(5):613–637, 2005. *2 citations pages 72, 73*

Gabriel Synnaeve and Pierre Bessière. A Bayesian model for RTS units control applied to StarCraft. In *Computational Intelligence and Games (CIG), 2011 IEEE Conference on*, pages 190–196. IEEE, 2011a. *cited page 7*

Gabriel Synnaeve and Pierre Bessière. A Bayesian model for opening prediction in RTS games with application to StarCraft. In *Computational Intelligence and Games (CIG), 2011 IEEE Conference on*, pages 281–288. IEEE, 2011b. *cited page 7*

Gabriel Synnaeve and Pierre Bessière. A Bayesian tactician. In *Computer Games Workshop at ECAI 2012*, pages 114–125, Montpellier, France, 2012a. *cited page 7*

Gabriel Synnaeve and Pierre Bessière. Special tactics: A Bayesian approach to tactical decision-making. In *Computational Intelligence and Games (CIG), 2012 IEEE Conference on*, pages 409–416. IEEE, 2012b. *cited page 7*

Richard Szeliski. Bayesian modeling of uncertainty in low-level vision. *International Journal of Computer Vision*, 5(3):271–301, 1990. *cited page 74*

Adriana Tapus and Roland Siegwart. Topological SLAM. In Pierre Bessière, Christian Laugier, and Roland Siegwart, editors, *Probabilistic Reasoning and Decision Making in Sensory-Motor Systems*, volume 46 of *Springer Tracts in Advanced Robotics*, pages 99–127. Springer-Verlag, Berlin Heidelberg, 2008. *cited page 7*

M.K. Tay, Kamel Mekhnacha, Cheng Chen, Manuel Yguel, and Christian Laugier. An efficient formulation of the Bayesian occupation filter for target tracking in dynamic environments. *International Journal of Vehicle Autonomous Systems*, 6(1-2):155–171, 2008. *cited page 7*

Joshua B. Tenenbaum, Charles Kemp, Thomas L. Griffiths, and Noah D. Goodman. How to grow a mind: Statistics, structure, and abstraction. *Science*, 331:1279–1285, 2011. *cited page 78*

Emanuel Todorov. Optimality principles in sensorimotor control. *Nature Neuroscience*, 7(9):907–915, 2004. *cited page 74*

James T Townsend. Theoretical analysis of an alphabetic confusion matrix. *Perception & Psychophysics*, 9(1):40–50, 1971. *cited page 37*

Pete C. Trimmer, Alasdair I. Houston, James A. R. Marshall, Mike T. Mendl, Elizabeth S. Paul, and John M. McNamara. Decision-making under uncertainty: biases and Bayesians. *Animal Cognition*, 14:465–476, 2011. *cited page 76*

Amos Tversky and Daniel Kahneman. The framing of decisions and the psychology of choice. *Science*, 211:453–458, 1981. *cited page 80*

Yoji Uno, Mitsuo Kawato, and Rika Suzuki. Formation and control of optimal trajectory in human multijoint arm movement: Minimum torque-change model. *Biological Cybernetics*, 61:89–101, 1989. *cited page 80*

Robert J. van Beers, Anne C. Sittig, and Jan J. Denier van der Gon. How humans combine simultaneous proprioceptive and visual position information. *Experimental Brain Research*, 111:253–261, 1996. *cited page 78*

Robert J. van Beers, Pierre Baraduc, and Daniel M. Wolpert. Role of uncertainty in sensorimotor control. *Phil. Trans. R. Soc. Lond. B*, 357:1137–1145, 2002. *cited page 80*

H. van der Kooij, R. Jacobs, B. Koopman, and Grootenboer H. A multisensory integration model of human stance control. *Biological Cybernetics*, 80:299–308, 1999. *cited page 78*

Kevin S. van Horn. Constructing a logic of plausible inference: a guide to Cox's theorem. *nternational Journal of Approximate Reasoning*, 34:3–24, 2003. *3 citations pages 6, 65, 84*

Iris Vilares, James D. Howard, Hugo L. Fernandes, Jay A. Gottfried, and Konrad P. Kording. Differential representations of prior and likelihood uncertainty in the human brain. *Current Biology*, 22:1641–1648, 2012. *cited page 75*

Benjamin T. Vincent. A tutorial on bayesian models of perception. *Journal of Mathematical Psychology*, 66:103–114, 2015a. *cited page 73*

Benjamin T. Vincent. Bayesian accounts of covert selective attention: A tutorial review. *Atten Percept Psychophys*, 77:1013–1032, 2015b. *cited page 73*

Paolo Viviani and Tamar Flash. Minimum-jerk, two-thirds power law, and isochrony: converging approaches to movement planning. *Journal of Experimental Psychology: Human Perception and Performance*, 21(1):32–53, 1995. *cited page 80*

Edward Vul, Noah D. Goodman, Thomas L. Griffiths, and Joshua B. Tenenbaum. One and done? optimal decisions from very few samples. *Cognitive Science*, 38:599–637, 2014. *cited page 72*

Yair Weiss, Eero P. Simoncelli, and Edward H. Adelson. Motion illusions as optimal percepts. *Nature Neuroscience*, 5:598–604, 2002. *3 citations pages 74, 76, 78*

Daniel M. Wolpert. Probabilistic models in human sensorimotor control. *Human Movement Science*, 26: 511–524, 2007. *cited page 73*

Daniel M. Wolpert and Zoubin Ghahramani. Computational principles of movement neuroscience. *Nature Neuroscience*, 3:1212–1217, 2000. *cited page 74*

Daniel M. Wolpert and Mitsuo Kawato. Multiple paired forward and inverse models for motor control. *Neural Networks*, 11(7-8):1317–1329, 1998. *cited page 77*

Daniel M. Wolpert, Zoubin Ghahramani, and Michael I. Jordan. An internal model for sensorimotor integration. *Science*, 269(5232):1880–1882, 1995. *cited page 78*

David R. Wozny, Ulrik R. Beierholm, and Ladan Shams. Human trimodal perception follows optimal statistical inference. *Journal of Vision*, 8(3):24, 1–11, 2008. *cited page 78*

Ronald R. Yager. On the Dempster-Shafer framework and new combination rules. *Information Sciences*, 41:93–137, 1987. *cited page 85*

Alan Yuille and Daniel Kersten. Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences, special issue: Probabilistic models of cognition*, 10(7), 2006. *cited page 74*

Carlos Zednik and Frank Jäkel. How does Bayesian reverse-engineering work? In *Proceedings of the 36th Annual Conference of the Cognitive Science Society (submitted to)*, 2014. *cited page 66*

L. H. Zupan, D. M. Merfeld, and C. Darlot. Using sensory weighting to model the influence of canal, otolith and visual cues on spatial orientation and eye movements. *Biological Cybernetics*, 86:209–230, 2002. *cited page 78*

APPENDIX A

Other projects: bullet list summaries

In this annex, I quickly describe projects that did not fit the main narrative of this manuscript, providing entry points, when available, to more detailed publications.

A.1 Bayesian modeling of robotic navigation

Collaborators: Pierre Bessière (ISIR, Paris), Emmanuel Mazer (LIG, Grenoble)
Supervised students: Eva Simonin (Master 2), Estelle Gilet (Master 1)
Publications: Diard et al. (2010a), Diard and Bessière (2008), Diard et al. (2005), Simonin et al. (2005), Diard et al. (2004a,b, 2003a), Diard (2003)

- In robotic navigation, we proposed to marry probabilistic mapping and bio-inspired mapping methods by defining maps as structured probabilistic representations of space. In this proposition, maps are modular, with building blocks called *Bayesian maps*, which are combined using *Bayesian map operators*.
- A Bayesian map π is a joint probability distribution $P(P_t\ L_t\ L_{t+1}\ A_t \mid \pi)$, with P_t a perception variable, L_t and L_{t+1} internal (location) variables at successive time steps, and A_t an action variable. Bayesian maps solve navigation tasks, using Bayesian inference, in the form of *elementary behaviors*, i.e., probabilistic questions of the form $P(A_t \mid X)$, with X a subset of variables of the Bayesian map.
- We defined three Bayesian map operators: the *Superposition* of Bayesian maps, the *Abstraction* of Bayesian maps, and the *Sensorimotor interaction* of Bayesian maps. They take Bayesian maps as input, and output new, well-defined Bayesian maps.
- In the Superposition operator, the internal space of the resulting map is the conjunction of internal spaces of input maps. Observation of the elementary behaviors of input maps allows to measure the "information collinearity" of their internal spaces, that is to say, detecting whether input maps are redundant (e.g., superposition of parallel 1D gradients) or not (e.g., superposition of orthogonal 1D gradients yielding grid-like maps).

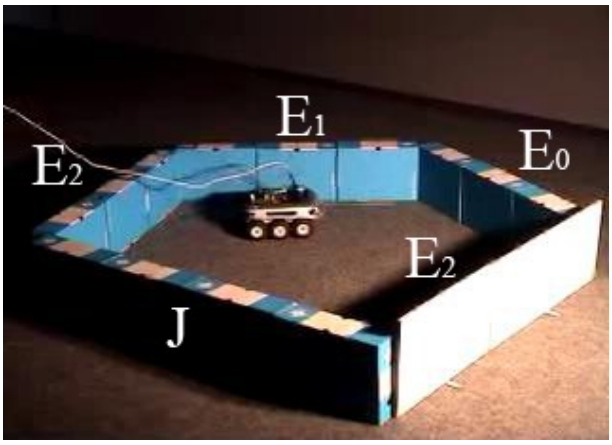

Figure A.1: **Illustration of the Sensorimotor Interaction of Bayesian Maps.** In an environment made of boxes and planks, a Koala mobile robot identifies the large-scale structure of the arena by identifying the relationship between wall-following and light sensing.

- In the Abstraction operator, the perception variable of the resulting, high-level map is the set of all variables of input, low-level maps. The internal variable of the resulting map is a set of symbols referring to each input map. The action variable of the resulting map is a set of symbols referring to low-level elementary behaviors. In this sense, the resulting map "observes" input maps using Bayesian model comparison, denotes locations as portions of the environment where they are good Bayesian models of sensorimotor interaction, and constructs high-level behaviors for navigating between abstract locations.

- In the Sensorimotor Interaction operator, input maps have the same action variable, and the resulting map observes the effect of the application of elementary behaviors on internal variables of input maps. In this sense, the resulting identify parts of the environment where input maps interact, during sensorimotor interaction with the environment, in a recognizable manner (these are the resulting map internal variable). In other words, it is perceptuo-motor fusion of modalities using action-oriented sensor models.

- In a robotic experiment, a complete hierarchy of Bayesian maps was programmed and incrementally learned by a Koala mobile robot. At the top of this hierarchy, a large-scale Bayesian map was obtained by Sensorimotor interaction of a proximity-based and a light-based Bayesian map. The larges-scale Bayesian map relied on implicit perception of the angle of obstacles relative to the light source, by the contingency of light sensing during wall-following due to cast shadows. This is illustrated Figure A.1.

A.2 Bayesian modeling of human navigation

Collaborators: Alain Berthoz (LPPA, Paris), Pierre Bessière (ISIR, Paris), Panagiota Panagiotaki (LPPA, Paris)
Publications: Diard et al. (2013a, 2009)

- In human navigation, we developed a Bayesian model of path memorization based on circular probability distributions, such as the von Mises probability distribution, which are the correct analogue of Gaussian probability distributions over circular spaces. von Mises probability distributions are defined by μ a central tendency parameter, and λ a concentration parameter. They are illustrated Figure A.2.

