

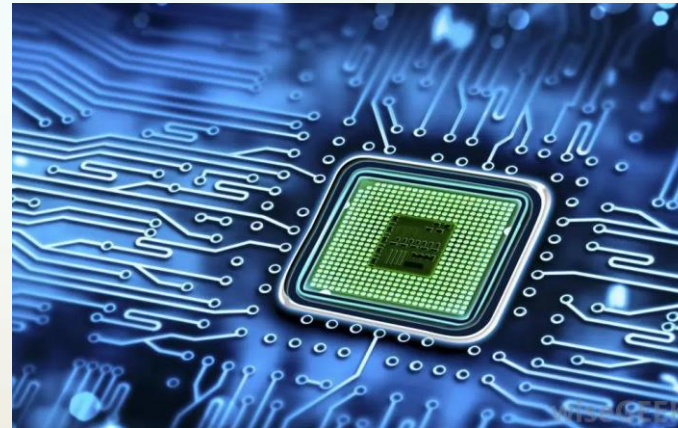
Aprendizaje Social

Ingeniería para problemas de Hoy

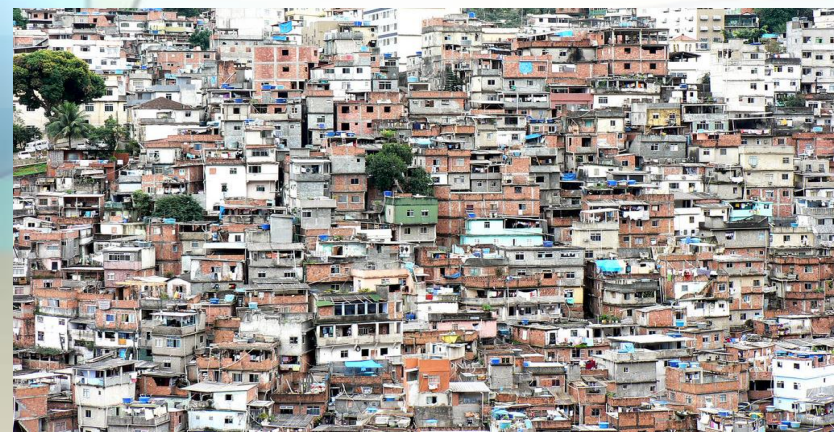
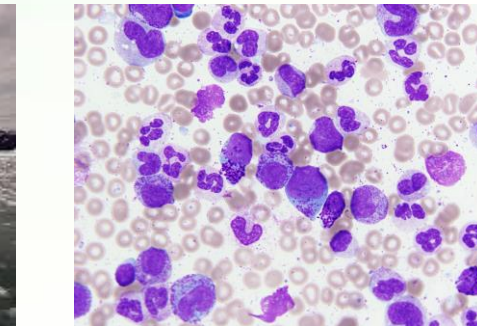


Marco Aurelio Alzate Monroy
Universidad Distrital
Maestría en Ciencias de la
Información y las Comunicaciones

Problemas de ingeniería del siglo XX

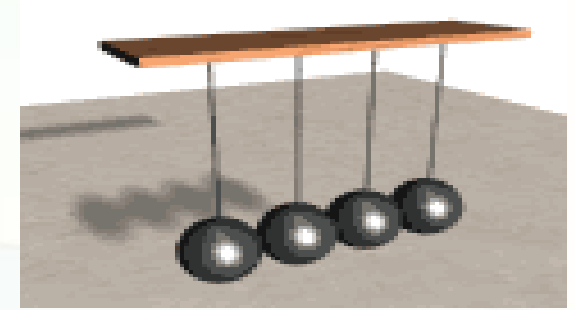
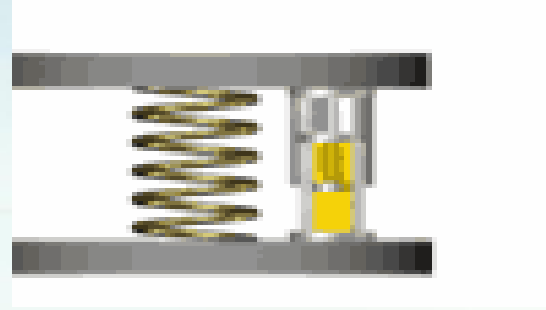
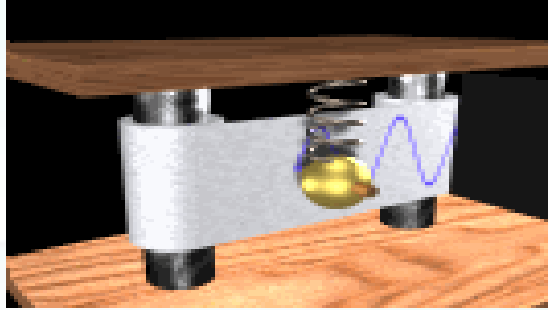
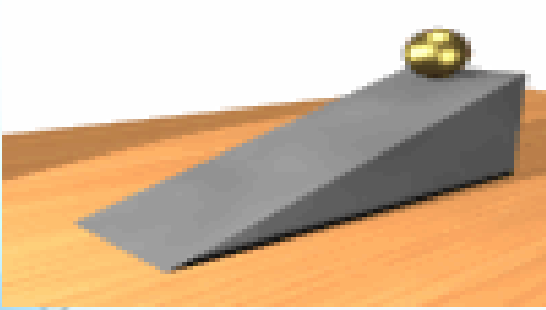


Retos de ingeniería del siglo XXI



Polución del aire, el agua y la tierra, Calentamiento global y cambio climático, Reducción de recursos, sobrepoblación y seguridad alimentaria, Disposición de desechos, Pérdida de biodiversidad y deforestación, Crecimiento de las ciudades, Salud pública, Democracia, Terrorismo, Epidemias, Cáncer, Conflictos sociales, respuesta a desastres, crisis financieras, energía sostenible, transporte urbano, ...

Ingeniería de sistemas simples pero complicados

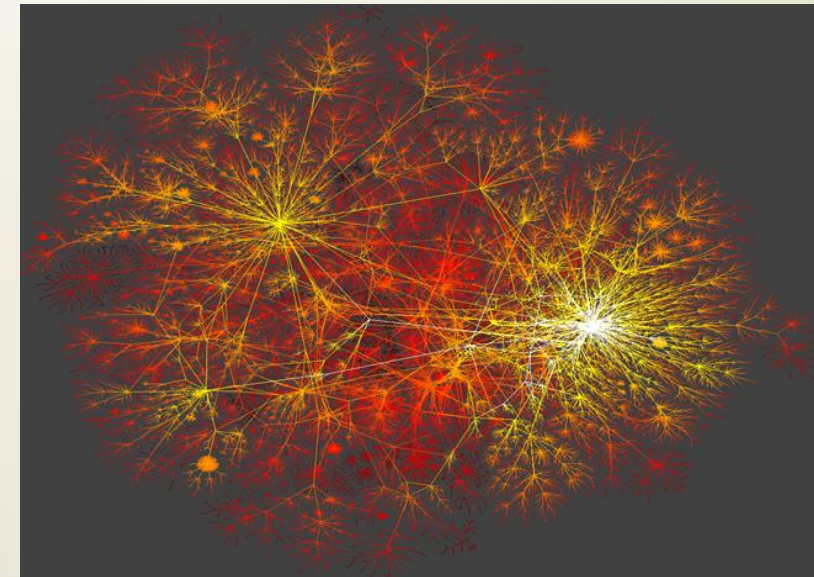
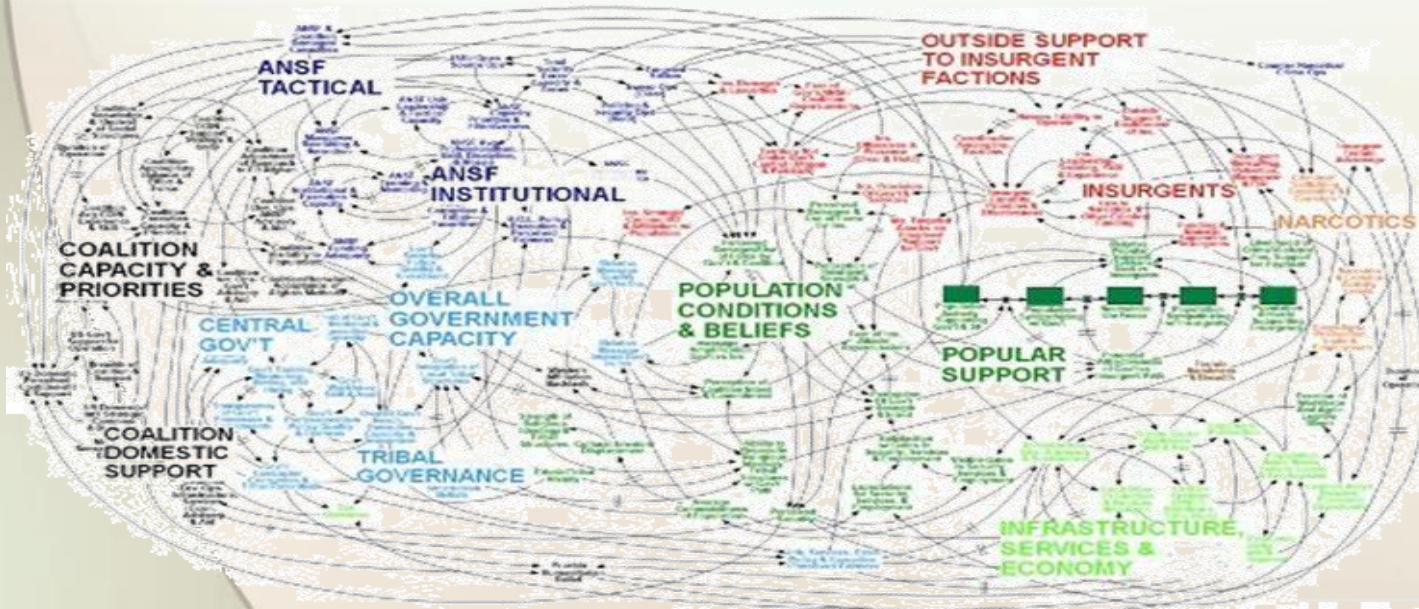
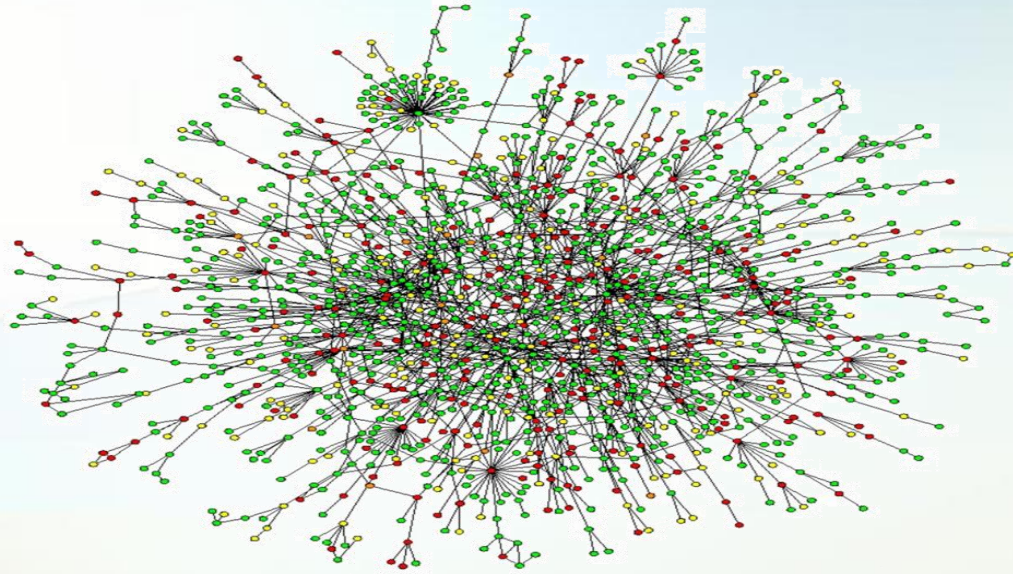


Los ingenieros conocemos o desarrollamos modelos matemáticos de sistemas dinámicos simples

