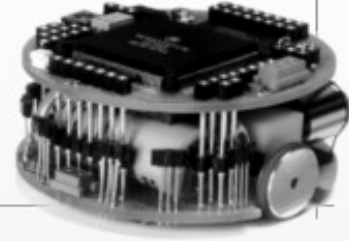


# VIDA MENTAL

## la autonomía de la conducta



*The question as to the nature of life, I believe, has been finally resolved, and is no longer a philosophical question. I hope something like this will happen to the so-called mind-body problem in the twenty-first century.*

JOHN SEARLE

*Now what makes the cell living? The soft organization of its inner events and occurrences. Thus, if we are looking for the fundamental laws, for the principle of life, we have to establish the connections of this soft organization.*

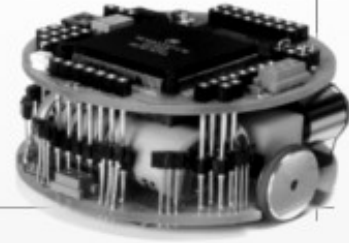
TIBOR GÁNTI

*Cognition presupposes the function of an organization for its own conservation and this is the first fundamental analogy with life.*

JEAN PIAGET



Xabier BARANDIARAN, 6 de Junio, UPM

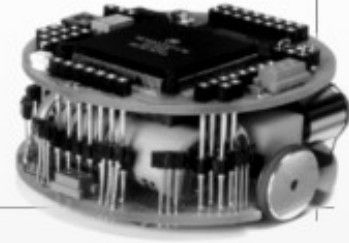


# Vida Mental

## la autonomía de la conducta



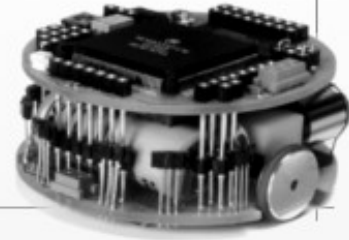
# Estructura



- El problema de la autonomía: identidad y normatividad
- Vida orgánica mínima: el metabolismo celular
- Vida Mental: La autonomía de la conducta
  - Hábitos
  - Organización de hábitos
- Robótica evolutiva
  - Métodos
  - Principio
  - Readaptación a la inversión visual



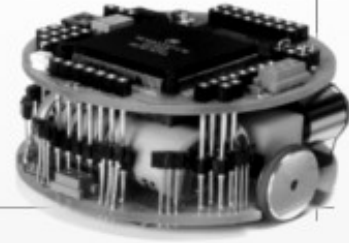
# El Problema de la Autonomía



- El problema de la autonomía es un problema que va mucho más allá de la *independencia* de un controlador remoto o de la capacidad de moverse sin grandes constricciones.
- *Auto-nomos*: dotarse uno mismo de sus propias normas
- *Auto*: identidad
  - ¿Cómo distinguir un sistema de su entorno?
  - ¿Cómo surge un sujeto de acciones?
- *Nomos*: normatividad
  - Dicotomía: Bueno/malo, adaptativo/no-adaptativo, verdadero/falso, correcto/equivocado... ¿Cómo justificar la naturaleza de esta dicotomía y de la norma que subyace?



# ¿A quién le importa?



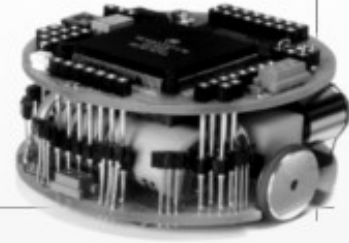
**INGENIERO:** Eso de la autonomía, la identidad y la normatividad... está muy bien como problema filosófico...pero **a mi robot no le importa!**

**FILÓSOFO:** Ese es precisamente el problema... que a tu robot no le importa! Si tu robot *falla* .... no le importa! Pero a los organismos vivos (incluyendo los becarios de robótica) sí nos importa. Somos capaces de definir por nosotros mismos qué es bueno o malo y nos preocupa fallar.