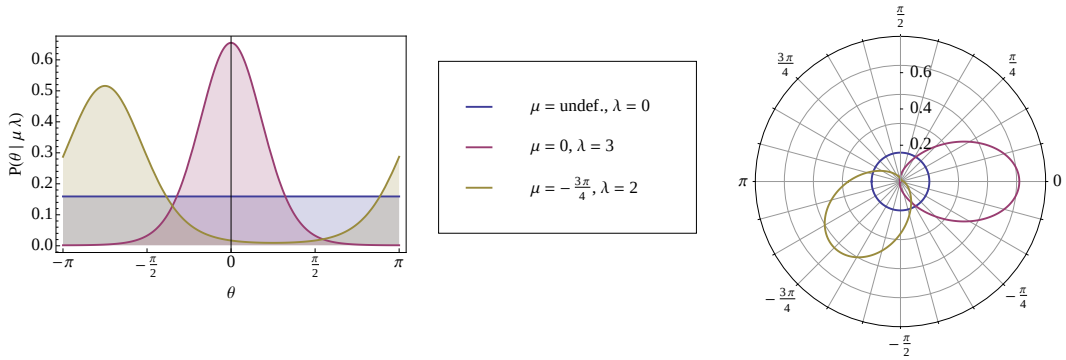

Figure A.2: **Examples of von Mises probability distributions for various μ, λ parameter values. Left:** on a linear scale on the $(-\pi, \pi]$ interval. **Right:** polar representation of the same distributions.

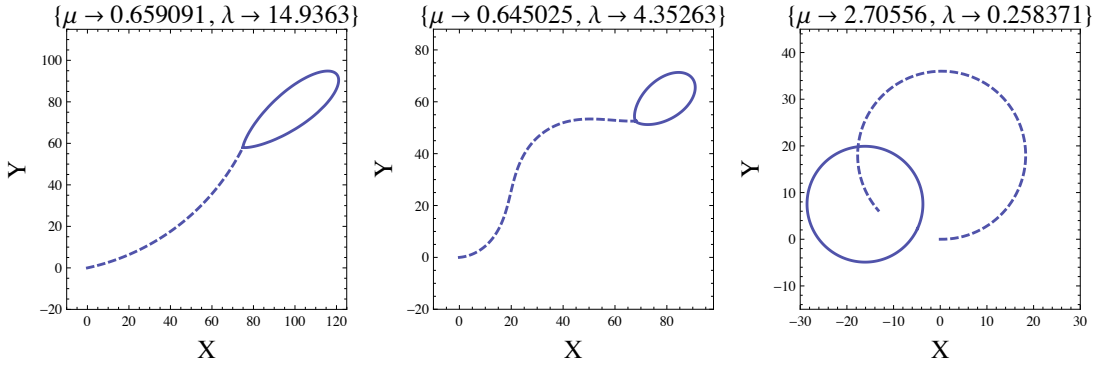

Figure A.3: **Path integration using von Mises probability distributions.** Each plot shows a path in a X, Y plane, starting from $(0, 0)$. Superposed to the path end points are polar representations of the von Mises learned along the path. Identification of the μ parameter provides a vector between the start and end point of the path; reversing this yields a homing vector. Identification of the λ parameter implicitly estimates path sinuosity. **Left:** small path sinuosity yields large λ. **Right:** large path sinuosity yields small λ. **Middle:** intermediary case.

- In this model, a navigator identifies the parameters of a von Mises probability distribution, from orientations experienced along a path. It was already known that this provided the navigator an angle toward the starting point of the path (by computing $\mu + \pi$). We have further demonstrated that this also provided the navigator an implicit measure of the distance D to the starting point of the path, provided path length N was measured (λ is is bijection with D/N). This yields "probabilistic" path integration (illustrated Figure A.3).

- We have demonstrated that, in memory models of paths, probabilistic representations and deterministic representations were indistinguishable, on theoretical grounds. This leads to comparatively study navigation strategies such as homing and path reproduction, and cue combination, in both the probabilistic and deterministic cases, in the pursuit of discriminative predictions.

- We have studied path memorization and path reproduction when landmarks are present in the environment. Our model predicts that, when landmarks are removed between path memorization and path reproduction, Bayesian inference yields a navigation strategy that goes in the general correct direction, but completely forgets the exact sequence of turns.

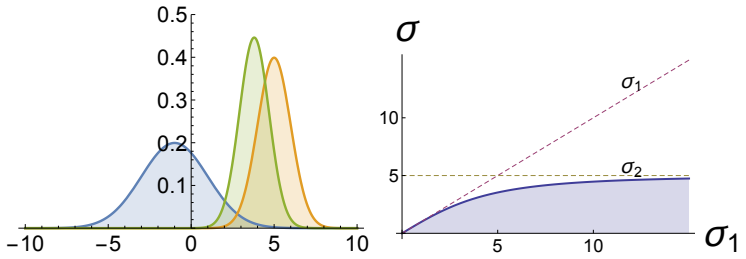

Figure A.4: **Bayesian cue combination in the linear case**, using Gaussian probability distributions, always decreases uncertainty, even in the case of inconsistant cues. **Left:** the blue and yellow Gaussian probability distributions, when combined using the classical sensor fusion model, yield the green distribution, which is intermediate and of smaller variance. **Right:** whatever the initial variances σ_1, σ_2, the final distribution has smaller variance σ.

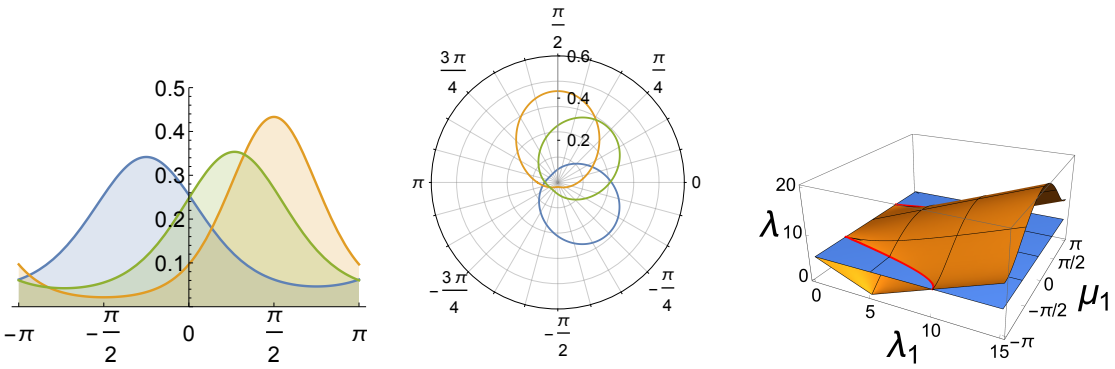

Figure A.5: **Bayesian cue combination in the orientation case**, using von Mises probability distributions, does not systematically decrease uncertainty. **Left, Middle:** combination of the blue and yellow von Mises distribution, which are inconsistent, yields the green distribution, which is more more concentrated than the yellow one. **Right:** concentration λ of the resulting distribution (yellow curve) as a function of the μ_1, λ_1 parameters of one of the combined distribution. (Parameters of the other combined distribution are $\mu_2 = 0$, $\lambda_2 = 5$ (blue plane), without loss of generality.)

This was illustrated in a virtual city experiment, where participants indeed managed to reach the goal in this test condition, but were unaware that they did not correctly reproduce the learned path (contrary to the control condition).

- Finally, we have studied cue combination in the case of orientation cues. We have demonstrated that, contrary to the linear case (Figure A.4), combining inconsistant cues does not always decrease uncertainty in orientation estimation (Figure A.5).

A.3 Bayesian modeling of human proxemics

Collaborators: Anne Spalanzani (LIG, Grenoble), Richard Palluel-Germain (LPNC, Grenoble), Nicolas Morgado (LPNC, Grenoble)
Supervised student: Marie-Lou Barnaud (Master 2)
Publication: Barnaud et al. (2014)

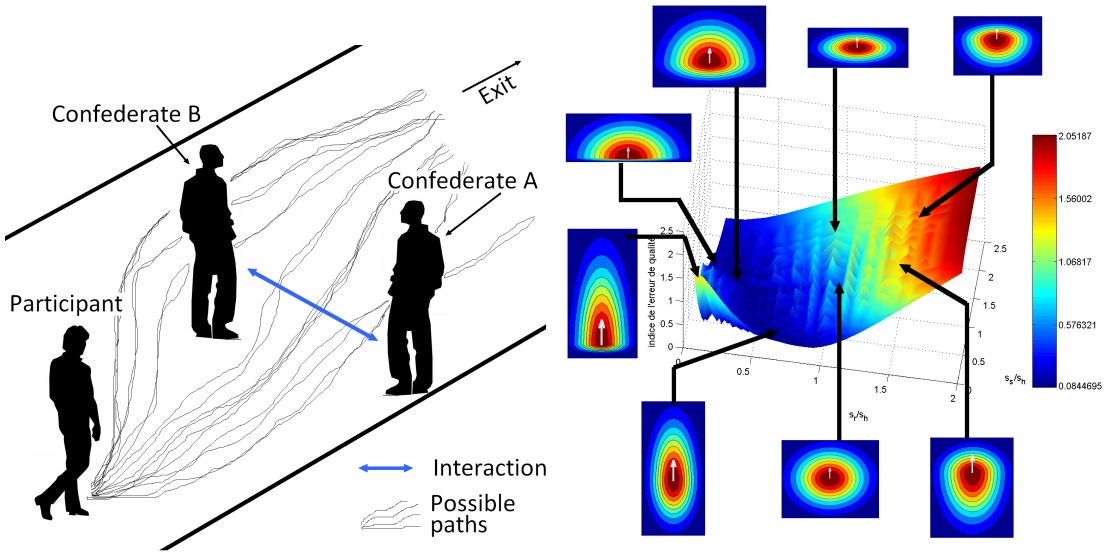

Figure A.6: **Bayesian modeling of personal and interaction social spaces. Left:** Setup of the Efran and Cheyne (1973) experiment. **Right:** Shapes of personal space models that have a good fit to experimental data are asymmetrical, elongated to the front (blue regions in the central plot). Purely circular and other shapes yield poor fit to experimental data (pale blue to red regions in the central plot).

- In human-aware robotic navigation, we have studied the use of social models of personal space and interaction space to constrain navigation strategies. We have proposed to ground such space representations in literature from social and cognitive psychology.

- We have developed a robotic simulation replicating a classical experiment of the domain, from Efran and Cheyne (1973). In this experiment, participants had to walk down a corridor, choosing whether they would pass between two confederates involved in an interaction, or avoid them and pass behind (Figure A.6, left).

- We have developed five model classes of human-aware navigation strategies, relying either on an asymmetric 2D Gaussian model of personal space (PS model), or on a 1D Gaussian model of interaction space (Normal IS model), or on a 1D constant model of interaction space (constant IS model), or on models combining personal and interaction spaces (PS+Normal IS, PS+Constant IS models).

- Model comparison shows that a personal space model is required to adequately fit experimental data. Furthermore, adding an interaction space to the personal space model of course increases complexity, but only marginally improves fit to data. This suggests that human-like robotic navigation can be obtained using only a PS model, at least in situations similar to our experimental conditions.

- Investigation of the parameter space of the PS model disproves the original proposal by Hall (1966) of personal spaces as concentric circles. In contrast, it strongly supports more recent proposals which have suggested asymmetrical shapes (Hayduk; 1981, Helbing and Molnar; 1995) (Figure A.6, right). Our approach could be used as a method for measuring proxemics spaces in social and cognitive psychology experiments.