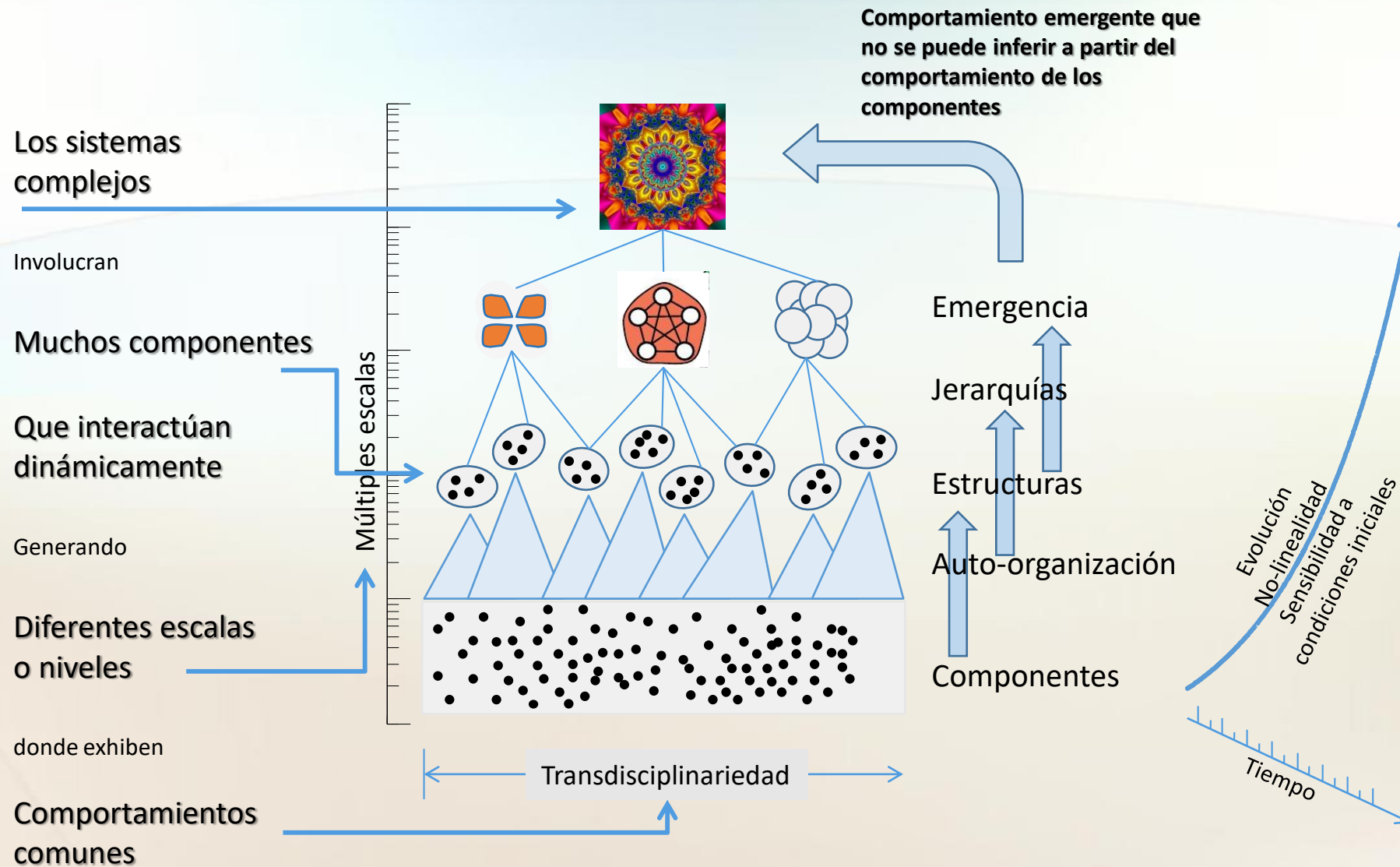


Con los cuales podemos diseñar sistemas complicados: Muchos componentes, cada uno con una funcionalidad específica, que no se adaptan individualmente ¿Tolerancia a fallas? ¡Redundancia! ¿Robustez? ¡Control en lazo cerrado! La causalidad sigue siendo precisa, el sistema debe ser predecible...

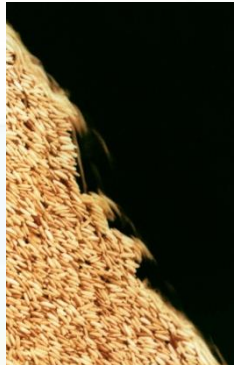
Modelos matemáticos de sistemas dinámicos Complejos



Sistemas complejos



Origen de la complejidad



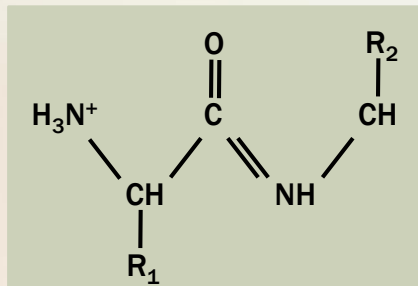
Componentes sencillos
Interacciones sencillas



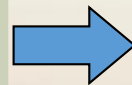
Comportamiento emergente
auto-organizado

SOC/EOC

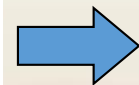
Self-organized criticality/
Edge of chaos
(propio de la mecánica estadística)



Componentes sencillos
Interacciones sencillas



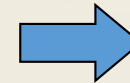
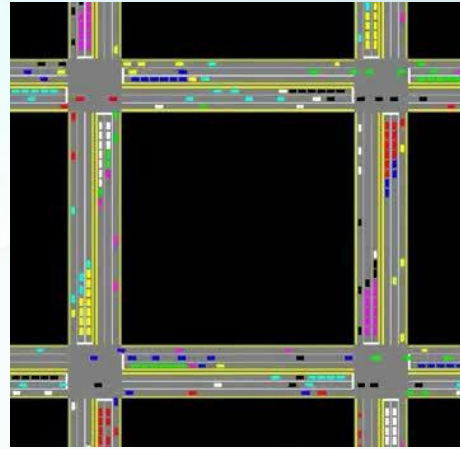
Comportamiento emergente
auto-organizado



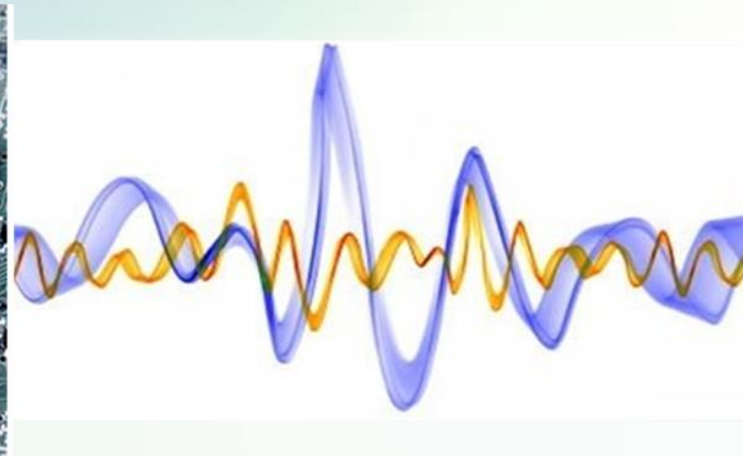
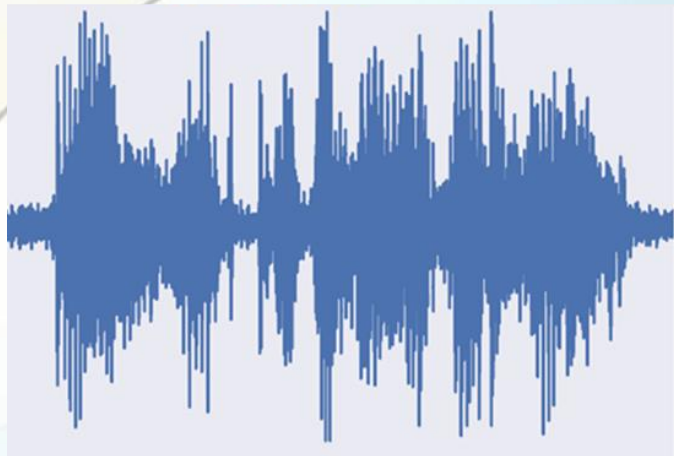
HOT

Highly Optimized Tolerance
(propio de sistemas optimizados)
Vida: Optimización por Evolución,
Tecnología: Optimización por Diseño

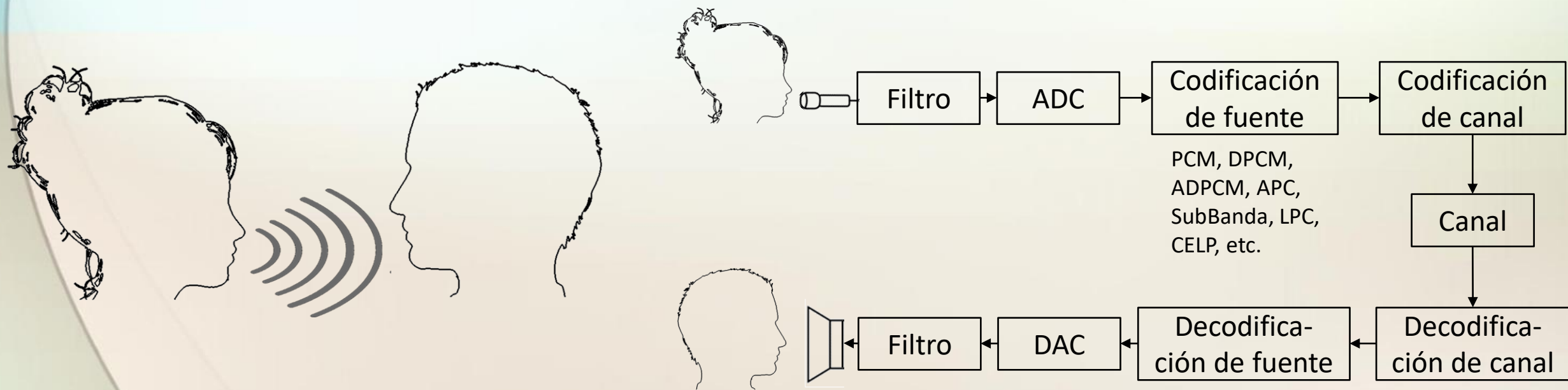
Algunos Sistemas complejos



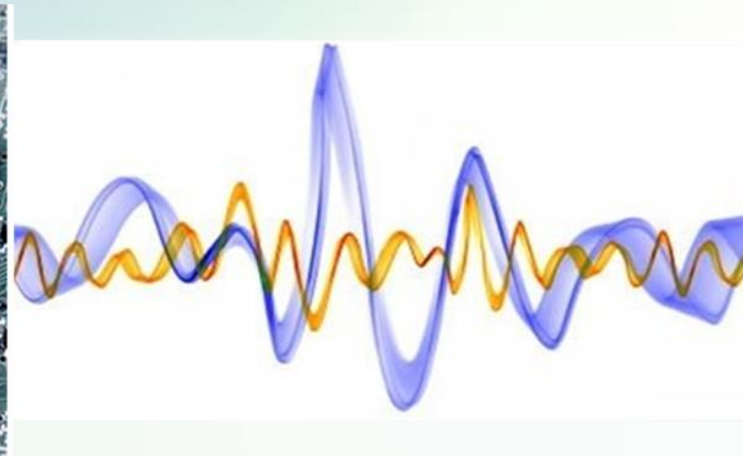
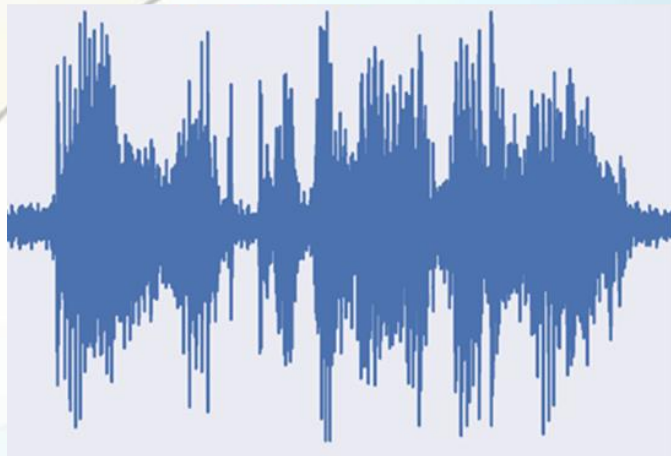
PROCESAMIENTO DIGITAL DE SEÑALES



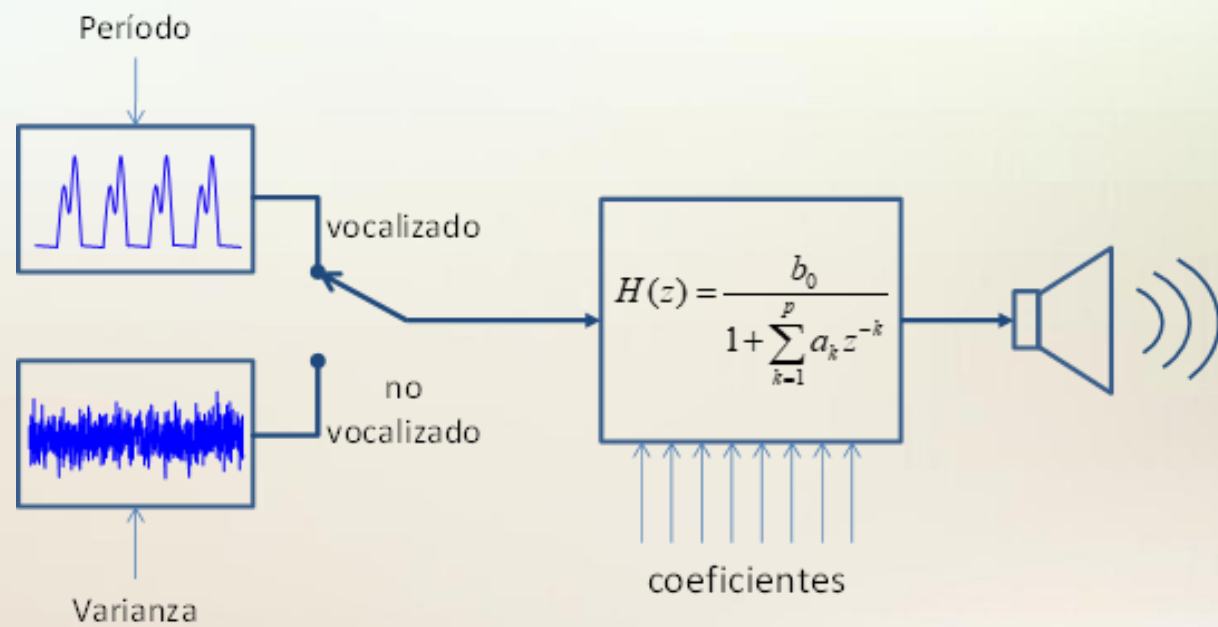
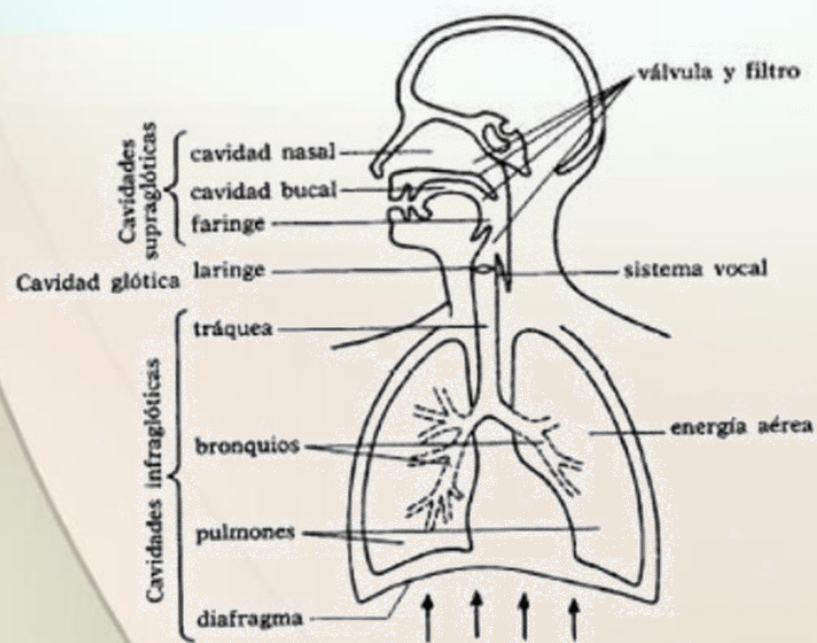
El DSP consiste, fundamentalmente, en extraer información de fenómenos físicos mediante procesos de inferencia