¿Cómo es eso posible? ¿Puede haber cognición genuina, significado, sin que te importe, sin posibilidad misma de definir qué es lo correcto o apropiado?



# ¿A quién le importa?



## → Roseblueth, Wiener & Bigelow (1943):

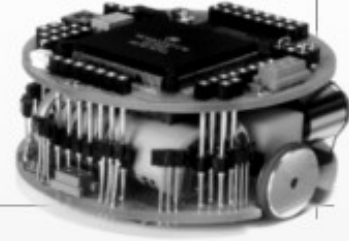
- Conducta teleológica es aquella que tiende a un estado relacional con el entorno a través de un mecanismo de corrección de error de feed-back negativo (ej.: servomecanismo de seguimiento visual (ametralladora antiaérea))
- Oscilaciones crecientes por error sobre-correcto... paralelismo con un paciente neurológico

## → Hans Jonas (1966):

- NO! al servomecanismo le da igual sobre-equivocarse, no se estresa, no le atañe. Al paciente SI!
- La teleología no puede ser definida desde fuera. Hay algo interno a la organización de lo viviente que hace que esté comprometido, preocupado por su propia existencia.



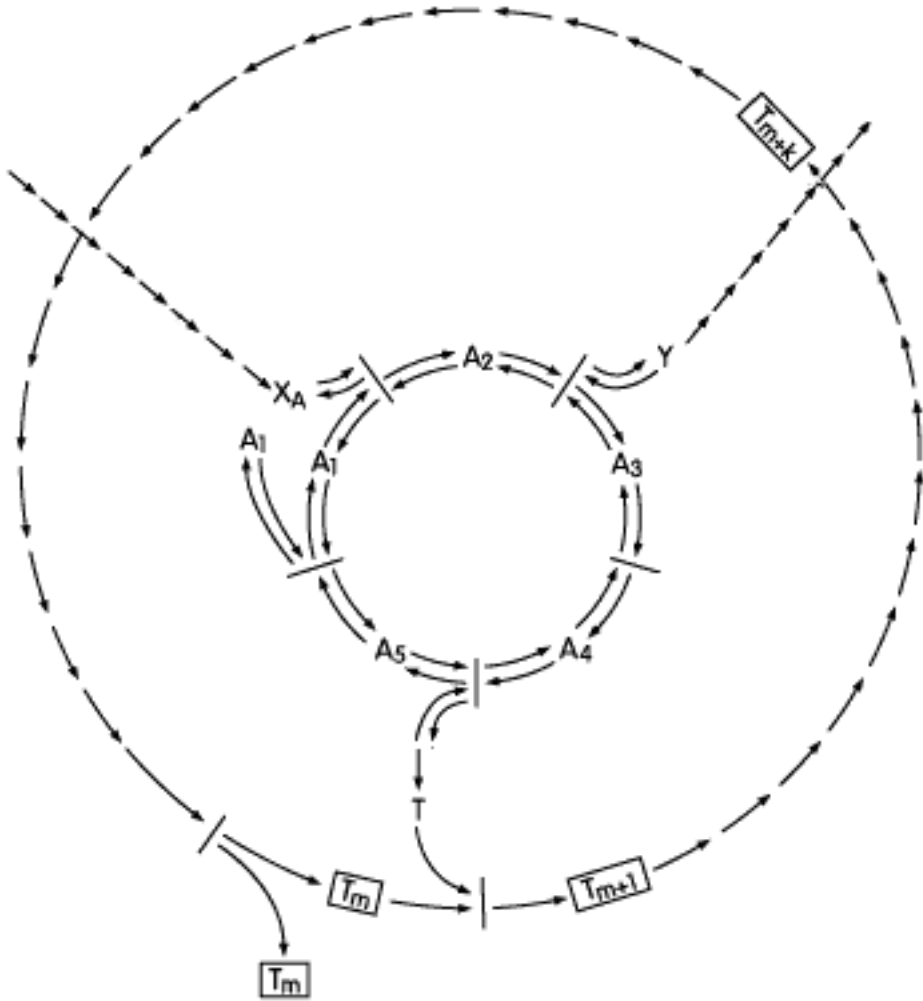
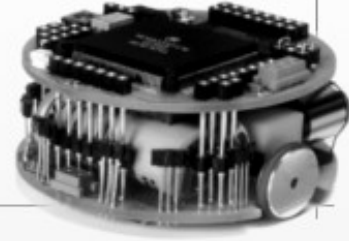
# La Vida como Autonomía Metabólica