A.4 Bayesian modeling of eye writing: BAP-EOL

Collaborators: Jean Lorenceau (LSP, Paris)
Supervised post-doc: Myriam Chanceaux (post-doc, LPNC, Grenoble)
Supervised students: Vincent Rynik (Master 1), Svetlana Meyer (Master 2)
Publications: Diard et al. (2013b), Chanceaux et al. (2014)

- In a collaborative project with Jean Lorenceau, we have explored an original object of study, cursive "eye writing" (Lorenceau; 2012). Eye writing is performed thanks to an illusion-inducing visual stimulus, eliciting illusory movement in the direction of eye movement. After training, users are able to generate smooth pursuit eye movements in the absence of visual target, and instead, along free-flow trajectories of their choice. Coupled with an eye-tracking device, eye writing and eye drawing are made possible.

- In this context, we have defined the Bayesian Action-Perception: Eye OnLine model (BAP-EOL), an adaptation of the BAP model (see Section 3.1) to the case of on-line eye-written character recognition. We have demonstrated that character recognition and novelty detection were possible, suggesting a potential use of the eye writing apparatus by motor impaired patients, as a free-form communication device.

- In proof-of-concept experiments, we have also extended the BAP model towards disability assessment, by augmenting the model for opportunistically observing fine-grained motor characteristics, during eye writing. Writer recognition was also performed, and shown to be very robust when based on strings of input letters. Finally, motor equivalence was studied, by defining measures invariant across effector change, and thus, characteristic of motor equivalence, and measures that are, instead, effector specific, such as effector mastery or feedback availability.

A.5 Bayesian modeling of performance: PARSEVAL

Collaborators: Muriel Lobier (Neuroscience Center, Helsinki), Christophe Poyet (Gate1, Grenoble)
Supervised engineer: Thierry Phénix (LPNC), Estelle Braillon (LPNC)
Supervised students: Jérémy Krein (Master 2), Justine Rexer (Master 2), Victor Sabatier (Master 2), Clément Vetroff (Master 1), Layla El Asri (Master 1), Mostapha Mahmoud (Master 1)
Publications: Diard et al. (2010b)
Software registration: PARSEVAL vα, inventors Diard, Lobier & El Asri, 2013, number IDDN.FR.001.470023.000.S.P.2013.000.20700; PARSEVAL v1.0, inventors Diard & Phénix, 2015, number pending.

- This was initiated as a research project, but because of its applicative potential, we then transformed it into an IT transfer project, in order to assess market potential and use cases. Grants and support were provided by GRAVIT Innovation (then Gate1 then GIFT), the IT transfer agency of Grenoble's academic and research entities. That is why, except for an initial, preliminary publication, no further paper was published. Instead, two software

Figure A.7: **Video game demonstrating the PARSEVAL algorithm. Left:** Screenshot of a game in progress. Bubbles appear in waves in varying numbers, move in varying speeds, and the time window where bursting them scores positive points also varies. This defines a 3D difficulty space, indicated at the bottom of the screen. **Right:** Menu screen showing the success rate, over time. In this example, a punitive 30% overall success rate was set, and attained after several minutes of play.

versions of the PARSEVAL algorithm were "registered" to the APP, the French Software registration agency (CNRS seldom patents softwares).

- We developed PARSEVAL, a Probabilistic Algorithm for Real-time Subject Evaluation. The context is multi-dimensional psychophysics for adaptive training and remediation software. The initial domain was a software for visuo-attentional training for dyslexic children.
- PARSEVAL is an algorithm which identifies user performance and performance profile in a multi-dimensional space, tracks user performance evolution over time (characterizing learning and learning speed, if any), and selects exercises, in an adaptive manner, in order to aim for a predefined success rate, balancing learning thanks to failures and motivation thanks to successes.
- PARSEVAL was implemented as a C++ library, and industrial partners were sought in the video game domain. To that end, a video game using PARSEVAL to automatically adapt the game's difficulty to the player's performance (see Figure A.7). The demonstrator was presented at the Game Connection professional expo in Paris (2014), and contacts are pursued.
- Current projects aim at integrating PARSEVAL in a "gamified" and personalized marketing application, with the maturation of the Cognidis start-up company in the lab.

A.6 Bayesian modeling of distinguishability of models: DEAD

Supervised student: Léo Lopez (Master 2)
Publications: Diard (2009)

- This research project concerns the modeling of the experimental-modeling loop in cognitive science, and more precisely the experimental design step. We consider the distinguishability of models, that is, the ability to discriminate between alternative theoretical explanations of experimental data, as a guide for experimental design.

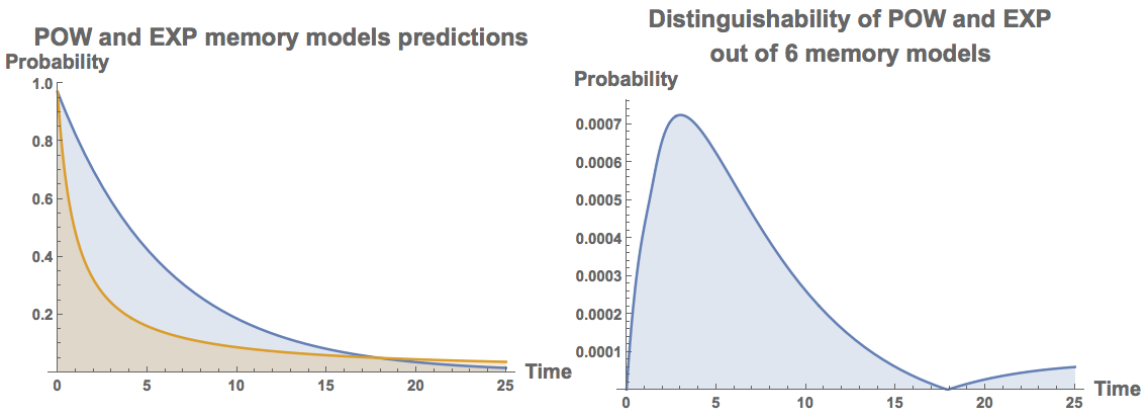

Figure A.8: **Bayesian distinguishability of models illustrated on memory retention. Left:** Prediction of the probability of correct recall p of items as a function of time elapsed t between memorization and recall, for two classes of models, the POW and EXP models. POW models are of the form $p = a(t+1)^{-b}$, EXP are $p = ae^{-bt}$. Here, the POW model has parameters $a = 0.97, b = 1.005$, the EXP model $a = 0.97, b = 0.165$ **Right:** Probability that models are distinguishable as a function of the experimental time set between memorization and recall. The points where both models predicted probabilities cross (at $t = 0$ and around $t = 18$) corresponds to indistinguishability cases.

- Many methods already exist in this domain. Most rely on the classical model comparison meta-model used for parameter identification, which is unsuited to the task. As a consequence, these methods have to "step outside" of the probabilistic framework, either by using the sample space (e.g., generation and cross-learning of virtual data), or by comparing probability values in ad-hoc manner (e.g., Bayes Factor).

- In contrast, we have defined an extension of the Bayesian model selection method that incorporates a measure of model distinguishability. This yields a fully Bayesian meta-model of the distinguishability of models, in which Bayesian inference can be carried out to answer any question, such as "where should the next experiment be to maximize distinguishability between model m_1 and m_2?", "between m_1 with parameters θ_1 and m_2 with parameters θ_2?", "between m_1 and m_2 out of a class of six alternate models?", etc.

- We have illustrated our method on the classical issue of discriminating between memory retention models. This is illustrated Figure A.8.

- We have also integrated our measure of Bayesian distinguishability of models to an adaptive experimental design algorithm. This is our Design of Experiments using Adaptive Distinguishability (DEAD) project. This acronym also stands for the current status of this project, due to lack of time, funding and interested students... This is unfortunate, as all theoretical groundwork has already been established.

APPENDIX B

Curriculum Vitæ

In the following pages, my long CV, in French, can be found.

Julien DIARD — Notice de titres et travaux

Né le 21 Avril 1976 à Saint Martin d'Hères, Isère (38)
Nationalité française – Marié
Adresse personnelle :
18, rue Camille Desmoulins 38400 Saint Martin d'Hères, France
Tél. : (+33) 4 56 00 92 86
E-mail : Julien.Diard@upmf-grenoble.fr
Site : `http://diard.wordpress.com/`

Situation actuelle

2010– Chargé de Recherche (CR1) au CNRS, affecté au LPNC, Grenoble (Laboratoire de Psychologie et NeuroCognition, UMR 5105 CNRS-UPMF-UJF-US).
Thème : **Modélisation Bayésienne de la perception et de l'action.**

Parcours – Diplômes

2005–2010 Chargé de Recherche (CR2) au CNRS, affecté au LPNC, Grenoble.

2004–2005 ATER à l'Université Joseph Fourier (Grenoble 1)
Recherches post-doctorales au laboratoire GRAVIR (Grenoble).

2003–2004 Séjours post-doctoraux à la National University of Singapore (Singapour) et au Laboratoire de Physiologie de la Perception et de l'Action (Collège de France, Paris).

2004 Qualifié aux fonctions de maître de conférences en 27e section (Informatique)

2003 Docteur et Moniteur en Informatique, Universités de Grenoble

1999 DEA (Bien) et Magistère (Très Bien) en Informatique, Université Grenoble 1

1998 Maîtrise d'Informatique (Très Bien) en mobilité internationale à UCSB

Publications Au 01/10/15.

Catégorie	Publiés, *in press*	
Articles de revues internationales	10	38 textes *peer-reviewed*
Articles de revues nationales / francophones	2	
Chapitres d'ouvrages collectifs	3	
Conférences internationales avec comité de lecture (article)	17	
Colloques internationaux avec comité de lecture	3	
Conférences francophones avec comité de lecture	3	
Conf. internationales avec comité de lecture (poster ou abstract)	3	
Colloques internationaux sans comité de lecture	3	
Rapports techniques	13	

Enseignements 573 heures équivalents TD (CM M2R, ATER, monitorat)

Année	Type	Public	Volume (eqTD)
2009–2015	CM	Master 2R	36 + 31,5 + 36 + 27 + 27 + 18 + 13,5
2004–2005	CM, TD, TP	Licence 1, 2, 3	192
1999–2002	CM, TD, TP	Licence	192

Encadrement d'étudiants (en cours)

Niveau d'études	Nb
Doctorat	4

Encadrement d'étudiants (précédemment)

Niveau d'études	Nb
Post-doctorat	1
Doctorat	3
Bac+5 : Master 2R, 3^e année d'école ingénieur, DEA	13
Bac+4 : Master 1, Magistère 2, 2^e année d'école ingénieur, FYP	11
Bac+3 : Licence 3	1

Situation actuelle

2010– Chargé de Recherche (CR1) au CNRS, affecté au LPNC, Grenoble (Laboratoire de Psychologie et NeuroCognition, UMR 5105 CNRS-UPMF-UJF-US).
Thème : **Modélisation Bayésienne de la perception et de l'action.**

Formation – Diplômes

2005–2010 Chargé de Recherche (CR2) au CNRS, affecté au LPNC, Grenoble (Laboratoire de Psychologie et NeuroCognition, UMR 5105 CNRS-UPMF-UJF-US). Recruté par la Commission Interdisciplinaire 45 « Cognition, langage, traitement de l'information, systèmes naturels et artificiels »

2003–2005

Recherches post-doctorales Projet international entre trois laboratoires :

- le Laboratoire de Physiologie de la Perception et de l'Action, au Collège de France à Paris (février-mars 2003), sous la direction du Prof. Alain Berthoz
- le Centre for Intelligent Products and Manufacturing Systems du Department of Mechanical Engineering, National University of Singapore (avril-décembre 2003), sous la direction de l'Associate Professor Marcelo-H. Ang Jr.
- le Laboratoire GRAVIR / IMAG - CNRS - INRIA, Grenoble (laboratoire de thèse)

Financements :

- février et mars 2003 : projet européen BIBA
- avril – décembre 2003 (à Singapour) : bourse de post-doctorat de l'INRIA International
- février – août 2004 : projet européen BIBA
- septembre 2004 – août 2005 : poste d'ATER à l'Université Joseph Fourier

Thème de recherche Modèles hiérarchiques probabilistes et bio-inspirés de la navigation, applications à la robotique mobile et à la modélisation du vivant. Etude de la **plausibilité biologique du formalisme de Carte Bayésienne** que j'ai défini dans mon doctorat.