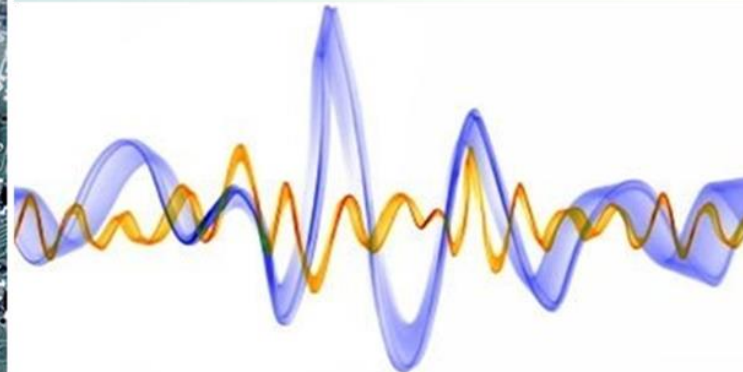
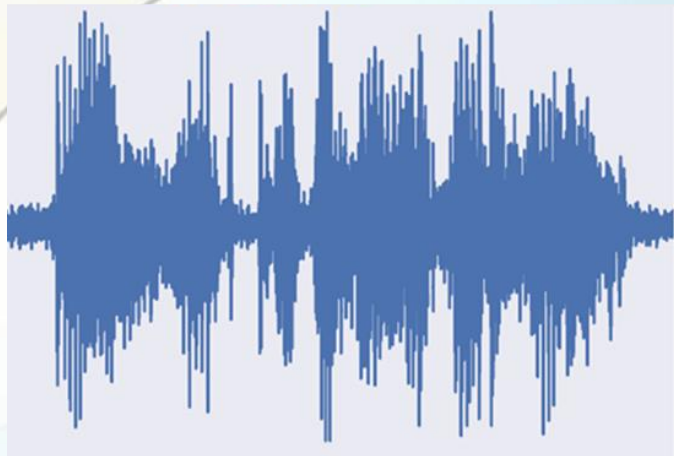
PROCESAMIENTO DIGITAL DE SEÑALES



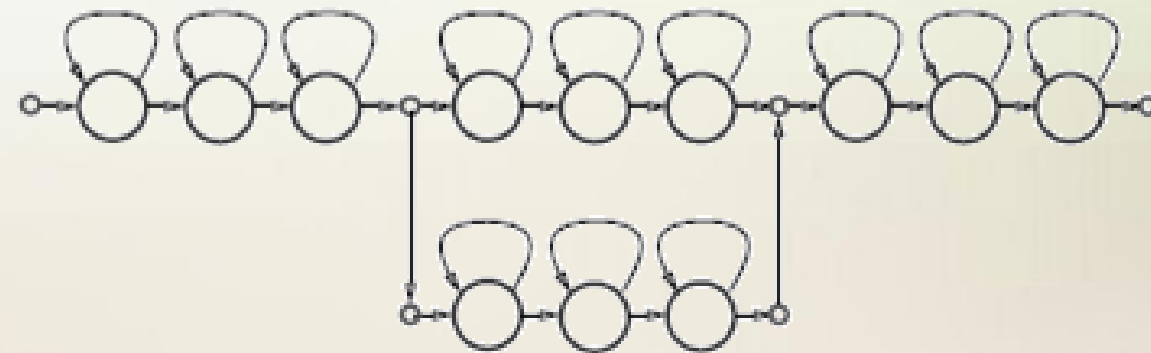
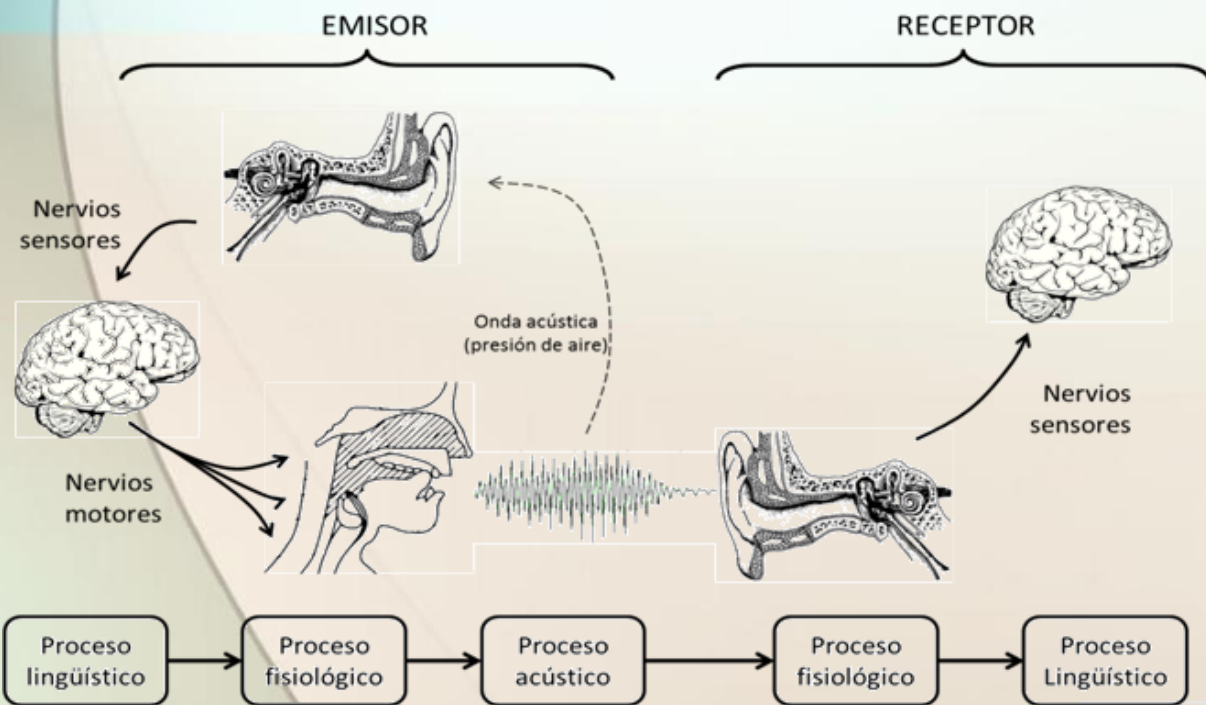
El DSP consiste, fundamentalmente, en extraer información de fenómenos físicos mediante procesos de inferencia



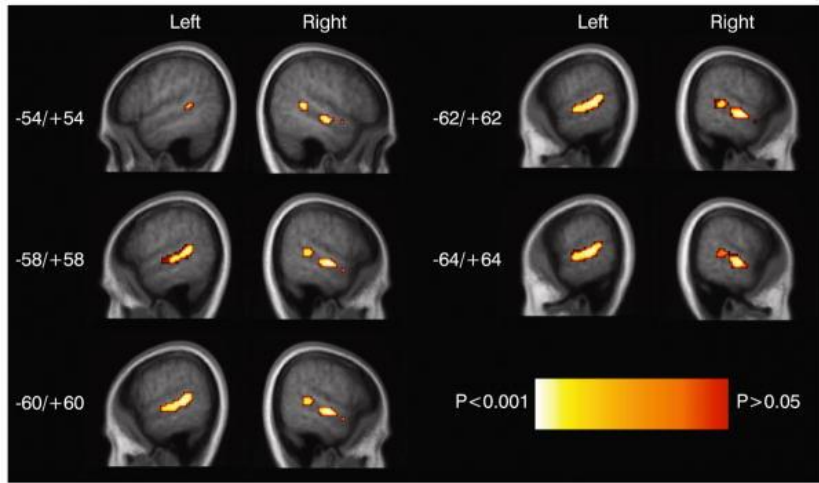
PROCESAMIENTO DIGITAL DE SEÑALES



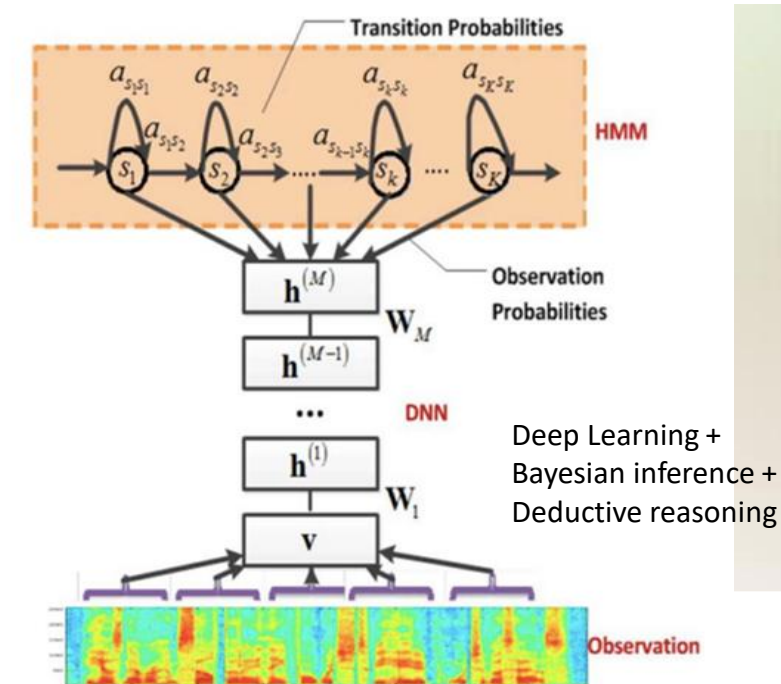
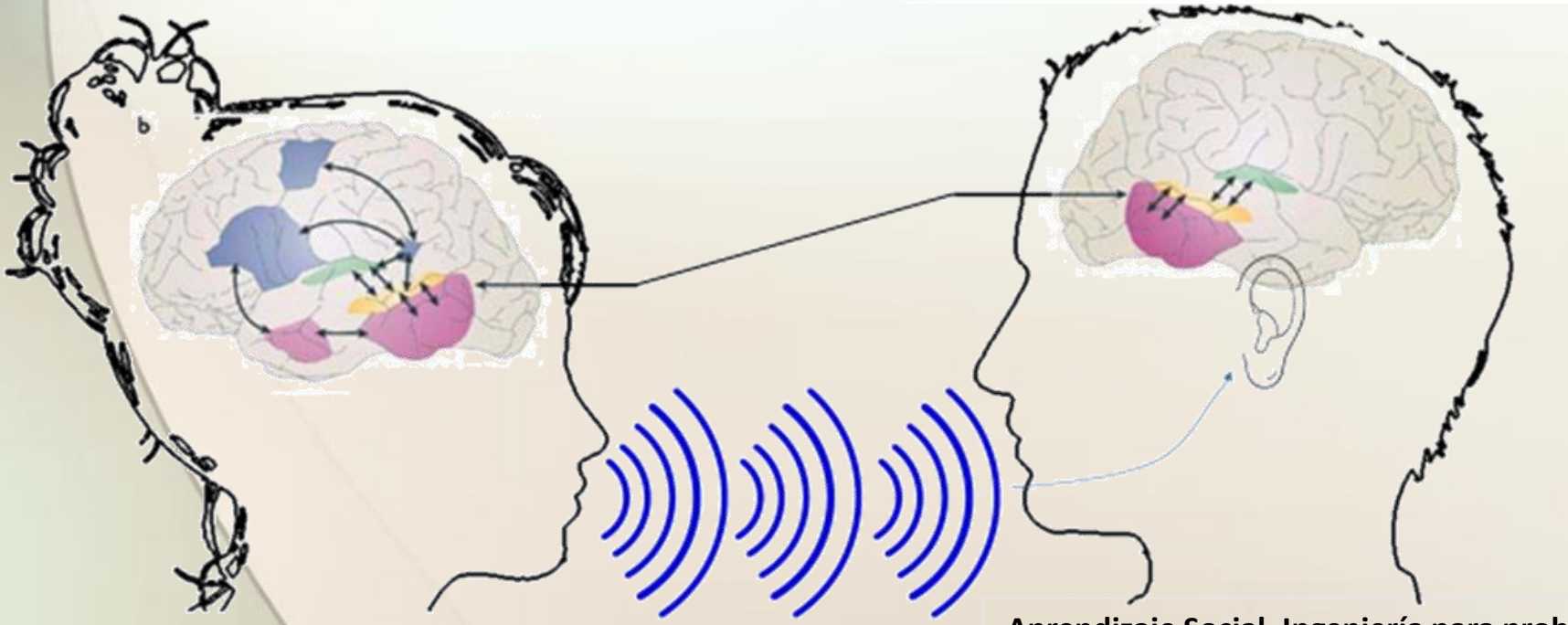
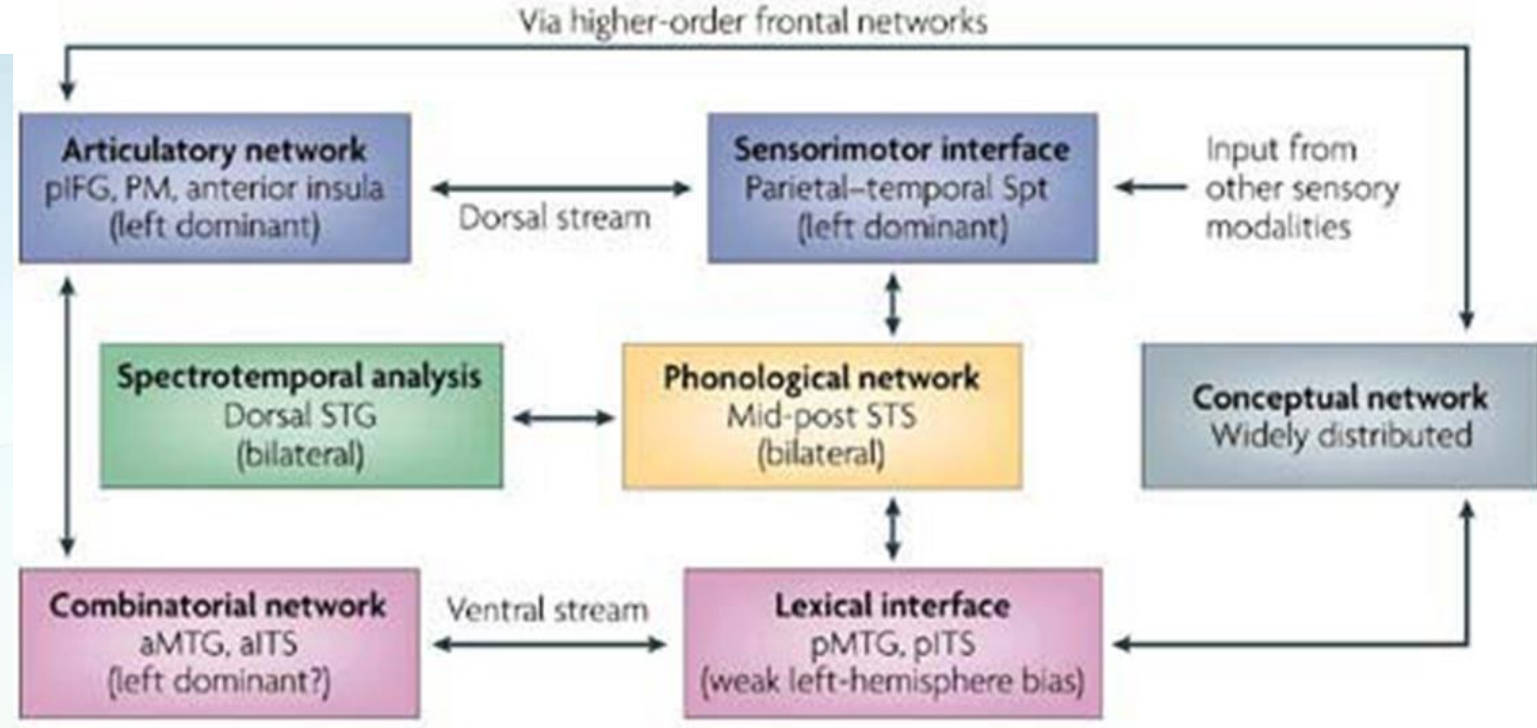
El DSP consiste, fundamentalmente, en extraer información de fenómenos físicos mediante procesos de inferencia



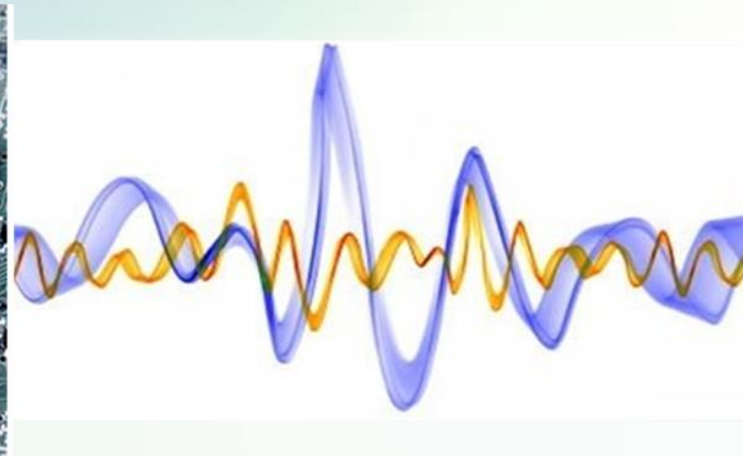
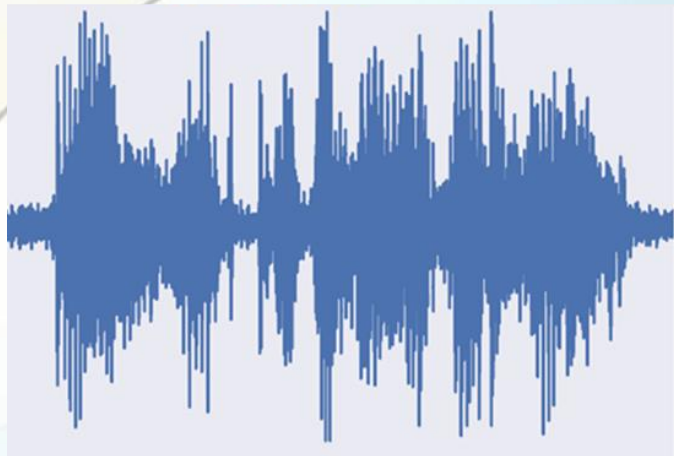
Modelo Lingüístico HMM



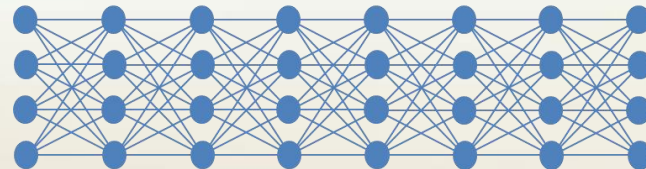
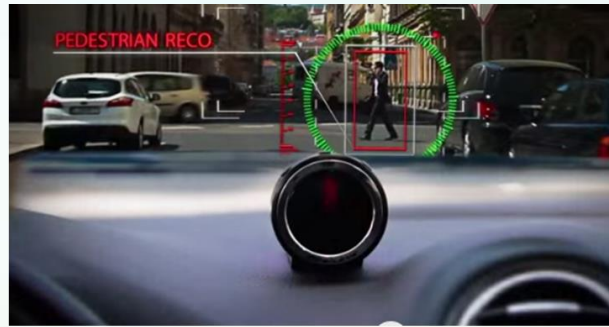
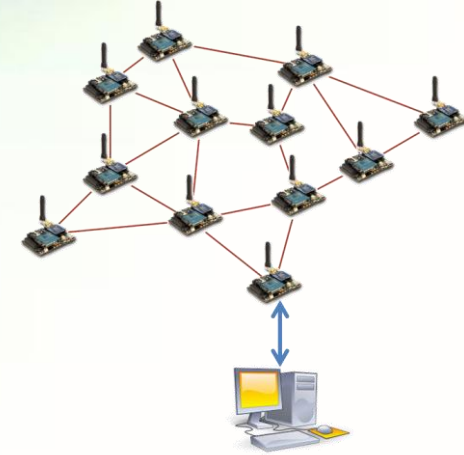
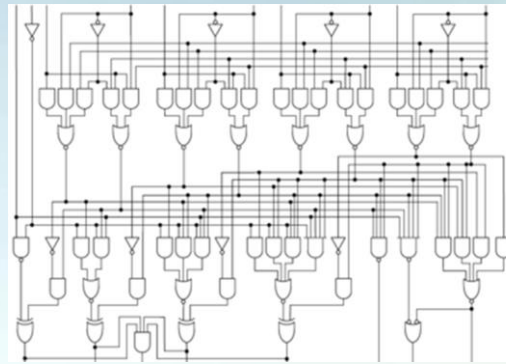
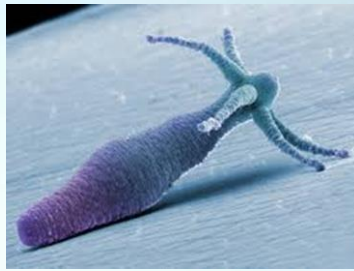
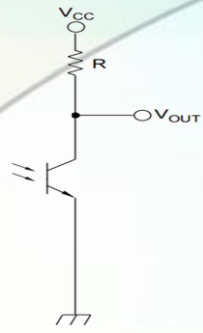
Z. Zheng, K. Munhall and I. Johnsrude "Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production" Journal of Cognitive Neuroscience, August 2010, volume 22, number 8m pp. 1770-1781