2005 Admis Classé 1e au concours CR2 CNRS, commission interdisciplinaire 45 « Cognition, Langage, Traitement de l'Information : Systèmes Naturels et Artificiels ».

2004 Admis Classé 6e au concours CR2 CNRS, commission interdisciplinaire 45 « Cognition, Langage, Traitement de l'Information : Systèmes Naturels et Artificiels » (4 postes ouverts, 1 désistement : classement final, 5e).

2004 Qualifié aux fonctions de maître de conférences en 27e section (Informatique)

1999–2003

Doctorat spécialité « Informatique, Systèmes et Communication », Institut National Polytechnique de Grenoble, soutenu le 27 janvier 2003 (3 ans et 4 mois de préparation)
Note : il n'y a plus de mentions aux doctorats de l'INPG
Financement : allocation de recherche ministérielle (MENRT)
Laboratoires : LEIBNIZ-IMAG et GRAVIR / IMAG - CNRS - INRIA, Grenoble
Directeur de thèse : Pierre Bessière ; co-directeur: Emmanuel Mazer
Titre : « La carte bayésienne – Un modèle probabiliste hiérarchique pour la navigation en robotique mobile. »

Monitorat Université Joseph Fourier de Grenoble (UJF) (voir page 11)

1998–1999

DEA Informatique, Système et Communications, UJF Grenoble, Mention Bien, 15,78/20 (classement : 2e/20)
Sujet : Apprentissage hiérarchique bayésien.

Magistère 3 Informatique, UJF Grenoble, Mention Très Bien
Sujet : Programmes bayésiens pour un robot koala.
Diplôme final de Magistère obtenu avec la Mention Très Bien

1997–1998

Maîtrise à UCSB Informatique, UJF Grenoble, en mobilité internationale à University of California at Santa Barbara (UCSB) Grade Point Average à UCSB : 3,96/4,0 (Dean's honors list – liste honorifique du président) Equivalent à l'UJF : Mention Très Bien, 17/20 (classement : 1er/≈120)

Magistère 2 Informatique, UJF Grenoble, Mention Très Bien
Sujet : Tableur probabiliste pour l'aide à la décision.

1996–1997

Licence Informatique, UJF Grenoble, Mention Bien, 15/20 (classement : 2e/≈120)

Magistère 1 Informatique, UJF Grenoble, Mention Très Bien
Sujet : Algorithme de Berry-Sethi.

1995–1996

DEUG A, section Mathématiques–Physique–Informatique, UJF Grenoble, Mention Très Bien (classement : 1er/≈40)

1992–1993

Baccalauréat série C, Lycée Champollion, Grenoble, Mention Assez Bien

Publications

Note : La plupart de ces références sont disponibles à `http://diard.wordpress.com/`.

Articles de revues internationales

[1] J.-F. Patri, J. Diard, and P. Perrier. Optimal speech motor control and token-to-token variability: A Bayesian modeling approach. *Biological Cybernetics*, in press.

IF 2014: 1.713. Publisher: Springer Science+Business Media (Intl)

[2] C. Moulin-Frier, J. Diard, J.-L. Schwartz, and P. Bessière. COSMO ("Communicating about Objects using Sensory-Motor Operations"): a Bayesian modeling framework for studying speech communication and the emergence of phonological systems. *Journal of Phonetics (special issue "On the cognitive nature of speech sound systems")*, in press.

IF 2014: 1.598. Publisher: Elsevier B.V. (Netherlands). Target paper of a special issue (most other papers of this issue discuss the contribution of this target paper).

[3] J. Diard, V. Rynik, and J. Lorenceau. A bayesian computational model for online character recognition and disability assessment during cursive eye writing. *Frontiers in Psychology*, 4:843, 2013.

IF 2013: 2.843. Publisher: Frontiers Media S.A. (Switzerland)

[4] R. Laurent, C. Moulin-Frier, P. Bessière, J.-L. Schwartz, and J. Diard. Integrate, yes, but *what* and *how*? A computational approach of sensorimotor fusion in speech perception. *Behavioral and Brain Sciences (commentary)*, 36(4):364–365, 2013.

IF 2013: 14.962. Publisher: Cambridge University Press (UK)

[5] J. Diard, P. Bessière, and A. Berthoz. Spatial memory of paths using circular probability distributions: Theoretical properties, navigation strategies and orientation cue combination. *Spatial Cognition & Computation: An Interdisciplinary Journal*, 13(3):219–257, 2013.

IF 2013: 1.000. Publisher: Taylor and Francis (UK)

[6] C. Moulin-Frier, R. Laurent, P. Bessière, J.-L. Schwartz, and J. Diard. Adverse conditions improve distinguishability of auditory, motor and perceptuo-motor theories of speech perception: an exploratory bayesian modeling study. *Language and Cognitive Processes*, 27(7–8):1240–1263, 2012.

IF 2012: 1.628. Publisher: Taylor and Francis (UK)

[7] E. Gilet, J. Diard, and P. Bessière. Bayesian action–perception computational model: Interaction of production and recognition of cursive letters. *PLoS ONE*, 6(6):e20387, 2011.

IF 2011: 4.092. Publisher: Public Library of Science (USA)

[8] J. Diard, E. Gilet, É. Simonin, and P. Bessière. Incremental learning of bayesian sensorimotor models: from low-level behaviors to the large-scale structure of the environment. *Connection Science*, 22(4):291–312, 2010.

IF 2010: 1.057. Publisher: Taylor and Francis (UK)

[9] F. Colas, J. Diard, and P. Bessière. Common bayesian models for common cognitive issues. *Acta Biotheoretica*, 58(2-3):191–216, 2010.

IF 2010: 1.707. Publisher: Springer Science+Business Media (Intl)

[10] O. Lebeltel, P. Bessière, J. Diard, and E. Mazer. Bayesian robot programming. *Autonomous Robots*, 16(1):49–79, 2004.

IF 2012: 1.908. Publisher: Springer Science+Business Media (Intl)

Articles de revues nationales / francophones

[1] E. Gentaz, H. Faineteau, E. Gilet, J. Bluteau, R. Palluel-Germain, and J. Diard. L'estimation kinesthésique des distances : études comportementales et analyse probabiliste. *L'Année Psychologique*, 110(3):453–492, 2010.

IF 2010: 0.28. Publisher: NecPlus / Université Paris Descartes (France)

[2] O. Lebeltel, P. Bessière, J. Diard, and E. Mazer. Programmation bayésienne des robots. *Revue d'Intelligence Artificielle*, 18(2):261–298, 2004.

Publisher: Hermès Science (France)

Chapitres d'ouvrages collectifs

[1] C. Moulin-Frier, J.-L. Schwartz, J. Diard, and P. Bessière. Emergence of articulatory-acoustic systems from deictic interaction games in a "vocalize to localize" framework. In A. Vilain, J.-L. Schwartz, C. Abry, and J. Vauclair, editors, *Primate communication and human language: Vocalisations, gestures, imitation and deixis in humans and non-humans*, Advances in Interaction Studies, pages 193–220. John Benjamins Pub. Co., Amsterdam / Philadelphia, PA, 2011.

[2] J. Diard and P. Bessière. Bayesian maps: probabilistic and hierarchical models for mobile robot navigation. In P. Bessière, C. Laugier, and R. Siegwart, editors, *Probabilistic Reasoning and Decision Making in Sensory-Motor Systems*, volume 46 of *Springer Tracts in Advanced Robotics*, pages 153–176. Springer-Verlag, 2008.

[3] D. Bellot, R. Siegwart, P. Bessière, A. Tapus, C. Coué, and J. Diard. Bayesian modeling and reasoning for real world robotics: Basics and examples. In F. Iida, R. Pfeifer, L. Steels, and Y. Kuniyoshi, editors, *Embodied Artificial Intelligence, International Seminar, Dagstuhl Castle, Germany, July 7-11, 2003, Revised Papers*, volume 3139 of *Lecture Notes in Computer Science*, pages 186–201. Springer-Verlag, 2004.

Conférences internationales avec comité de lecture

[1] M. Chanceaux, V. Rynik, J. Lorenceau, and J. Diard. Writer recognition in cursive eye writing: a Bayesian model. In P. Bello, M. Guarini, M. McShane, and B. Scassellati, editors, *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, pages 2014–2019, Austin, TX, 2014. Cognitive Science Society.

Sélectivité : 515/714 (72%)

[2] R. Laurent, J.-L. Schwartz, P. Bessière, and J. Diard. A computational model of perceptuo-motor processing in speech perception: learning to imitate and categorize synthetic CV syllables. In *Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013)*, pages 2797–2801, Lyon, France, 2013.

Nominé pour le *Best Student paper award*. Sélectivité : 52% de 1549 papiers

[3] E. Gilet, J. Diard, R. Palluel-Germain, and P. Bessière. Bayesian action-perception loop modeling: Application to trajectory generation and recognition using internal motor simulation. In A. Mohammad-Djafari, J.-F. Bercher, and P. Bessière, editors, *Proceedings of the Thirtieth International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (Maxent 2010)*, pages 59–66, Melville, New-York, USA, 2010. American Institute of Physics Conference Proceedings.

Sélectivité non-précisée

[4] J. Diard, M. Lobier, and S. Valdois. Bayesian modeling of human performance in a visual processing training software. In V. Duffy, editor, *Advances in Applied Digital Human Modeling*,

Advances in Human Factors and Ergonomics Series, pages 467–476, Boca Raton, Florida, 2010. CRC Press, Taylor & Francis Group.

Sélectivité non-précisée

[5] C. Moulin-Frier, J.-L. Schwartz, J. Diard, and P. Bessière. A unified theoretical bayesian model of speech communication. In V. Duffy, editor, *Advances in Applied Digital Human Modeling*, Advances in Human Factors and Ergonomics Series, pages 457–466, Boca Raton, Florida, 2010. CRC Press, Taylor & Francis Group.

Sélectivité non-précisée

[6] J. Diard. Bayesian model comparison and distinguishability. In *Proceedings of the International Conference on Cognitive Modeling (ICCM 09)*, pages 204–209, 2009.

Sélectivité non-précisée

[7] J. Diard, P. Panagiotaki, and A. Berthoz. Biomimetic bayesian models of navigation: How are environment geometry-based and landmark-based strategies articulated in humans? In *Proceedings of the International Conference on Cognitive Modeling (ICCM 09)*, pages 210–215, 2009.