PROCESAMIENTO DIGITAL DE SEÑALES

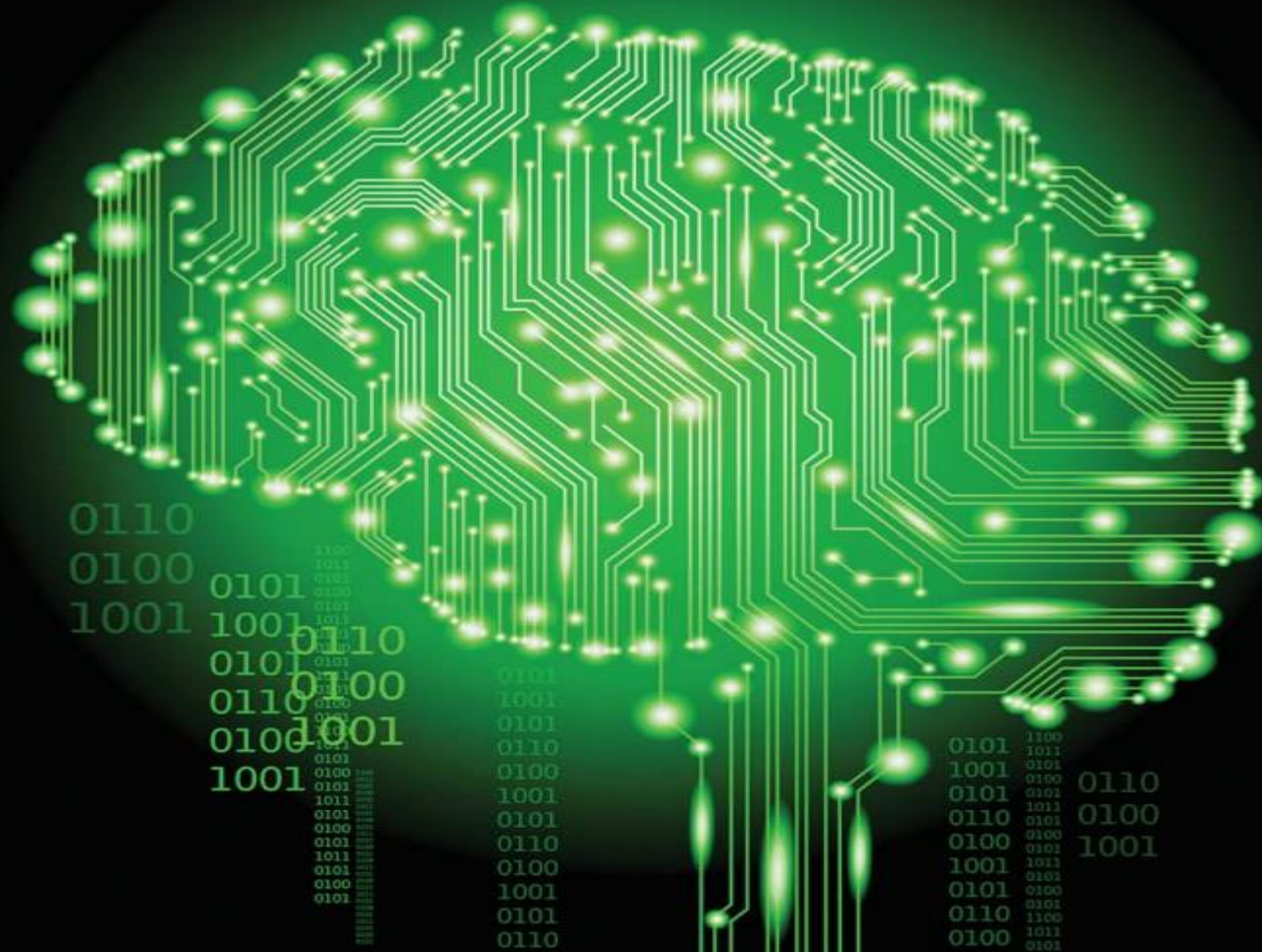


Procesamiento
de Información
en Sistemas
Cognitivos



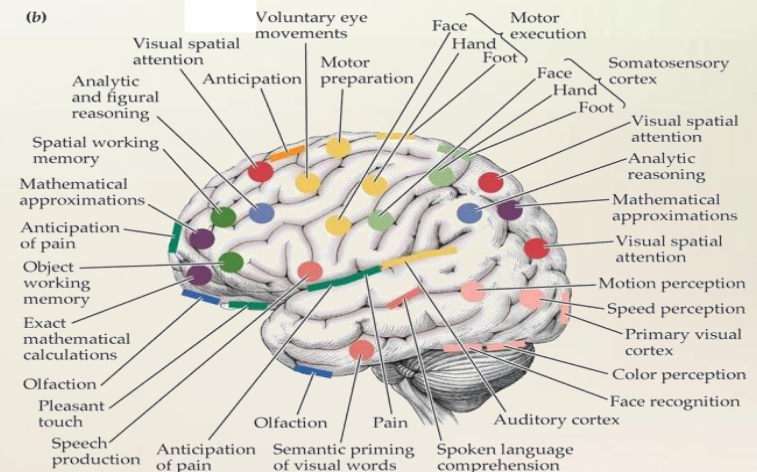
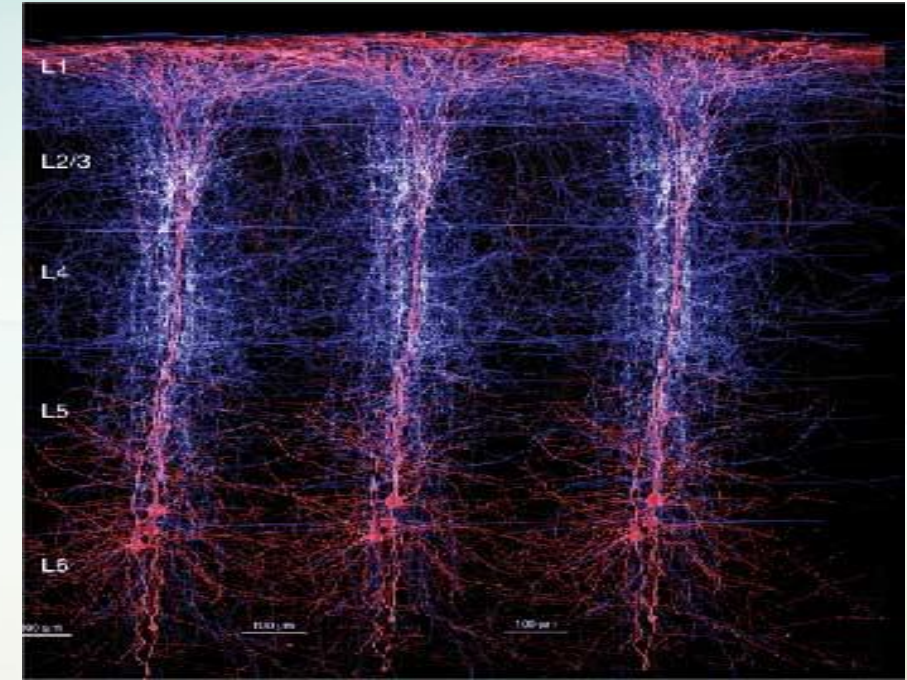
Con el estado actual de las técnicas de procesamiento digital de señales podemos construir agentes cognitivos de gran complejidad

El cerebro humano como paradigma de sistema cognitivo



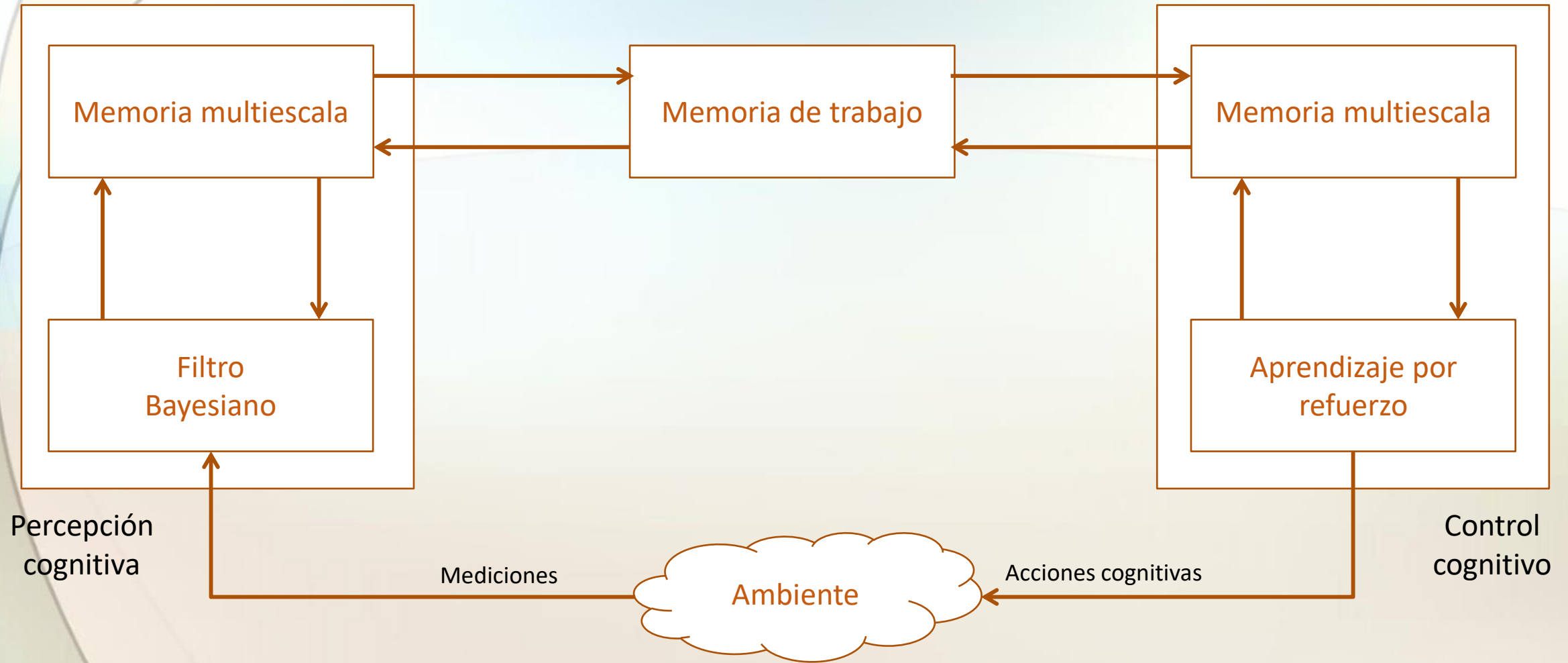
El cerebro humano como paradigma de sistema cognitivo

- Joaquín Fuster
- Un altísimo porcentaje del procesamiento de información en el cerebro se realiza en la corteza cerebral
- La corteza cerebral está organizada en minicolumnas (ordenamientos verticales de entre 80 y 240 neuronas). Existen cerca de 200 millones de minicolumnas.
- Las columnas corticales parecen ser unidades jerárquicas de reconocimiento de patrones. La memoria y el conocimiento se representan en la interacción distribuida, y altamente sobrelapada, de estas minicolumnas corticales.
- Aunque no tienen una estructura uniforme, todas parecen usar un algoritmo básico de procesamiento de información: Diferentes tipos de entradas sensoriales (visual, auditiva, etc.) se codifican de una manera estándar antes de alimentar este algoritmo básico de procesamiento.
- Las funciones cognitivas (**percepción, atención, memoria, inteligencia, lenguaje y acción**) se basarían en transacciones neuronales al interior de las minicolumnas y entre ellas.
- El paradigma de Fuster se convierte en un marco de referencia para los sistemas dinámicos cognitivos.



BIOLOGICAL PSYCHOLOGY, Fourth Edition, Figure 1.11 (Part 2) © 2004 Sinauer Associates, Inc.

Estructura de un agente cognitivo



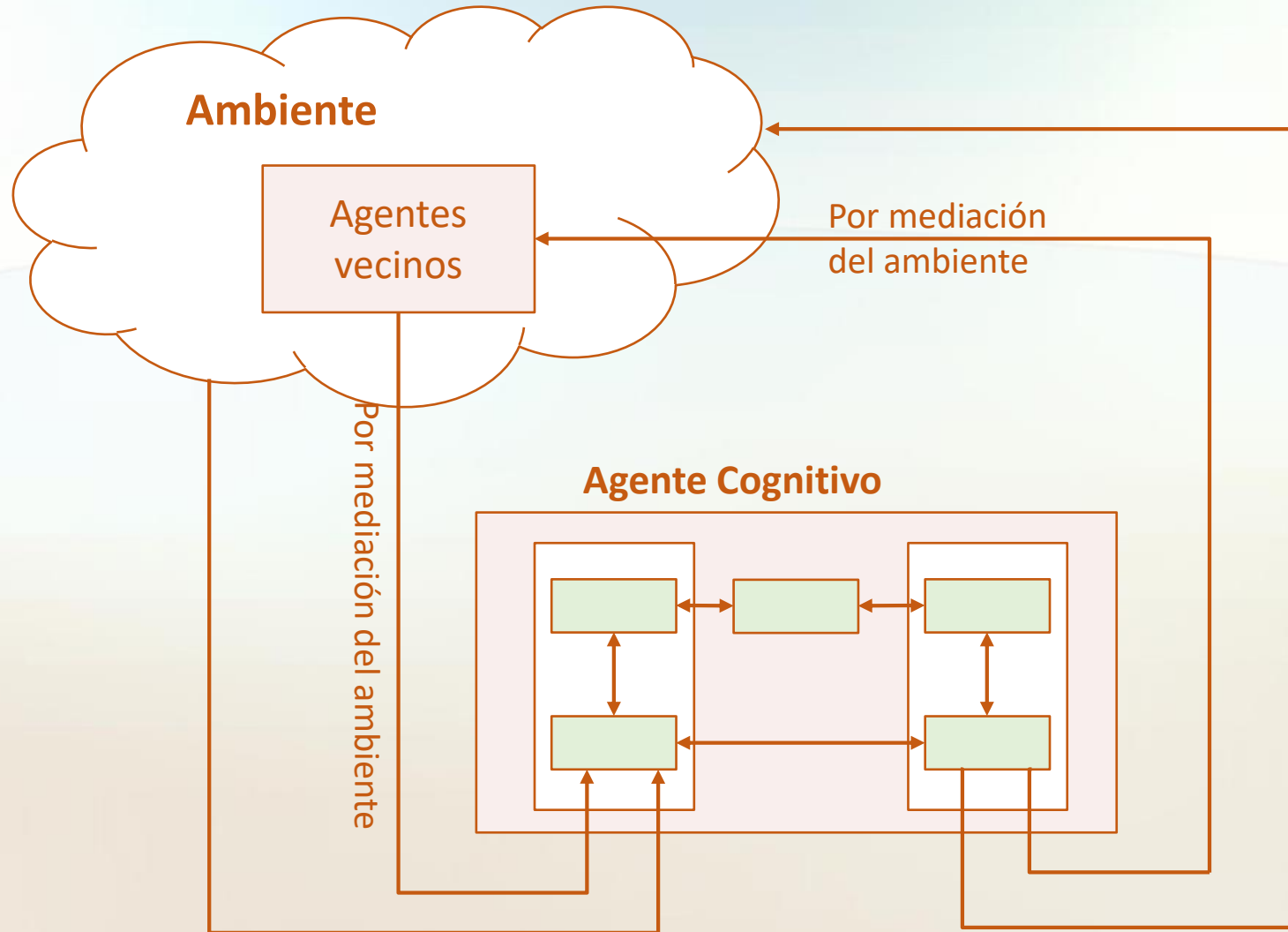
Agente Cognitivo

Un Agente Cognitivo es

- Un **Sistema Complejo**
- Capaz de desarrollar un **comportamiento emergente**
- Al **procesar información** en el transcurso del tiempo
- Mediante las siguientes funciones
 - **Percepción** del ambiente
 - **Aprendizaje** del ambiente
 - **Adaptación** a sus variaciones estadísticas
 - Construcción de un **modelo predictivo** sobre un conjunto prescrito de aspectos de interés del ambiente
 - Desarrollo de **reglas de comportamiento** que le permitan actuar sobre el ambiente
- Realizadas en **tiempo real**
- Con el propósito de llevar a cabo algunas **tareas prescritas**
- De manera **efectiva, eficiente y confiable**
- A pesar de las **incertidumbres en el ambiente.**

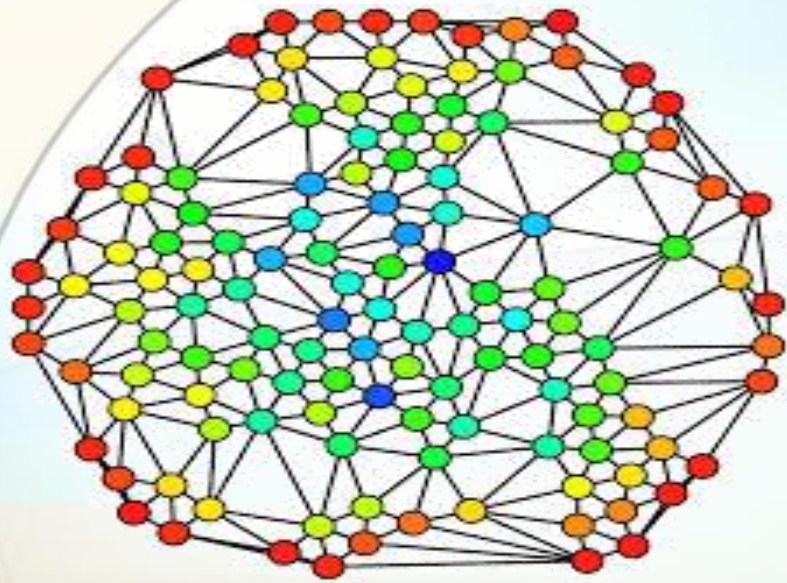
→ **Técnicas de DSP (incluyendo CI y DL)**