Sélectivité non-précisée

[8] S. Lallée, J. Diard, and S. Rousset. Multiple Object Manipulation: is structural modularity necessary? a study of the MOSAIC and CARMA models. In *Proceedings of the International Conference on Cognitive Modeling (ICCM 09)*, pages 306–311, 2009.

Sélectivité non-précisée

[9] É. Simonin and J. Diard. BBPRM: a behavior-based probabilistic roadmap method. In *Proceedings of the 2008 IEEE International Conference on Systems, Man and Cybernetics*, pages 1719–1724, 2008.

Sélectivité : 618/≈1000 (≈ 62%)

[10] J. Diard, P. Bessière, and E. Mazer. Merging probabilistic models of navigation: the bayesian map and the superposition operator. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS05)*, pages 668–673, 2005.

Sélectivité : 663/1204 (55 %)

[11] É. Simonin, J. Diard, and P. Bessière. Learning Bayesian models of sensorimotor interaction: from random exploration toward the discovery of new behaviors. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS05)*, pages 1226–1231, 2005.

Sélectivité : 663/1204 (55 %)

[12] J. Diard, P. Bessière, and E. Mazer. Hierarchies of probabilistic models of navigation: the bayesian map and the abstraction operator. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA04)*, pages 3837–3842, New Orleans, LA, USA, 2004.

Sélectivité : 858/1457 (59 %)

[13] J. Diard, P. Bessière, and E. Mazer. A theoretical comparison of probabilistic and biomimetic models of mobile robot navigation. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA04)*, pages 933–938, New Orleans, LA, USA, 2004.

Sélectivité : 858/1457 (59 %)

[14] J. Diard, P. Bessière, and E. Mazer. A survey of probabilistic models, using the bayesian programming methodology as a unifying framework. In *The Second International Conference on Computational Intelligence, Robotics and Autonomous Systems (CIRAS 2003)*, Singapore, December 2003.

Sélectivité : 171/243 (70 %)

[15] J. Diard and O. Lebeltel. Bayesian programming and hierarchical learning in robotics. In Meyer, Berthoz, Floreano, Roitblat, and Wilson, editors, *SAB2000 Proceedings Supplement Book*, pages 140–149, Honolulu, USA, 11–16 2000. International Society for Adaptive Behavior.

Sélectivité : 98/150 (65 %)

[16] O. Lebeltel, J. Diard, P. Bessière, and E. Mazer. A bayesian framework for robotic programming. In A. Mohammad-Djafari, editor, *Twentieth International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (Maxent 2000)*, pages 625–637, Melville, New-York, USA, 2000. American Institute of Physics Conference Proceedings.

Sélectivité non-précisée

[17] J. Diard and O. Lebeltel. Bayesian learning experiments with a khepera robot. In Löffler, Mondada, and Rückert, editors, *Experiments with the Mini-Robot Khepera : Proceedings of the 1st International Khepera Workshop*, pages 129–138, Germany, 1999. HNI-Verlagsschriftenreihe, band 64.

Sélectivité non-précisée

Colloques internationaux avec comité de lecture (workshops en marge de conférences)

[1] M.-L. Barnaud, N. Morgado, R. Palluel-Germain, J. Diard, and A. Spalanzani. Proxemics models for human-aware navigation in robotics: Grounding interaction and personal space models in experimental data from psychology. In *Proceedings of the 3rd IROS'2014 workshop "Assistance and Service Robotics in a Human Environment"*, 2014.

[2] E. Gilet, J. Diard, and P. Bessière. Bayesian modelling of a sensorimotor loop: application to handwriting. In *Proceedings of the NIPS 2008 workshop "Principled theoretical frameworks for the perception-action cycle"*, 2008.

[3] J. Diard, P. Bessière, and E. Mazer. Hierarchies of probabilistic models of space for mobile robots : the bayesian map and the abstraction operator. In *Reasoning with Uncertainty in Robotics (IJCAI'03 Workshop)*, 2003.

Conférences francophones avec comité de lecture

[1] R. Laurent, J.-L. Schwartz, P. Bessière, and J. Diard. COSMO, un modèle bayésien de la communication parlée : application à la perception des syllabes (COSMO, a bayesian model of speech communication, applied to syllable perception) [in french]. In *Actes de la conférence conjointe JEP-TALN-RECITAL 2012, volume 1: JEP*, pages 305–312, Grenoble, France, June 2012. ATALA/AFCP.

[2] E. Gilet, J. Diard, and P. Bessière. Bayesian modelling of sensorimotor systems: application to handwriting. In *NeuroComp 08*, pages 203–206, 2008.

[3] C. Moulin-Frier, J.-L. Schwartz, J. Diard, and P. Bessière. Emergence du langage par jeux déictiques dans une société d'agents sensori-moteurs en interaction. In *Journées d'Etude de la Parole*, 2008.

Posters ou abstracts dans des conférences internationales avec comité de lecture

[1] T. Phenix, S. Valdois, and J. Diard. Bayesian modelling of visual attention in word recognition: simulating optimal viewing position. In *Proceedings of the 19th Conference of the European Society for Cognitive Psychology (ESCOP)*, 2015.

Also reused for the EDISCE day, and won the *Best Poster Award*; also reused for the Workshop "Probabilistic Inference and the Brain", at Collège de France, Paris (September, 2015);

[2] M.-L. Barnaud, J. Diard, P. Bessière, and J.-L. Schwartz. COSMO, a Bayesian computational model of speech communication: Assessing the role of sensory vs. motor knowledge in speech perception. In *5th International Conference on Development and Learning and on Epigenetic Robotics (ICDL-Epirob)*, pages 248–249, 2015.

Also reused for the Workshop "Probabilistic Inference and the Brain", at Collège de France, Paris (September, 2015)

[3] J.-F. Patri, J. Diard, J.-L. Schwartz, and P. Perrier. A Bayesian framework for speech motor control. In *Progress in Motor Control X*, 2015.

Also reused for the Workshop "Probabilistic Inference and the Brain", at Collège de France, Paris (September, 2015)

[4] M.-L. Barnaud, R. Laurent, P. Bessière, J. Diard, and J.-L. Schwartz. Modeling concurrent development of speech perception and production in a Bayesian framework. In *Workshop on Infant Language Development (WILD)*, 2015.

[5] C. Moulin-Frier, R. Laurent, P. Bessière, J.-L. Schwartz, and J. Diard. Noise and inter-speaker variability improve distinguishability of auditory, motor and perceptuo-motor theories of speech perception: an exploratory Bayesian modeling study. In *Proceedings of the 9th International Seminar on Speech Production (ISSP 11)*, Montréal, Québec, Canada, 2011.

[6] J. Diard, P. Panagiotaki, A. Berthoz, and P. Bessière. Modeling human spatial navigation components and their relationship using bayesian sensori-motor schemas. In *Proceedings of the 12th World Congress of Psychophysiology*, 2004.

[7] P. Panagiotaki, J. Diard, A. Lehmann, and A. Berthoz. Hierarchical cognitive mechanisms of memory encoding during spatial navigation on humans: Psychophysical studies with virtual reality. In *Proceedings of the 12th World Congress of Psychophysiology*, 2004.

Textes ou posters dans des colloques internationaux sans comité de lecture

[1] C. Moulin-Frier, J.-L. Schwartz, J. Diard, and P. Bessière. Emergence of a language through deictic games within a society of sensory-motor agents in interaction. In *Proceedings of the 8th International Seminar on Speech Production (ISSP 08)*, 2008.

[2] J. Diard, P. Bessière, and E. Mazer. Combining probabilistic models of space for mobile robots: the bayesian map and the superposition operator. In M. Armada and P. Gonz·lez de Santos, editors, *Proceedings of the 3rd IARP (International Advanced Robotics Programme) International Workshop on Service, Assistive and Personal Robots - Technical Challenges and Real World Application Perspectives*, pages 65–72, Madrid, Spain, October 14–16 2003. Industrial Automation Institute (IAI-CSIC) Spanish Council for Scientific Research.

[3] J. Diard, P. Bessière, and O. Lebeltel. Robotic programming and space representation using a unified bayesian framework. In D. Pélisson, C. Prabland, and Y. Rossetti, editors, *Proceedings of the International Symposium on Neural control of space coding and action production (poster)*, 2001.

Rapports techniques

[1] E. Gilet, J. Diard, and P. Bessière. D2.19: Bayesian modeling of sensori-motor system: an application to handwriting. BACS european project deliverable 2.19 "Bayesian Approach to Cognitive Systems", European Community, "Information Society Technologies" Programme, 2010.

[2] M. Lafon, J. Diard, M. Chevaldonne, D. Paillot, and A. Berthoz. Rep. on the bayesian modelling and model selection in the bayes city experiment. BACS european project deliverable 2.16 "Bayesian Approach to Cognitive Systems", European Community, "Information Society Technologies" Programme, 2009.

[3] J. Diard, M. Lafon, and A. Berthoz. Bayesian modelling of visual cue combination in human path memorization tasks. BACS european project deliverable 2.6 "Bayesian Approach to Cognitive Systems", European Community, "Information Society Technologies" Programme, July 2009.

[4] J. Diard. Bayesian model comparison and distinguishability. BACS european project deliverable 3.5 "Bayesian Approach to Cognitive Systems", European Community, "Information Society Technologies" Programme, 2009.

[5] M. Lafon, J. Diard, M. Chevaldonne, D. Paillot, and A. Berthoz. Human path memorization for navigation in the virtual bayes city experiment. BACS european project deliverable 2.9 "Bayesian Approach to Cognitive Systems", European Community, "Information Society Technologies" Programme, July 2008.

[6] P. Panagiotaki, J. Diard, A. Lehmann, and A. Berthoz. Human cognitive strategies of navigation in the magnapolis experiment. BACS european project deliverable 6.6 "Bayesian Approach to Cognitive Systems", European Community, "Information Society Technologies" Programme, 2007.

[7] J. Diard, P. Panagiotaki, and A. Berthoz. Bayesian modelling of human homing in large-scale urban environments – bayesian modelling of path integration using von mises probability distributions. BACS european project deliverable 2.3 "Bayesian Approach to Cognitive Systems", European Community, "Information Society Technologies" Programme, December 2007.

[8] F. Colas, J. Diard, and P. Bessière. Review report of bayesian approaches to fusion multimodality, conflicts, ambiguities, hierarchies and loops – common bayesian models for common cognitive issues. BACS european project deliverable 3.2 "Bayesian Approach to Cognitive Systems", European Community, "Information Society Technologies" Programme, January 2007.

[9] J. Diard, P. Panagiotaki, and A. Berthoz. Are landmark-based and geometry-based human navigation strategies hierarchically articulated? a bayesian modelling perspective. BIBA european project deliverable d21, section 6.3 "Bayesian Inspired Brain and Artefacts: using probabilistic logic to understand brain function and implement life-like behavioural co-ordination", European Community, "Information Society Technologies" Programme, 2005.

[10] P. Panagiotaki, J. Diard, and A. Berthoz. Geometric encoding during human navigation used as 'back-up' mechanism in case of sudden absence of familiar landmarks: An immersive vr study. BIBA european project deliverable d21, section 6.3 "Bayesian Inspired Brain and Artefacts: using probabilistic logic to understand brain function and implement life-like behavioural co-ordination", European Community, "Information Society Technologies" Programme, 2005.

[11] J. Diard, P. Panagiotaki, A. Berthoz, and P. Bessière. The bayesian map: a probabilistic and hierarchical model of navigation. BIBA european project deliverable 15, section 6.3 "Bayesian Inspired Brain and Artefacts: using probabilistic logic to understand brain function and implement life-like behavioural co-ordination", European Community, "Information Society Technologies" Programme, 2004.