Sistemas dinámicos cognitivos



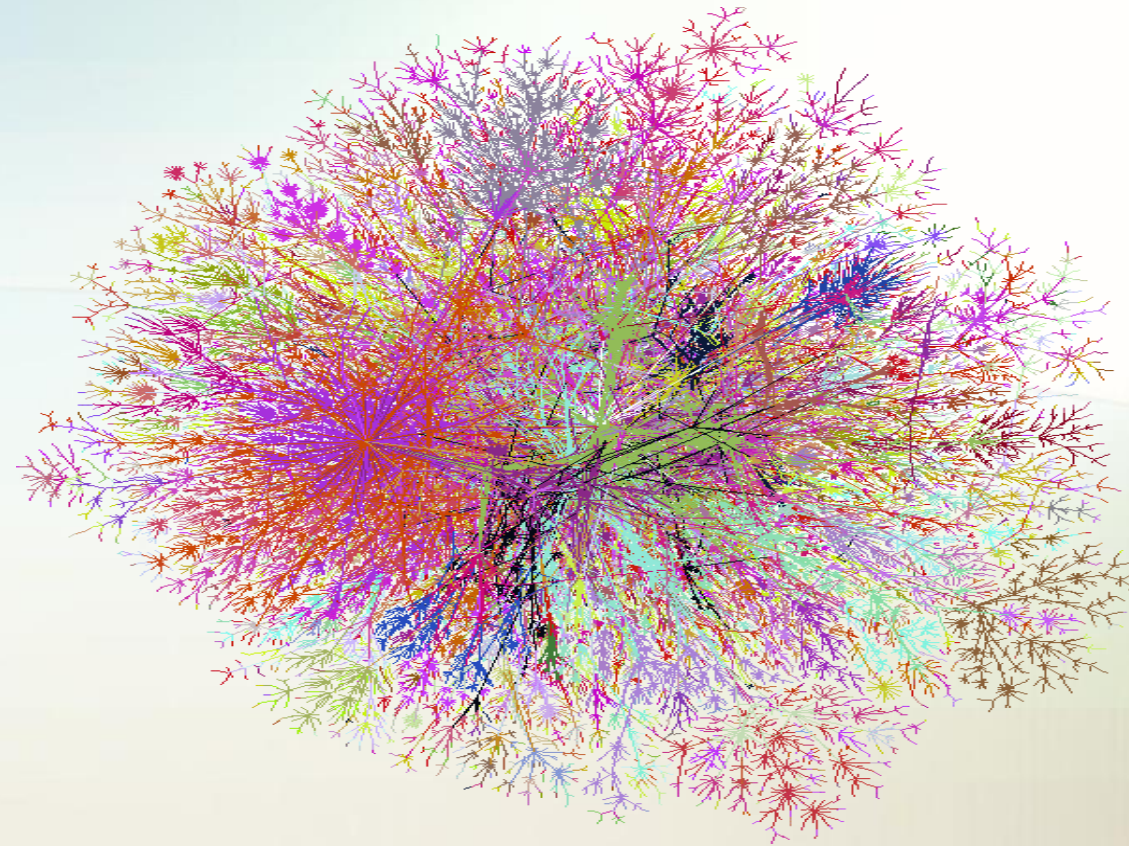
Redes cognitivas

- Agentes distribuidos espacialmente (nodos) con relaciones entre ellos (enlaces) que forman una red, una topología de conexiones.
- Los agentes tienen capacidades de percepción y acción. Entre ellas, median capacidades adicionales de memoria, inteligencia, atención y lenguaje.
- Los agentes se pueden mover y la topología puede cambiar con el tiempo
- Los agentes perciben su entorno, asocian los estímulos con modelos mentales de su realidad y actúan de acuerdo con el estímulo y el modelo. En el proceso pueden aprender ajustando su propio modelo del mundo.
- Los agentes cooperan localmente mediante procesamiento distribuido de información a través de la red.
- Logran procesamiento descentralizado de información, optimización descentralizada (de objetivos globales especificados o no), inferencia estadística descentralizada de datos remotos, etc.



Agentes cognitivos interactuando en redes
Inteligencia colectiva superior a la suma de las
inteligencias individuales

¿Procesamiento digital de señales sobre grafos?



Procesamiento de información en redes complejas

¡Modelo matemático de aprendizaje social!

Procesamiento de señales

- Extraer información de fenómenos físicos mediante procesos de inferencia
- Voz, imágenes, video, comunicaciones, percepción remota, ingeniería biomédica, interfaces cerebro-máquina, etc.
- Control (técnico) de los medios a través de los cuales la gente se informa
- Inferencia de qué es lo que valora la gente y cómo se influencia sus creencias
- Si el procesamiento de señales condujo al desarrollo de las redes de comunicaciones a través de las cuales las personas comparten sus opiniones, sus sentimientos, sus conocimientos, con sinceridad o sin ella ¿No es hora de que el procesamiento de señales se preocupe por estas aplicaciones sociales?
- ¿Cómo es que los individuos afectan las decisiones de otros individuos? ¿Cómo es que las sociedades generan líderes cuya aceptación crece y decrece con el tiempo? ¿Cómo es que las sociedades conducen sus creencias y sus valores?

Procesamiento Digital de Señales entre agentes autónomos distribuidos

- ¿Cómo y por qué aparece el comportamiento organizado a nivel de grupo, a partir de las interacciones entre agentes autónomos?
- ¿Qué topologías permiten la emergencia de orden a nivel superior a partir de interacciones de nivel inferior?
- ¿Cómo se procesa la información durante el proceso de difusión de conocimiento que conduce a este fenómeno?
- Diferentes disciplinas intentan dilucidar diferentes aspectos de estas preguntas fundamentales (biología evolutiva, etología y sociología, biología física, ...)
- En DSP se estudian estrategias descentralizadas de procesamiento de información para dotar a las redes complejas con capacidades de adaptación y aprendizaje en tiempo real.

IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS

Editorial

A PUBLICATION OF
THE IEEE SIGNAL PROCESSING SOCIETY
THE IEEE COMMUNICATIONS SOCIETY
THE IEEE COMPUTER SOCIETY



MARCH 2015 VOLUME 1 NUMBER 1 ITSIBW (ISSN 2373-776X)

Editorial	<i>P. M. Djurić</i>	1
PAPERS		
<i>Adaptation, Detection, Estimation, and Learning</i>		
Information-Sharing Over Adaptive Networks With Self-Interested Agents	<i>C.-K. Yu, M. van der Schaar, and A. H. Sayed</i>	2
Diffusion and Superposition Distances for Signals Supported on Networks	<i>S. Segarra, W. Huang, and A. Ribeiro</i>	20
Wireless Compressive Sensing Over Fading Channels With Distributed Sparse Random Projections	<i>T. Wimalajeewa and P. K. Varshney</i>	33
Distributed Widely Linear Kalman Filtering for Frequency Estimation in Power Networks	<i>S. Kanna, D. H. Dini, Y. Xia, S. Y. Hui, and D. P. Mandic</i>	45
<i>Communications, Networking, and Sensing</i>		
A Scalable Feedback Mechanism for Distributed Nullforming With Phase-Only Adaptation	<i>A. Kumar, R. Mudumbai, S. Dasgupta, M. M.-U. Rahman, D. R. Brown III, U. Madhow, and T. P. Bidigare</i>	58
EDICS—Editor's Information Classification Scheme		71
Information for Authors		72

THE PAST decade has witnessed an outbreak in research on techniques and methodologies for information processing over networks. Much of this progress has been propelled by advances in network systems and in equipment and methodologies for the collection and analysis of data.

Network systems are generally composed of diverse interconnected elements and give rise to many interesting signal and information-processing challenges. In general, the network elements are distributed in space and they are subjected to information collected locally and from their neighbors. The elements can then use the available information to solve a variety of tasks, such as estimation, detection, prediction, monitoring, and filtering, and to make decisions in a distributed manner. These operations can be repeated over and over again to enable learning, adaptation, and even adjustment of the network topology and the manner by which the agents interact with each other.

The key to understanding network systems and to improving their design are methods for signal and information processing over networks. On the one hand, distributed statistical inference, learning, adaptation, optimization, understanding of relationships, resource allocation, and diffusion of information

with time and space are some of the key fundamental topics that arise in this context. On the other hand, applications abound in many fields including sensing networks, communications, smart grids, transportation networks, life sciences, the Internet and the Internet of Things, robotics, wireless health, the stock market and finance, and ecology, as well as in the modeling of animal and human behavior.

In 2015, the IEEE Signal Processing Society is launching a new journal, *Transactions on Information and Signal Processing Over Networks*. This is its first issue. The new journal is shared with the Communications Society (20%) and the Computer Society (10%). We are confident that it will grow to become an important venue for disseminating research of the highest scholarly standards. We are delighted that we are starting it and are optimistic and excited about its prospects.

PETAR M. DJURIĆ, *Professor*
Department of Electrical and Computer
Engineering
Stony Brook University
Stony Brook, NY 11794 USA

a positive left eigenvector v whose corresponding eigenvalue is equal to one.⁴ Left multiplying both sides of (7) by v ,

$$v^T z_t(\theta) = v^T z_{t-1}(\theta) + \sum_{i=1}^n v_i a_{ii} \left(\frac{\ell_i(\omega_i, t|\theta)}{\ell_i(\omega_i, t|\theta^*)} - 1 \right) z_{i,t-1}(\theta).$$

Thus,

$$\begin{aligned} E^* [v^T z_t(\theta) | \mathcal{F}_{t-1}] \\ &= v^T z_{t-1}(\theta) + \sum_{i=1}^n v_i a_{ii} \\ &\quad \times \left(E^* \left[\frac{\ell_i(\omega_i, t|\theta)}{\ell_i(\omega_i, t|\theta^*)} | \mathcal{F}_{t-1} \right] - 1 \right) z_{i,t-1}(\theta) \\ &= v^T z_{t-1}(\theta) \end{aligned}$$

where the second equality is since

$$E^* \left[\frac{\ell_i(\omega_i, t|\theta)}{\ell_i(\omega_i, t|\theta^*)} | \mathcal{F}_{t-1} \right] = \sum_{s_i \in S_i} \ell_i(s_i|\theta^*) \frac{\ell_i(s_i|\theta)}{\ell_i(s_i|\theta^*)} = 1.$$

Since $v^T z_t(\theta) \geq 0$, this also implies that

$$E^* [v^T z_t(\theta)] = E^* [v^T z_t(\theta)] = v^T z_0(\theta) < \infty.$$

Therefore, $v^T z_t(\theta)$ is a martingale with respect to the filtration \mathcal{F}_{t-1} . It is also non-negative since v and $z_t(\theta)$ are positive as a result of the Perron-Frobenius theorem and our assumption above, respectively. Hence, by the martingale convergence theorem, $v^T z_t(\theta)$ converges P^* -almost surely. Consequently, the martingale difference converges to zero P^* -almost surely; i.e.,

$$\sum_{i=1}^n v_i a_{ii} \left(\frac{\ell_i(\omega_i, t|\theta)}{\ell_i(\omega_i, t|\theta^*)} - 1 \right) z_{i,t-1}(\theta) \rightarrow 0 \quad \text{as } t \rightarrow \infty.$$

On the other hand, since $\ell(\cdot|\theta^*)$ has full support, for all $s \in S$ the event $\{\omega_t = s$ for infinitely many $t\}$ has P^* -probability one. Therefore, for all $(s_1, \dots, s_n) \in S_1 \times \dots \times S_n$,

$$\sum_{i=1}^n v_i a_{ii} \left(\frac{\ell_i(s_i|\theta)}{\ell_i(s_i|\theta^*)} - 1 \right) z_{i,t-1}(\theta) \rightarrow 0 \quad \text{as } t \rightarrow \infty$$

with P^* -probability one.

Now choose the profile $\hat{s}_\theta = (\hat{s}_{1,\theta}, \dots, \hat{s}_{n,\theta}) \in S$ where $\hat{s}_{i,\theta} \in \arg \max_{s_i \in S_i} \ell_i(s_i|\theta)/\ell_i(s_i|\theta^*)$. Since $\ell_i(\cdot|\theta)$ is a probability distribution over S_i , we can conclude that $\ell_i(\hat{s}_{i,\theta}|\theta)/\ell_i(\hat{s}_{i,\theta}|\theta^*) \geq 1$ with equality if and only if agent i cannot distinguish θ from θ^* , i.e., $\theta \in \Theta_i^*$. Therefore, since a_{ii} and v_i are positive,

$$\left(\frac{\ell_i(\hat{s}_{i,\theta}|\theta)}{\ell_i(\hat{s}_{i,\theta}|\theta^*)} - 1 \right) z_{i,t-1}(\theta) \rightarrow 0 \quad \text{as } t \rightarrow \infty \quad \forall i \in \mathcal{N}$$

with P^* -probability one. Consequently, $z_{i,t}(\theta)$ converges to zero P^* -almost surely for all i such that $\theta \in \Theta \setminus \Theta_i^*$. Since $\theta \in \Theta \setminus \Theta^*$, there exists at least one such agent; call it i . Since $z_{i,t}(\theta)$ converges to zero and $z_{j,t}(\theta)$ is non-negative for all $j \in \mathcal{N}_i$, (7) implies that $z_{j,t}(\theta)$ also converges to zero

⁴This is a consequence of the Perron-Frobenius theorem. For more on this see, for instance, Chapter 2 of [15].

P^* -almost surely for all $j \in \mathcal{N}_i$. Note that this happens even if $\theta \in \Theta_i^*$. A similar argument implies that $z_{k,t}(\theta)$ converges to zero for all $k \in \mathcal{N}_j$ and $j \in \mathcal{N}_i$. Proceeding inductively and using the strong connectivity assumption implies that $z_{j,t}(\theta)$ converges to zero P^* -almost surely for all $j \in \mathcal{N}$.