[12] P. Bessière, J.-M. Ahuactzin, O. Aycard, D. Bellot, F. Colas, C. Coué, J. Diard, R. Garcia, C. Koike, O. Lebeltel, R. LeHy, O. Malrait, E. Mazer, K. Mekhnacha, C. Pradalier, and A. Spalanzani. Survey: Probabilistic methodology and techniques for artefact conception and development. Technical Report RR-4730, INRIA Rhône-Alpes, Montbonnot, France, 2003.

[13] O. Lebeltel, P. Bessière, J. Diard, and E. Mazer. Bayesian robot programming. Technical Report 01, Les Cahiers du Laboratoire Leibniz, Grenoble, France, Mai 2000.

Rapports de stages

[1] J. Diard. *La carte bayésienne – Un modèle probabiliste hiérarchique pour la navigation en robotique mobile.* Thèse de doctorat, Institut National Polytechnique de Grenoble, Grenoble, France, Janvier 2003.

[2] J. Diard. Programmes bayésiens pour un robot koala. Rapport de Troisième année de Magistère d'Informatique à l'Université Joseph Fourier, Septembre 1999.

[3] J. Diard. Apprentissage hiérarchique bayésien. Rapport de dea « informatique, systèmes et communications », Université Joseph Fourier, Grenoble, France, Juin 1999.

[4] J. Diard. Tableur probabiliste pour l'aide la décision. Rapport de Seconde année de Magistère d'Informatique à l'Université Joseph Fourier, Septembre 1998.

[5] J. Diard. Algorithme de Berry-Sethi. Rapport de Première année de Magistère d'Informatique à l'Université Joseph Fourier, Août 1997.

Activités d'enseignement

2014–2015 Master 2R Sciences Cognitives, « Cognition Bayésienne », 18h

2014–2015 Master 2R Psychologie Cognitive, « Introduction à la modélisation Bayésienne », 6h

2013–2014 Master 2R Sciences Cognitives, « Cognition Bayésienne », 15h

2013–2014 Master 2R Psychologie Cognitive, « Introduction à la modélisation Bayésienne », 6h

2012–2013 Master 2R Sciences Cognitives, « Cognition Bayésienne », 18h

2012–2013 Master 2R Psychologie Cognitive, « Introduction à la modélisation Bayésienne », 6h

2011–2012 Master 2R Sciences Cognitives, « Cognition Bayésienne », 12h

2011–2012 Master 2R Psychologie Cognitive, « Introduction à la modélisation Bayésienne », 6h

2010–2011 Master 2R Sciences Cognitives, « Cognition Bayésienne », 12h

2010–2011 Master 2R Psychologie Cognitive, « Introduction à la modélisation Bayésienne », 6h

2009–2010 Master 2R Sciences Cognitives, « Cognition Bayésienne », 6h

2009–2010 Master 2R Psychologie Cognitive, « Modélisation Bayésienne des systèmes sensori-moteurs », 6h

2008–2009 Master 2R Sciences Cognitives, « Cognition Bayésienne », 6h

2008–2009 Master 2R Psychologie Cognitive, « Modélisation Bayésienne des systèmes sensori-moteurs », 3h

2004–2005 Attaché Temporaire d'Enseignement et de Recherche (ATER) à l'UJF, Grenoble. Enseignements donnés sur l'antenne de Valence de l'UJF. **Total : 192 h eqTD**

- **96h TD et TP** « INF 110V : Méthodes informatiques et techniques de programmation », Licence Sciences et Technologie, un groupe de 1e année (48h) et un groupe de 2e année (48h).
- **13h responsabilité** Responsable des Travaux Dirigés d'Expérimentation pour le module INF 110V.
- **30h TD et TP** « INF 350V : Programmation Orientée Objet », Licence Sciences et Technologie, un groupe de 3e année.
- **53h Cours, TD, TP, Projet** « INF 242V : Programmation logique et contraintes », Licence Sciences et Technologie, un groupe de 2e année.

2001–2002 Troisième année de Monitorat à l'UJF, Grenoble **Total : 64 h eqTD**

- **44h Cours/TD et TP** « Méthodes informatiques pour les disciplines scientifiques », 2e année de DEUG Sciences de la Matière (SM)
- **18h Cours/TD** « Langages et Programmation 2 : Probabilités », Licence d'Informatique, UFR Informatique et Mathématiques Appliquées (IMA)

2000–2001 Seconde année de Monitorat à l'UJF, Grenoble **Total : 64 h eqTD**

- **44h Cours/TD et TP** « Méthodes informatiques pour les disciplines scientifiques », 2e année de DEUG SM
- **24h Projet** de programmation de fin d'année, Licence d'Informatique, UFR IMA

1999–2000 Première année de Monitorat à l'UJF, Grenoble **Total : 64 h eqTD**

44h Cours/TD et TP « Méthodes informatiques pour les disciplines scientifiques », 2e année de DEUG Sciences de la Matière (SM)

18h TD « Outils Formels pour l'Informatique 2 : Analyse Syntaxique », Licence d'Informatique / MST ESI, UFR IMA

1998–1999 Tutorat en DEUG première année à l'UJF, Grenoble

mars–mai 1998 Correcteur de copies à UCSB pour un cours d'architecture matérielle d'ordinateurs (cs 154 : computer architecture. Prof K. Schauser ; schauser@cs.ucsb.edu)

1996 Cours de soutien, organisés par la PEEP, en Mathématiques et Sciences Physiques, classes de Seconde au Lycée des Eaux Claires, Grenoble

Activités d'encadrement (en cours)

Doctorat Marie-Lou Barnaud. Doctorante école doctorale EDISCE. Co-encadrement avec Jean-Luc Schwartz (GIPSA-Lab). **Titre de travail** : « Understanding how speech unites the sensory and motor streams, and how speech units emerge from perceptuo-motor interactions ». **Financement** : ERC Jean-Luc Schwartz. (septembre 2014–)

Doctorat Jean-François Patri. Doctorant école doctorale EDISCE. Co-encadrement avec Pascal Perrier (GIPSA-Lab). **Titre de travail** : « Modéles bayésiens et contrôle moteur de la parole ». **Financement** : ERC Jean-Luc Schwartz. (octobre 2014–)

Doctorat Thierry Phénix. Doctorant école doctorale EDISCE. Co-encadrement avec Sylviane Valdois. **Titre de travail** : « EVAPlus : Modèle probabiliste d'évaluation de la performance et de l'apprentissage, application à l'apprentissage de la lecture ». **Financement** : Allocation de Recherche de la Fondation de France (janvier 2014–).

Doctorat Svetlana Meyer. Doctorante école doctorale EDISCE. Co-encadrement avec Sylviane Valdois. **Titre de travail** : « Modélisation bayésienne de la lecture et de son apprentissage ». **Financement** : Allocation de Recherche ministérielle (octobre 2015–).

Activités d'encadrement (précédemment)

Post-doctorat Myriam Chanceaux **Titre** : « Motor control and writer identifiability in eye-writing ». **Financement** : Projet ANR EOL (2013–2014).

Doctorat Raphaël Laurent. Directeur de thèse principal, grâce à un agrément ponctuel à diriger un doctorat sans HDR (école doctorale EDMSTII). Co-encadrement avec Jean-Luc Schwartz et Pierre Bessière. **Titre** : « COSMO : un modèle bayésien des interactions sensori-motrices dans la perception de la parole ». **Financement** : Allocation de Recherche ministérielle (2010–2013), poster ATER (2013–2014). Soutenance le 8 octobre 2014.

Doctorat Clément Moulin-Frier. Doctorant école doctorale EDISCE. Co-encadrement avec Jean-Luc Schwartz (encadrement effectif à 20%). **Titre** : « Rôle des relations perception-action dans la communication parlée et l'émergence des systèmes phonologiques : étude, modélisation computationnelle et simulations ». **Financement** : Allocation de Recherche ministérielle. Soutenance le 15 juin 2011 (2007–2011).

Doctorat Estelle Gilet. Doctorante de l'école doctorale Mathématiques, Sciences et Technologies de l'Information, Informatique, Université Joseph Fourier, Grenoble. Co-encadrement avec Pierre Bessière (encadrement effectif à 80%). **Titre** : « Modélisation Bayésienne d'une boucle perception-action : application à la lecture et à l'écriture ». **Financement** : bourse du projet européen BACS. Soutenance le 02 octobre 2009 (2006–2009).

Doctorat Eva Simonin. Doctorante de l'école doctorale Mathématiques, Sciences et Technologies de l'Information, Informatique, Université Joseph Fourier, Grenoble. Co-encadrement avec Pierre Bessière (encadrement effectif à 100%). **Titre** : « Apprentissages et modèles probabilistes bio-inspirés pour la navigation en robotique mobile ». **Financement** : Allocation de Recherche MENRT. (2004–2005 : thèse interrompue par l'étudiante).

3^e année d'école d'ingénieur Svetlana Meyer. École Nationale Supérieure de Cognitique (ENSC), Bordeaux. **Titre de travail** : « Equivalence motrice et écriture avec les yeux » **Financement** : Projet ANR EOL (2014–2015).

Master 2R Marie-Lou Barnaud, Master 2R Information, Cognition, Apprentissages, spécialité Sciences Cognitives. Co-encadrement avec Anne Spalanzani (Laboratoire d'Informatique de Grenoble). **Titre** : « Intégration des contraintes sociales de l'action dans la navigation robotique » (2013–2014).

Master 2R Adrien Delmas, Master 2R Information, Cognition, Apprentissages, spécialité Sciences Cognitives. Co-encadrement avec Jean-Luc Schwartz et Raphaël Laurent (GIPSA-Lab). **Titre** : « Modélisation d'idiosyncrasies dans l'apprentissage de la parole » (2013–2014).

Master 2R Jean-François Patri, Master 2R Information, Cognition, Apprentissages, spécialité Sciences Cognitives. Co-encadrement avec Pascal Perrier (GIPSA-Lab). **Titre** : « Reformulation Bayésienne d'un modèle de contrôle optimal pour la planification motrice de gestes de parole » (2013–2014).

Master 2R Léo Lopez, Master 2R Informatique, Université Blaise Pascal, Clermont-Ferrand. **Titre** : « Comparaison et distinguabilité Bayésienne des modèles » (2012–2013).

3^e année d'école d'ingénieur Jérémy Krein, Justine Rexer et Victor Sabatier. Co-encadrement avec Christophe Poyet (GRAVIT-Innovation). Stage ManInTec, **Titre** : « Thief Escape : design et développement d'un jeu vidéo démonstrateur de l'algorithme PARSEVAL » (2012–2013).

Master 2R Raphaël Laurent, co-encadrement avec Pierre Bessière et Jean-Luc Schwartz, Master international d'Informatique, MOSIG. **Titre** : « A computational model to study quantitatively motor, sensory, and sensorimotor model responses in Speech Recognition » (2009–2010).

Master 2R Stéphane Lallée, « Implémentation et comparaisons de deux modèles de contrôle moteur, MOSAIC et CARMA : modularité structurelle versus représentation interne des objets. », Master Information, Cognition, Apprentissages, spécialité Sciences Cognitives, Institut National Polytechnique de Grenoble. Co-encadrement avec Stéphane Rousset. (2007–2008).

Master 2R Clément Moulin-Frier, « Jeux déictiques dans une société d'agents sensori-moteurs », Master Information, Cognition, Apprentissages, spécialité Sciences Cognitives, Institut National Polytechnique de Grenoble. Co-encadrement avec Jean-Luc Schwartz . (2006–2007).

Master 2R & Magistère 3 Estelle Gilet, Master Mathématique, Informatique, Université Joseph Fourier, Grenoble. Co-encadrement avec Pierre Bessière et Jean-Luc Schwartz (encadrement effectif à 100%). **Titre** : « Modèles bayésiens et sélection de modèles de perception pour les systèmes sensori-moteurs. ». (2005–2006)

Master 2R Eva Simonin, Master Mathématique, Informatique, Université Joseph Fourier, Grenoble. Co-encadrement avec Pierre Bessière (encadrement effectif à 100%). **Titre** : « Apprentissages et cartes bayésiennes ». **Sujet** : Etude et expérimentation en robotique mobile d'une méthode originale d'apprentissage de cartes de l'environnement (transformation d'un comportement en carte bayésienne). [Septembre 2003–Juin 2004]

2^e année d'école d'ingénieur Vincent Rynik, Ecole Centrale Lyon. **Titre** : « Modèle computationnel bayésien pour la reconnaissance on-line de caractères et l'évaluation du handicap – Application à l'écriture avec les yeux ». **Financement** : Projet ANR EOL (2012–2013).

Master 1 Informatique Clément Vetroff, TER M1 Université Joseph Fourier. **Titre** : « PARSEVAL : Algorithme Probabiliste d'Evaluation Temps-Réel de la Performance d'un Sujet » (2012–2013).

Master 1 Maxime Frecon, Master 1 Psychologie, co-encadrement avec Richard Palluel-Germain. **Titre** : « Simulation motrice et reconnaissance de lettres » (2009–2011).

2e année ENSIMAG Layla El Asri, co-encadrement avec Muriel Lobier. (2009–2010). **Titre** : « Probabilistic modeling of a subject's performance space. Application to adaptive software ».

Master 1 Mostapha Mahmoud, Master 1 ICPS, co-encadrement avec Muriel Lobier. **Titre** : « Modélisation probabiliste des performances d'un sujet et de son apprentissage – Application à un logiciel de rééducation de la dyslexie ». (2008–2009)

Master 1 Jonatan Wentzel, Master Information Cognition Apprentissages - Ingénierie de la Communication Personne-Système, Université Pierre Mendes-France, Grenoble. Co-encadrement avec Edouard Gentaz (encadrement effectif à 20%). **Titre** : « PsychBar ». (2006)

Magistère 2 Estelle Gilet, Magistère 2 Informatique, UJF, Grenoble. Co-encadrement avec Eva Simonin (encadrement effectif à 50%). **Titre** : « Apprentissage de relations sensori-motrices pour la superposition de cartes bayésiennes ». **Sujet** : Expérimentation robotique de superposition et apprentissage de cartes bayésiennes basées sur deux modalités sensorielles différentes (gradient lumineux et proximité des obstacles). [Septembre 2004–Août 2005]

Magistère 2 Laurent Goethals, Magistère 2 Informatique, UJF, Grenoble (encadrement à 100%). **Titre** : « Manipulation robotique bio-inspirée ». **Sujet** : Continuation de l'expérience de simulation d'un algorithme de contrôle hiérarchique pour manipulation d'objets, et réécriture bayésienne d'un modèle de Gomi, Wolpert et Kawato (Mosaic). Projet en collaboration avec Etienne Burdet de l'Université Nationale de Singapour (NUS). [Septembre 2004–Août 2005]

Final Year Project (équiv. stage de maîtrise) Lee Kao Hsiung, Faculty of Engineering de la National University of Singapore (NUS). Co-encadrement avec Marcelo Ang (encadrement effectif à 100%). **Titre** : « Biologically-inspired Bayesian Mobile Robot Navigation ». **Sujet** : Simulation d'un scénario robotique de poursuite de proies ; utilisation de l'opérateur de superposition de cartes bayésiennes. [Août 2003–Avril 2004]

Final Year Project (équiv. stage de maîtrise) Ang Ghee Koon, Faculty of Engineering de la National University of Singapore (NUS). Co-encadrement avec Etienne Burdet (encadrement effectif à 80%). **Titre** : « Biomimetic Skillfull Robot Manipulation ». **Sujet** : Simulation d'un algorithme de contrôle hiérarchique pour manipulation d'objets. Réécriture bayésienne d'un modèle de Gomi, Wolpert et Kawato (Mosaic). [Août 2003–Avril 2004]

Magistère 2 Frédéric Raspail, Magistère 2 Informatique, UJF, Grenoble. Co-encadrement avec Olivier Aycard (encadrement effectif à 80%). **Titre** : « Localisation par HMMs sur robot Koala ». **Sujet** : Expérimentation robotique de localisation et apprentissage d'un modèle probabiliste dans un couloir pour robot mobile. [Septembre 1999–Août 2000]

Magistère 1 Stéphane Bissol, Magistère 1 Informatique, UJF, Grenoble. Co-encadrement avec Pierre Bessière (encadrement effectif à 80%). **Titre** : « Module de visualisation de fonctions ». **Sujet** : Développement d'un module pour une maquette de tableur probabiliste. [Septembre 1998–Août 1999]

Projets de recherche

Porteur Projet de valorisation SATT GIFT, maturation, Projet « Cognidis », 2015–2016, 115 k€

Co-porteur Fondation de France, appel d'offres « Recherche sur le développement cognitif et les apprentissages de l'enfant » ; Projet « EVAPlus : Modèle probabiliste d'évaluation de la performance et de l'apprentissage, application à l'apprentissage de la lecture », co-porté avec Sylviane Valdois, 2014–2017, 117 k€

Porteur Projet de valorisation GRAVIT, maturation, Projet « PARSEVAL, probabilistic algorithm for real-time subject evaluation », 2013–2014, 35 k€

Porteur Soutien du Conseil Scientifique de l'Université Pierre-Mendès-France (Grenoble 2), Projet « Distinguabilité bayésienne des modèles, application à la mesure de performance des sujets », 2013, 3,5 k€

Porteur Pôle Grenoble Cognition, Projet « IPAS – Integrating Perception and Action in Speech », 2013, 4,5 k€

Partenaire ANR, appel d'offres « TecSan, Technologies pour la Santé », Projet « EOL (Eye On-Line) : Directed pursuit for eye-writing »; porteur : J. Lorenceau, 2012–2015, 402 k€.

Porteur Projet de valorisation GRAVIT, pré-maturation, Projet « PARSEVAL, probabilistic algorithm for real-time subject evaluation », 2011–2013.

Partenaire Pôle Grenoble Cognition, Projet « Modélisation bayésienne de la perception et de la production de la parole : simulation comparative et distinguabilité des modèles », 2011, 3 k€

Partenaire Membre du projet européen BACS. Responsable scientifique du WorkPackage 3, « Bayesian approach to cognitive systems: Fusion, multi-modality, conflicts, ambiguities, hierarchies, loops and, stability in a Bayesian context ». Partenaires : ETH (Suisse), INRIA - GRAVIR (France), CdF - LPPA (France), MPI - BC (Allemagne), ISC (France), IDIAP (Suisse), HUG (Suisse), MIT (USA), UniCoimbra (Portugal), ProBayes (France), BlueBotics (Suisse), EDF (France), 2006–2009.

Partenaire Projet de recherche PICS-CNRS 2612 « Visual perception of the 3D space and applications in robotics ». Partenaires : SERI (Singapour), NUS (Singapour), CNRS-GRAVIR & INRIA Rhône-Alpes (France), 2004–2006.

Partenaire Membre du projet européen BIBA (IST–2001-32115 Bayesian Inspired Brain and Artefacts : Using probabilistic logic to understand brain function and implement life-like behavioural co-ordination). Partenaires : GRAVIR (France), LPPA (France), EPFL (Suisse), UCL Gatsby (GB), UCAM (GB), MIT (USA), Bluebotics (Suisse), Probayes (France), 2001–2005.

Responsabilités collectives diverses

Relecteur pour des conférences internationales : CosSci 2015, CosSci 2014, SAB 2014, CosSci 2013, ICRA 2013, Ro-Man 2012, SAB 2012, AIM 12, SAB 2010, ICIT 10, AHFE/ADHM 10, HSI 09, AIM 09, CIRAS 09, INDIN 09, CJCSC 09, SMC 09, ICCM 09, SMC 08, IROS 07, ISIE 07, CIRAS 2005, IJCAI 2005, IROS 2005, ICRA 2005, IROS 2004, IROS 2002

2015 Associate Editor pour The 5th Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics (IEEE ICDL-Epirob 15).

2014 Technical program committee et relecteur pour SAB 2014.

2014 Relecteur pour *Human Movement Science*.

2013 Relecteur pour *Connection Science*.

2013 Expertise pour le CNRS, appel d'offre PEPS site grenoblois.

2012 Relecteur pour le *Journal of Computer Science and Technology*.

2012 Expertise pour le Ministry of Science and Innovation (MSI) de Nouvelle Zélande.

2012 Technical program committee et relecteur pour SAB 2012.

2011 Relecteur pour *Adaptive Behavior*.

2010 Relecteur pour l'*Année Psychologique*.

2010 Expertise ANR, appel d'offres « Blanc ».

2010 Relecteur pour un numéro spécial de *Cognitive Systems Research*.

2010 Technical program committee et relecteur pour SAB 2010.

2009 Expertise dossiers allocations de recherche Région Bretagne.

2008 Membre du Technical Program Committee, relecteur et Session Chair pour la IEEE International Conference on Systems, Man and Cybernetics 08 (SMC 08), Singapour, 12–15 octobre 2008.

2008 Organisateur principal du workshop « Bayesian Models and Contraction Theory » dans le cadre du projet européen BACS, 13–15 janvier 2008, Chamonix, France

2006 Relecteur pour un numéro spécial de *Psych Research*.

3 août 2005 Co-président de séance (*session chairman*) de la session WPII-7 (Behavior Learning) à la IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS05), Edmonton, Canada.

2004–2005 Co-rédacteur du projet européen BACS (Bayesian Approach to Cognitive Systems), Appel d'offre IST Cognitive Systems 2005.

30 août – 1er septembre 2004 Organisateur principal du workshop « Navigation » pour le projet européen BIBA (Collège de France, Paris).

18 décembre 2003 Co-président de séance (*session chairman*) de la session PS07 (Signal Processing) à la Second International Conference on Computational Intelligence, Robotics and Autonomous Systems (CIRAS 03), Singapour.

avril–décembre 2003 Coordonnateur pour les collaborations entre le laboratoire CIPMAS de la National University of Singapore et la France. Par exemple :

- co-organisation d'un séminaire et de la visite du Prof. Pissaloux (Laboratoire de Robotique de Paris) ;
- co-organisation du stage au LPPA d'un doctorant de Marcelo Ang (Mana Saedan, financement Egide attribué par l'Ambassade de France à Singapour).

2003 Co-rédacteur d'un projet de recherche PICS déposé au CNRS en 2003. Partenaires : SERI (Singapour), NUS (Singapour), CNRS-GRAVIR & INRIA Rhône-Alpes (France). Titre : « Visual perception of the 3D space and applications in robotics ». Responsable d'une partie du sous-projet 3.

1999–2003 Responsable et administrateur d'une base de données d'articles scientifiques. Cette base fut pendant une partie de ma thèse celle du PRASC (Pôle Rhône-Alpes de Sciences Cognitives).

Formations diverses

Cours Formations Mathematica, 2011.