We can now prove the result for the case that z_0 is not necessarily non-negative. Note that z_0 can be decomposed as $z_0 = z_0^+ - z_0^-$, where z_0^+ and z_0^- are non-negative. Since $\Phi(\omega; t)$ is a linear map, z_t is given by

$$z_t = z_t^+ - z_t^- = \Phi(\omega; t)z_0^+ - \Phi(\omega; t)z_0^-$$

which by what we just proved goes to zero on a set of P^* -probability one as $t \rightarrow \infty$. ■

The next lemma shows that $f_t(\omega; x) = \varphi_t(\omega; x) - M_t(\omega)x$, as defined in (6), is negligible—uniformly in t and ω —when x is small.

For any $s \in S$ and $\theta \in \Theta$, let $G_\theta(s) : \mathbb{R}^{|\Theta \setminus \Theta^*|} \rightarrow \mathbb{R}^{n \times n}$ be the diagonal matrix-valued function whose i th diagonal element is given by

$$G_{i,\theta}(s; x) = a_{ii} \left(\frac{\ell_i(s_i|\theta)}{g_i(s_i; x)} - \frac{\ell_i(s_i|\theta)}{\ell_i(s_i|\theta^*)} \right) \quad (8)$$

where $g_i(s_i; x_i)$ is defined as

$$g_i(s_i; x_i) = \ell_i(s_i|\theta^*) \left(1 - \sum_{\theta \in \Theta \setminus \Theta^*} x_i(\theta) + \sum_{\theta \in \Theta \setminus \Theta^*} \ell_i(s_i|\theta) x_i(\theta) \right)$$

It is straightforward to show that $f_{t,\theta}(\omega; x) = G_\theta(\omega; x)x(\theta)$ for all $\theta \in \Theta \setminus \Theta^*$, where $f_{t,\theta}(\omega; x)$ is the θ component of $f_t(\omega; x)$.

Lemma 2: There exist a neighborhood V of the origin and a constant $K_\theta > 0$ such that if $x, y \in V$, then

$$\|G_\theta(s; x) - G_\theta(s; y)\| \leq K_\theta \|x - y\| \quad \forall s \in S.$$

Proof: Some elementary algebra results in

$$G_{i,\theta}(s; x) - G_{i,\theta}(s; y) = \left[\frac{a_{ii} \ell_i(s_i|\theta)}{g_i(s_i; x) g_i(s_i; y)} \right] \times (g_i(s_i; y) - g_i(s_i; x)).$$

Note that, for all $s_i \in S_i$, $g_i(s_i; x_i)$ is a continuous function of x_i which is equal to $\ell_i(s_i|\theta^*)$ when $x_i = 0$. Therefore, for any sufficiently small $\epsilon > 0$, there exists a neighborhood $U_i(s_i)$ of the origin such that if $x_i, y_i \in U_i(s_i)$, then

$$\left| \frac{a_{ii} \ell_i(s_i|\theta)}{g_i(s_i; x) g_i(s_i; y)} \right| \leq \frac{a_{ii} \ell_i(s_i|\theta)}{[\ell_i(s_i|\theta^*) - \epsilon]^2}.$$

Choose ϵ such that $\epsilon < \ell_i(s_i|\theta^*)$ for all $s_i \in S_i$. Let V_i be the intersection, over $s_i \in S_i$, of the corresponding $U_i(s_i)$. If $x_i, y_i \in V_i$, then

$$\left| \frac{a_{ii} \ell_i(s_i|\theta)}{g_i(s_i; x) g_i(s_i; y)} \right| \leq \max_{s_i \in S_i} \frac{a_{ii} \ell_i(s_i|\theta)}{[\ell_i(s_i|\theta^*) - \epsilon]^2} \quad \forall s_i \in S_i. \quad (9)$$

Define $k_{i,\theta} > 0$ as the right hand side of the above inequality. On the other hand,

$$g_i(s_i; y_i) - g_i(s_i; x_i) = \sum_{\theta \in \Theta \setminus \Theta^*} (\ell_i(s_i|\theta^*) - \ell_i(s_i|\theta)) \times (x_i(\theta) - y_i(\theta)).$$

Therefore,

$$\begin{aligned} |g_i(s_i; y_i) - g_i(s_i; x_i)| \\ &\leq \sum_{\theta \in \Theta \setminus \Theta^*} |(\ell_i(s_i|\theta^*) - \ell_i(s_i|\theta))| |x_i(\theta) - y_i(\theta)| \\ &\leq |\Theta \setminus \Theta^*| \left[\max_{\theta \in \Theta \setminus \Theta^*} |\ell_i(s_i|\theta^*) - \ell_i(s_i|\theta)| \right. \\ &\quad \left. \times \max_{\theta \in \Theta \setminus \Theta^*} |x_i(\theta) - y_i(\theta)| \right] \\ &\leq k'_i \|x_i - y_i\|_\infty \end{aligned} \quad (10)$$

where k'_i is the non-negative constant defined as

$$k'_i = |\Theta \setminus \Theta^*| \max_{s_i \in S_i, \theta \in \Theta \setminus \Theta^*} |\ell_i(s_i|\theta^*) - \ell_i(s_i|\theta)|.$$

Note that $k'_i > 0$ if $\Theta_i^* \neq \Theta$. Combining (9) and (10), we can conclude that if $x_i, y_i \in V_i$, then

$$|G_{i,\theta}(s; x) - G_{i,\theta}(s; y)| \leq k_{i,\theta} k'_i \|x_i - y_i\|_\infty.$$

Therefore, if $x, y \in V_1 \times \dots \times V_n$ for all $i \in \mathcal{N}$, then

$$\begin{aligned} \|G_\theta(s; x) - G_\theta(s; y)\| &= \max_{i \in \mathcal{N}} |G_{i,\theta}(s; x) - G_{i,\theta}(s; y)| \\ &\leq \max_{i \in \mathcal{N}} [k_{i,\theta} k'_i \|x_i - y_i\|_\infty] \\ &\leq \left[\max_{i \in \mathcal{N}} k_{i,\theta} k'_i \right] \left[\max_{i \in \mathcal{N}} \|x_i - y_i\|_\infty \right] \\ &= \left[\max_{i \in \mathcal{N}} k_{i,\theta} k'_i \right] \|x - y\|_\infty \\ &\leq K_\theta \|x - y\| \end{aligned}$$

for some constant $K_\theta > 0$, where the last inequality is due to the equivalence of the norms. Letting $V = V_1 \times \dots \times V_n$ completes the proof. ■

We are now ready to prove Proposition 3. The basic idea of the proof is similar to the indirect method of Lyapunov: the linear part of the dynamical system $\varphi(\omega)$ is asymptotically (and hence exponentially) stable (by Lemma 1); furthermore, its non-linear part is uniformly sufficiently small (by Lemma 2). Therefore, $\varphi(\omega)$ inherits the exponential stability of its linear part in a neighborhood of the origin.

We cannot use the classic result of Lyapunov to prove the proposition, however, since it requires the linearized dynamical system to be uniformly asymptotically stable; whereas Lemma 1, which makes use of the martingale convergence theorem, only proves its asymptotic stability. Recent results by Barreira and Valls on the other hand show that exponential stability is preserved under perturbations, given conditions that are much more general than uniform asymptotic stability [16]. We use one such result (Theorem 1 in Appendix C) to

prove the proposition. The theorem relies on a condition known as Lyapunov regularity which we verify using the celebrated multiplicative ergodic theorem of Oseledets (Theorem 2 in Appendix D).

Proof of Proposition 3: We prove the proposition by using Theorem 1 in Appendix C to show that the dynamical system $\varphi(\omega)$ is exponentially stable for P^* -almost all ω . We first verify that the premises of Theorem 1 are satisfied. Let $\Phi(\omega; t)$ be the evolution operator corresponding to $M(\omega)$. Oseledets' Theorem (Theorem 2 in Appendix D) implies that $\Phi(\omega; t)$ is Lyapunov regular for P^* -almost all $\omega \in \Omega$.⁵ Lemma 1 implies that $\lambda_1(\omega)$, the top Lyapunov exponent of $\Phi(\omega; t)$, is negative on a set of P^* -probability one. Next, we show that f satisfies the condition in (27) with $\gamma = 1$. By the sub-multiplicativity of the norm and the triangle inequality,

$$\begin{aligned} \|f_{t,\theta}(\omega; x) - f_{t,\theta}(\omega; y)\| \\ &= \|G_\theta(\omega; x)x(\theta) - G_\theta(\omega; y)y(\theta)\| \\ &= \|G_\theta(\omega; y)(x(\theta) - y(\theta)) + (G_\theta(\omega; x) - G_\theta(\omega; y))x(\theta)\| \\ &\leq \|G_\theta(\omega; y)\| \|x(\theta) - y(\theta)\| \\ &\quad + \|G_\theta(\omega; x) - G_\theta(\omega; y)\| \|x(\theta)\| \\ &= \|G_\theta(\omega; y) - G_\theta(\omega; 0)\| \|x(\theta) - y(\theta)\| \\ &\quad + \|G_\theta(\omega; x) - G_\theta(\omega; y)\| \|x(\theta)\| \end{aligned}$$

where the last equality is since $G_\theta(\omega; 0) \equiv 0$. By Lemma 2 and the above inequality,

$$\|f_{t,\theta}(\omega; x) - f_{t,\theta}(\omega; y)\| \leq K_\theta \|x(\theta) - y(\theta)\| (\|x(\theta)\| + \|y(\theta)\|).$$

Therefore,

$$\|f_t(\omega; x) - f_t(\omega; y)\| \leq K \|x - y\| (\|x\| + \|y\|)$$

for some $K > 0$.⁶

We have shown that the dynamical system $\varphi(\omega)$ satisfies the assumptions of Theorem 1 for ω in a set of P^* -probability one. Therefore, we can conclude that, for sufficiently small $\epsilon > 0$ and P^* -almost all $\omega \in \Omega$, there exists a neighborhood $U(\omega)$ of the origin and $C(\omega) > 0$ such that for all $\tilde{\mu}_{t_0} \in U(\omega)$ and $t \geq t_0$,

$$\|\tilde{\mu}_t\| \leq C(\omega) e^{(\lambda_1(\omega) + \epsilon)(t - t_0) + 2\epsilon t_0} \|\tilde{\mu}_{t_0}\| \quad (11)$$

where $\lambda_1(\omega) < 0$ is the top Lyapunov exponent of $M(\omega)$. On the other hand, Proposition 1 implies that for P^* -almost all ω and any neighborhood V of the origin, there exists $T(\omega)$ such that $\tilde{\mu}_{t_0} \in V$ for all $t_0 \geq T(\omega)$. Therefore, taking \limsup 's of logarithms of both sides of (11) and noting that ϵ is arbitrary implies that

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \log \|\tilde{\mu}_t\| \leq \lambda_1(\omega)$$

with P^* -probability one.

⁵Strictly speaking we also need $M_t(\omega)$ to be invertible for all $\omega \in \Omega$ and $t \in \mathbb{N}$; however, this requirement is satisfied generically.

⁶Note that even though the condition is proved only in a neighborhood U of the origin, this does not pose a problem as there exists $T(\omega)$ such that $x_i \in U$ for all $t \geq T(\omega)$ with P^* -probability one.

Claro, son Transactions y Journals del IEEE : ¡La ingeniería como debe ser!...

¿y el estudio de los sistemas complejos como debe ser?

Algunas aplicaciones de DSP en adaptación y aprendizaje sobre redes complejas:

- Modelos de difusión de creencias en redes sociales, Chamley y Scaglione
- Aprendizaje en redes mediante juegos de información incompleta, Eksin, Molabi y Rivero.
- Aprendizaje social y juegos bayesianos en procesamiento de señales multiagente, Krishnamurthy y Poor.
- Procesamiento de señales en grafos, Shuman, Narang, Ortega, Frossard
- Inferencia distribuida sobre redes, Kar and Moura
- Etc.

Algunas aplicaciones de DSP en adaptación y aprendizaje sobre redes complejas:

- **Modelos de difusión de creencias en redes sociales, Chamley y Scaglione**
- Aprendizaje en redes mediante juegos de información incompleta, Eksin, Molabi y Rivero.
- Aprendizaje social y juegos bayesianos en procesamiento de señales multiagente, Krishnamurthy y Poor.
- Procesamiento de señales en grafos, Shuman, Narang, Ortega, Frossard
- Inferencia distribuida sobre redes, Kar and Moura
- Etc.

Veamos un ejemplo muy por encima:

Modelos para la difusión de creencias en redes sociales



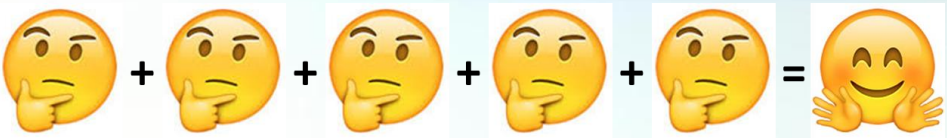
¿Cuánto dinero hay en esta jarra?
(hay cerca de tres monedas de \$500
por cada moneda de \$1000)

Marquis de Condorcet (1785)
abogando por la democracia y el voto
popular.