Cours « Méthodologie et épistémologie », Master 2 Psychologie Cognitive, UPMF (Grenoble 2), auditeur libre (10/06)

Cours « Algorithmic Motion Planning » à NUS, Singapour, auditeur libre (08 – 11/03)

Ecole d'hiver Seconde du projet européen BIBA, Combloux, France (16 – 22/01/03)

Ecole d'été Première du projet européen BIBA, Moudon, Suisse (01 – 05/07/02)

Cours « Philosophie des Sciences », UJF Grenoble (01 – 04/02)

Ecole d'hiver Première du projet européen BIBA, Chichilianne, France (18 – 20/02/02)

Ecole d'été « Summer School on Image and Robotics », INRIA Rhône-Alpes, 70 heures (26/06 – 7/07/00)

Activités de communication – Développements de logiciels

Conf invitée Table ronde de vulgarisation, « Sciences & SF », avec Alain Damasio (auteur de SF, « La Horde du Contrevent ») et Mathias Echenay (Editions La Volte), vidéo en ligne sur le site Echosciences `http://www.echosciences-grenoble.fr/actualites/table-ronde-sciences-sf-avec-damasio-barberi-et-diard`, 18/04/14

Séminaire « COSMO, un modèle d'interaction perceptuo-motrice dans la parole », workshop « Robotique et interaction », GIPSA-LAb, 18/03/2014

Conf invitée « Raisonnement probabiliste », conférence invitée aux JNRR 2013 (Journées Nationales de la Recherche en Robotique), 17/10/13

Séminaire « Modélisation bayésienne : des equations à l'utilisateur », Journée du Pôle Grenoble Cognition, 04/06/2013

Séminaire « Bayesian modeling of cognition or Modeling Bayesian cognition? », séminaire invité du GIPSA-Lab, 07/01/13

Logiciel Dépôt à l'Agence pour la Protection des Programmes (APP) du logiciel PARSEVAL, « Probabilistic Algorithm for Real-time Subject EVALuation », co-auteurs : Julien Diard, Muriel Lobier et Layla El Asri, numéro FIST-CNRS DL 05304-01, 2012, dépôt effectif le 21/11/2013 sous le n° IDDN.FR.001.470023.000.S.P.2013.000.20700.

Séminaire « Adverse conditions improve distinguishability of auditory, motor and perceptuo-motor theories of speech perception: an exploratory Bayesian modeling study », séminaire invité du GIPSA-Lab, 08/12/11

Séminaire « Modélisation Bayésienne de l'interaction entre perception et action : simuler la production de gestes pour mieux percevoir, simuler la reconnaissance de gestes pour mieux produire », Journée du Pôle Grenoble Cognition, 31/06/11

Séminaire Julien Diard et Richard Palluel-Germain, « Mirror Neuron Theory of Action Understanding », atelier biblio du LPNC, 15/06/2010

Séminaire Julien Diard,« Is your hierarchical Bayesian model layered, or abstracted? », workshop du WP3 BACS, 20/11/2009

Séminaire Julien Diard, « Modélisation Bayésienne et mesures d'erreurs », atelier biblio du LPNC, 7/10/08

Séminaire Estelle Gilet, Julien Diard et Pierre Bessière, « Modélisation Bayésienne d'un système sensori-moteur : Application à l'écriture », Journée du LPNC, 1er juillet 2008

Séminaire Julien Diard, Estelle Gilet, Eva Simonin et Pierre Bessière, « Incremental Learning of Bayesian Sensorimotor Models: from Low-level Behaviors to the Large-scale Structure of the Environment », Workshop BACS « Bayesian models and Contraction theory », Janvier 2008, Chamonix, France

Tutoriel Plateforme AFIA, Association Française d'Intelligence Artificielle, Tutoriel « Modélisation, apprentissage et inférence Bayésiens », Grenoble, 2/07/07

Séminaire Estelle Gilet et Julien Diard ; « Bayesian modeling and model selection for sensorimotor systems », projet PICS/CNRS, Paris, 23/11/06

Séminaire Julien Diard ; « Introduction à la modélisation Bayésienne », séminaire invité au laboratoire LIS / INPG, Grenoble, 21/11/06

Séminaire Panagiota Panagiotaki et Julien Diard ; « Cognitive Strategies of Spatial Encoding in Humans », kick-off meeting du projet européen BACS, Tübingen, Allemagne, 14/03/06

Séminaire Matthieu Lafon et Julien Diard ; « Projet Bayes City », LPPA, Paris, 09/03/06

Séminaire Alain Berthoz et Julien Diard ; « Are landmark-based and geometry-based human navigation strategies hierarchically articulated? A Bayesian modelling perspective », Bayesian Cognition workshop, Paris, 16/01/06

Séminaire Julien Diard ; « Programmation et modélisation bayésienne : navigation robotique et humaine », séminaire de l'UFR Informatique et Mathématiques Appliquées, UJF, Grenoble, 01/12/05

Séminaire Julien Diard ; « Learning hierarchies of Bayesian Maps for robotic navigation », projet PICS/CNRS, Singapour 25/11/05

Séminaire Julien Diard ; « Représentations probabilistes de l'espace : la Carte Bayésienne. Applications à la robotique mobile et à la modélisation du vivant », séminaire de l'équipe SIAMES, IRISA de Rennes, 14/04/2005

Séminaire Julien Diard ; « Le projet européen BIBA : Bayesian Inspired Brains and Artefacts » ; Journée du Pôle Rhône-Alpes de Sciences Cognitives (PRASC), Grenoble, 17/12/2004.

Séminaire Julien Diard ; « Probabilistic, bio-inspired and hierarchical models of navigation – applications to cognitive modeling and robotic manipulation » ; Journée du projet international de collaboration scientifique (PICS CNRS) LPPA-GRAVIR-NUS, Grenoble, 9/11/2004.

Séminaire Julien Diard ; « Modèles hiérarchiques probabilistes et bio-inspirés de la navigation ; applications à la modélisation du vivant et à la robotique mobile » ; séminaire invité au Laboratoire de Psychologie et NeuroCognition, Grenoble, 30 mars 2004.

Logiciel Pendant le stage post-doctoral, développement d'un logiciel d'expérimentation en psychophysique pour l'étude des capacités humaines de navigation. Utilisation du logiciel de réalité virtuelle Virtools (Virtools SA).

Candidature Prix de thèse Spécif 2003.

Séminaires 3 présentations aux écoles du projet européen BIBA (18/02/02 à Chichilianne, 01/07/02 à Moudon (Suisse), 16/01/03 à Combloux).

Logiciel Pendant la thèse, développement d'une maquette d'architecture de contrôle robotique basé sur les Cartes Bayésiennes. Utilisation du robot mobile Koala (K-Team) et d'une librairie d'inférence probabiliste.

Séminaire Julien Diard ; « The Bayesian Map : A probabilistic and hierarchical model for mobile robot navigation » ; séminaire invité à l'Université Nationale de Singapour (NUS), 03 décembre 2002.

Séminaire Julien Diard ; « Bayesian Maps and Navigation » ; groupe Modèle, Collège de France, 12 juin 2002.

Poster Julien Diard, Pierre Bessière et Olivier Lebeltel ; « Robotic Programming and space representation using a unified bayesian framework » ; International Symposium on Neural Control of Space Coding and Action Production ; March 22-24 2001.

Poster Une version abrégée du précédent a été présentée à la Journée RESCIF « La cognition : de l'artificiel au naturel » ; Réseau de sciences cognitives d'Ile-de-France ; Collège de France, 26/27 octobre 2001.

Démonstration robotique grand public Fête de la science 2001, Grenoble

Posters accompagnant les articles parus à MaxEnt 2000 et SAB 2000.

Démonstration robotique SAB 2000.

Séminaires internes Plusieurs séminaires internes à l'équipe Laplace et à l'INRIA Rhône-Alpes (étudiants de la halle robotique), également au LPPA.

Logiciel Pendant le DEA, développement d'une maquette d'architecture de contrôle robotique intégrant un apprentissage hiérarchique de comportements. Utilisation du robot mobile Khepera (K-Team).

Démonstrations robotiques grand public ComputerNacht, Paderborn Allemagne, 1999 (évènement en marge de IKW 99) ; TEC99, Grenoble.

Logiciel Pendant le Magistère, développement d'une librairie de visualisation de fonctions intégrée à un tableur probabiliste (2e année). Développement de comportements bayésiens pour le robot mobile Koala (3e année).

Atouts – Langues

- Anglais parlé et écrit courant (séjour de 9 mois (1997–1998) à Santa Barbara, Californie ; séjour de 8 mois (2003) à Singapour)
- Sports : volley-ball (en club ; pendant mon post-doc à NUS, j'ai été membre de la « NUS Staff Team »)
- Loisirs : littérature fantastique et science-fiction, en langue anglaise ; musiques progressives

Titre — Modélisation Bayésienne Algorithmique en Science Cognitive

Résumé — Dans le domaine de la modélisation des systèmes sensorimoteurs, qu'ils soient artificiels ou naturels, nous nous intéressons à la définition et à l'étude de modèles probabilistes structurés des fonctions et représentations cognitives. Dans ce but, nous utilisons le formalisme de la Programmation Bayésienne, développé initialement dans le domaine de la programmation robotique. Il offre un langage mathématiquement unifié pour exprimer et manipuler des connaissances, dans des modèles arbitrairement complexes.

Nous l'appliquons à la modélisation cognitive, pour obtenir des Modèles Bayésiens Algorithmiques de plusieurs systèmes perceptifs et moteurs. De cette manière, nous définissons le modèle BAP pour la lecture et l'écriture de lettres cursives isolées, le modèle BRAID pour la reconnaissance de mots, le modèle COSMO-Emergence pour l'émergence de codes de communication, le modèle COSMO-Perception pour la perception des syllabes, le modèle COSMO-Production pour la production de séquences de phonèmes.

Nous discutons enfin la place de la Modélisation Bayésienne Algorithmique dans le panorama actuel de la modélisation bayésienne en Science Cognitive, défendant le besoin d'une distinction claire entre les explications computationnelles et algorithmiques des fonctions cognitives, et proposant une méthodologie de comparaison de modèles pour explorer et contraindre les propriétés de modèles probabilistes complexes, d'une manière systématique.

Mots clés : Modélision bayésienne, Programmation bayésienne, Science Cognitive, lecture et écriture, perception et production de la parole.

Titre — Bayesian Algorithmic Modeling in Cognitive Science

Abstract — In the domain of modeling sensorimotor systems, whether they are artificial or natural, we are interested in defining and studying structured probabilistic models of cognitive functions and cognitive representations. To do so, we use the Bayesian Programming framework, originally developed in the domain of robotic programming. It provides a mathematically unified language to express and manipulate knowledge, in arbitrarily complex models.

We apply it to cognitive modeling, obtaining Bayesian Algorithmic Models of several perception and action systems. We thus define the BAP model for isolated cursive letter reading and writing, the BRAID model for word recognition, the COSMO-Emergence model for communication code emergence, the COSMO-Perception model for syllable perception, and the COSMO-Production model for phoneme sequence production.

We then discuss the place of Bayesian Algorithmic Modeling in the current panorama of Bayesian modeling in Cognitive Science, arguing for the need for a clear distinction between computational and algorithmic accounts of cognitive functions, and advocating a model comparison methodology for exploring and constraining the properties of complex probabilistic models in a formally principled manner.

Keywords — Bayesian modeling, Bayesian Programming, Cognitive Science, reading and writing, speech perception and production.