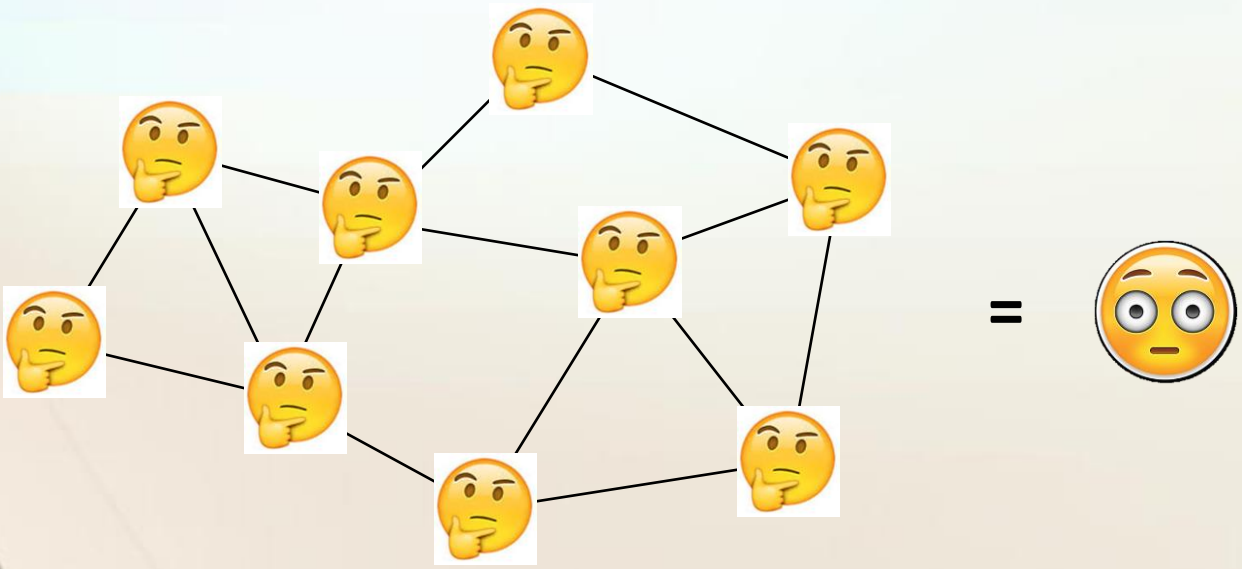
Francis Galton (1905) “Vox Populi”

Jason Silva (2015), juegos mentales de
NatGeo

La opinión de cada individuo está sujeta a errores, de manera que al promediar muchas opiniones, por la ley de los grandes números, se obtiene un estimador más preciso



Pero ¿qué tal la perturbación inducida por conocer la opinión de otros individuos?



La diferencia es clave para los mecanismos de aprendizaje social

- La cantidad de dinero es una v.a. $x \in X$
- Hay n agentes
- Cada agente tiene un modelo mental de acuerdo con su experiencia y su conocimiento, que se representa mediante las señales $\{s_i, i=1,2,\dots,n\}$
- La opinión expresada por un agente es una acción sobre la sociedad que lleva una señal que alimenta el proceso de aprendizaje social.
- La acción puede ser su propia señal, pero no necesariamente es así pues puede haber otros intereses o fallas de comunicación.

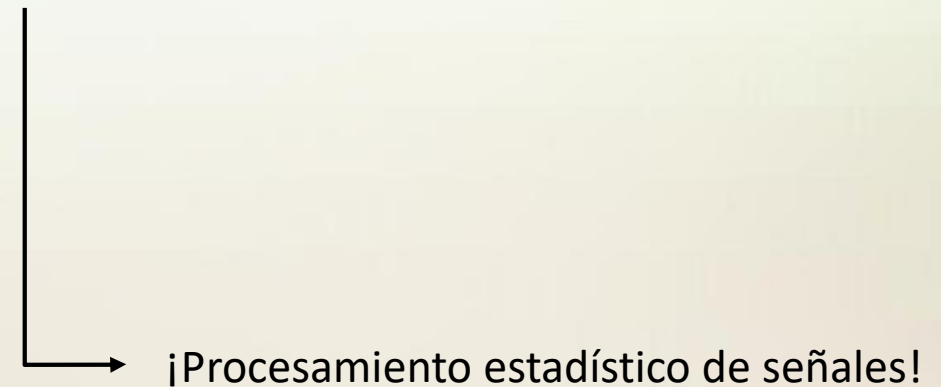
$x \sim \mathcal{N}(\theta_0, \sigma_0^2)$ Distribución a priori del valor real

$s_i = \mathcal{N}(x, \sigma_i^2)$ Distribución de cada señal individual

$a_i = E_i[x | \text{info}_i]$ Las acciones de los agentes son sus señales (¡agentes sinceros! 🤗)

Modelo secuencial:

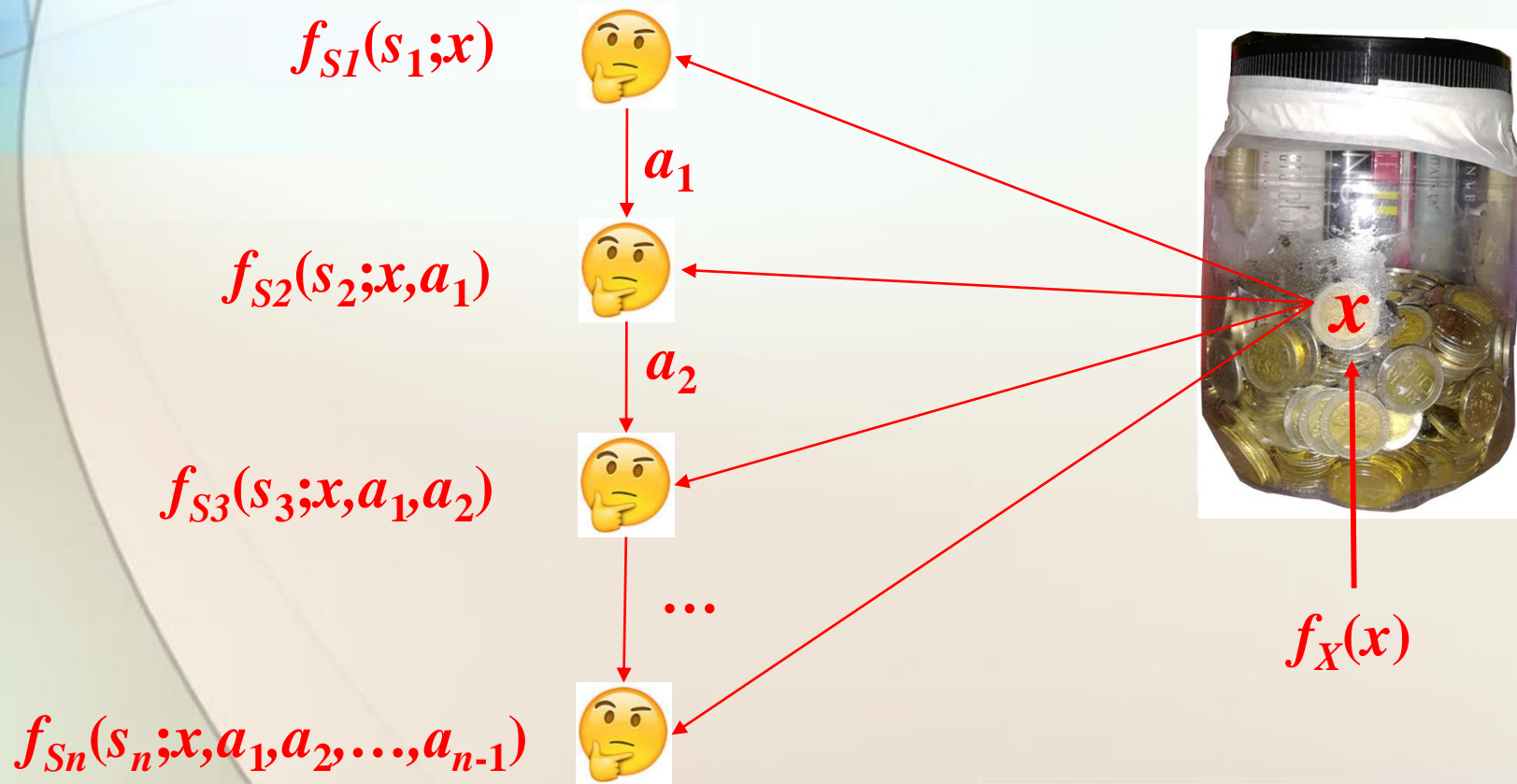
- Cada agente dispone de su información privada y la historia de las acciones de los agentes que lo precedieron
- Los agentes son racionales porque usan la regla de Bayes para combinar su información privada con la información pública y escogen la acción que maximiza (según sus creencias) alguna función de ganancia que depende del verdadero valor de x y de su propia acción.



$x \sim \mathcal{N}(\theta_0, \sigma_0^2)$ Distribución a priori del valor real

$s_i = \mathcal{N}(x, \sigma_i^2)$ Distribución de cada señal individual

$a_i = E_i[x | \text{info}_i]$ Las acciones de los agentes son sus señales



$$\gamma_i = \frac{\sigma_0^2}{\sigma_i^2} \quad \text{Relación señal a ruido}$$

Estimador de mínimo error cuadrado promedio

$$E[x | s_1, \dots, s_n] = \frac{\sum_{i=1}^n \gamma_i s_i}{\sum_{k=0}^n \gamma_k}$$

$$\gamma_i = \frac{\sigma_0^2}{\sigma_i^2} \quad \text{Relación señal a ruido}$$

Estimador de mínimo error cuadrado promedio

$$E[x | s_1, \dots, s_n] = \frac{\sum_{i=1}^n \gamma_i s_i}{\sum_{k=0}^n \gamma_k}$$

La varianza del error disminuye de acuerdo con el factor

$$\rho_n = \frac{1}{\sum_{k=0}^n \gamma_k} = \frac{\rho_{n-1}}{1 + \gamma_n \rho_{n-1}}$$

El simple promedio ponderado permite calcular un estimado de $E[x | s_1, \dots, s_n]$ con una varianza del error $\text{MMSE}_n = \rho_n \sigma_0^2 < \sigma_0^2$

La precisión del estimador aumenta con la población : El estimado se actualiza recursivamente pues cada iteración es un mensaje que un agente social envía al resto de la población.

Si $\tau_n = \rho_n / \rho_{n-1}$, el mejor estimado (y la acción óptima) del n -ésimo agente es la combinación convexa

$$\theta_n^* = (1 - \tau_n) s_n + \tau_n \theta_{n-1}^*$$

Si $\tau_n = \rho_n / \rho_{n-1}$, el mejor estimado (y la acción óptima) del n -ésimo agente es la combinación convexa

$$\theta_n^* = (1 - \tau_n) s_n + \tau_n \theta_{n-1}^*$$

Después de la n -ésima adivinación, el modelo de creencia de la población, $f_n(x|a_1, \dots, a_{n-1}, s_n)$ se actualiza así:

$$x \sim \mathcal{N}(\theta_n^*, \rho_n \sigma_0^2)$$

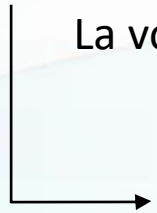
Conclusión interesante:

τ_n , la mejora en el MMSE de cada actualización, que tiende a 1, se puede interpretar como la confianza en el estimado histórico agregado en comparación con la información privada añadida por cada individuo... Así de poderoso puede ser el aprendizaje social.

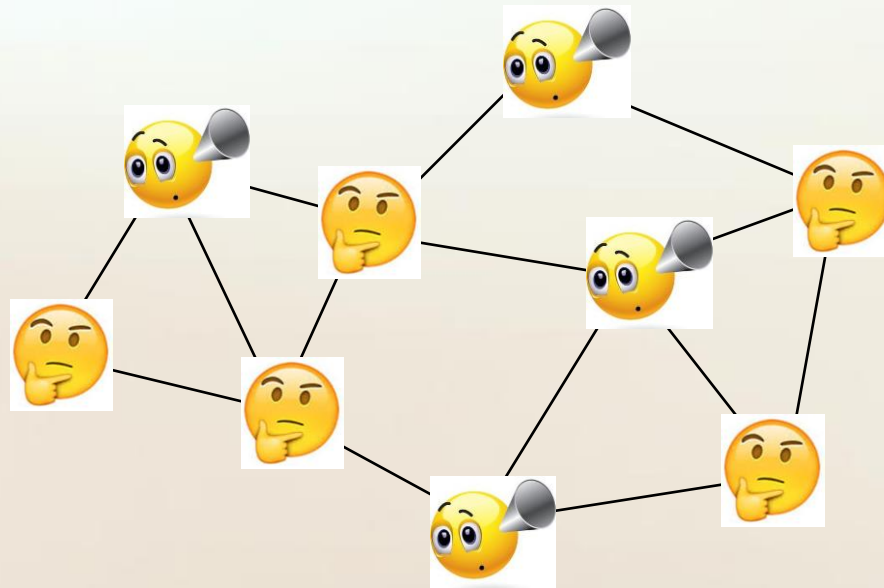
Si la capacidad de comunicación no es infinita, la teoría de información indica que la reducción en el error es

$$\rho_n = \frac{\rho_{n-1}}{1 + \gamma_n \rho_{n-1}} + \frac{\sigma_0^2}{1 + \gamma_n} \cdot \frac{1 - \rho_{n-1}}{\gamma_n + \rho_{n-1}} \cdot 2^{-R_{n-1}}$$

De manera que el error ya no se reducirá como $1/n$ y se limita la máxima exactitud alcanzable :
La voz del pueblo ya no sería la voz de Dios



Aún en una sociedad de gente honesta, la falta de infraestructura de comunicación puede generar dificultades democráticas.



¿Modelo secuencial?

- Los acuerdos se logran promediando opiniones entre grupos de vecinos
- Los agentes actúan múltiples veces en eventos aleatorios y asincrónicos
- La estructura de la red se ajusta con el tiempo y no sólo intercambia información sino también premios y costos
- No se trata de hacer una estimación correcta sino en ponerse de acuerdo
- Las interacciones se basan en intereses particulares
- ...
- Los modelos desde el procesamiento de señales sobre grafos podrían llegar a revelar algunos aspectos útiles sobre la capacidad de una sociedad para mezclar su información y sus evidencias de una manera exitosa, de manera que llegue a comportarse como una inteligencia colectiva racional, superior a la suma de las inteligencias individuales.

¡Gracias!