- Los seres vivos son capaces de construirse y repararse a sí mismos constantemente, en condiciones alejadas del equilibrio y definiendo sus propias fronteras de forma activa.
- Ser (estructura) y Hacer (función) son interdependientes
- ¿Y en las máquinas? NO, las estructuras de construyen fuera del contexto organizativo robótico, se ensamblan y luego se ponen a funcionar.
- La vida como forma de organización circular y automantenida: Robert Rosen, Hans Jonas, Maturana y Varela, Ruiz-Mirazo & Moreno, Tibor Ganti, ...



# La Vida como Autonomía Metabólica

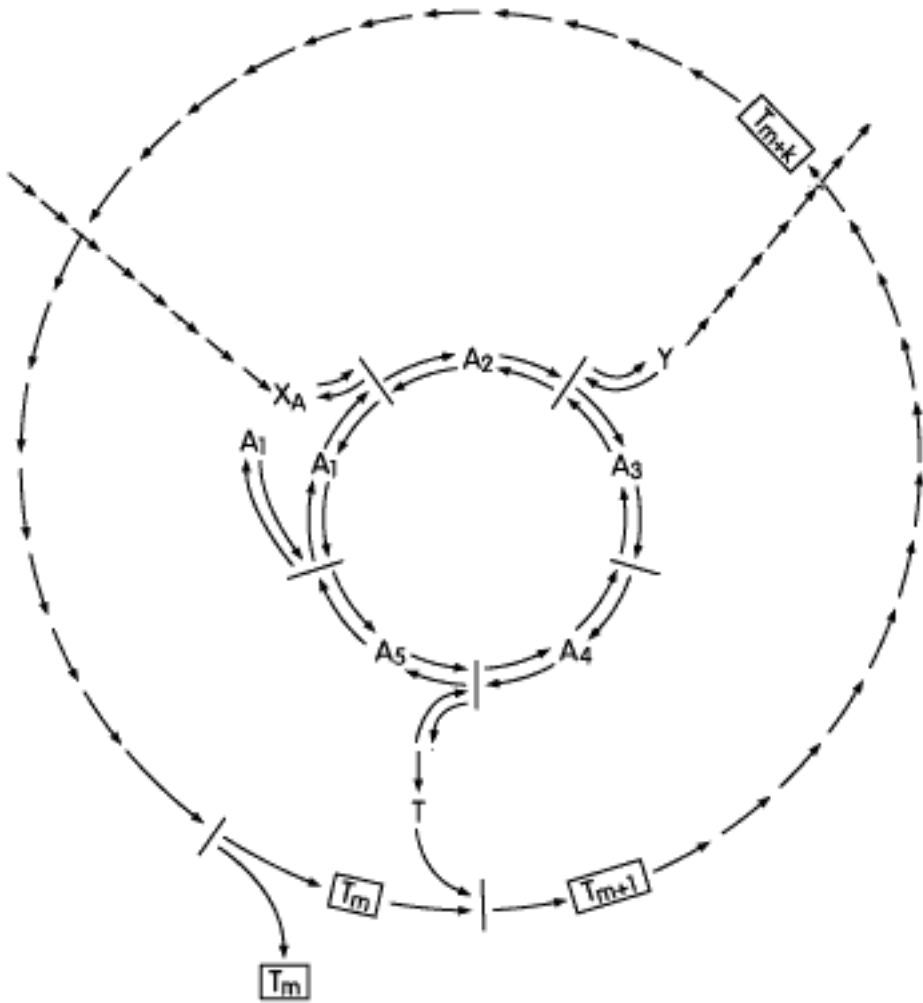
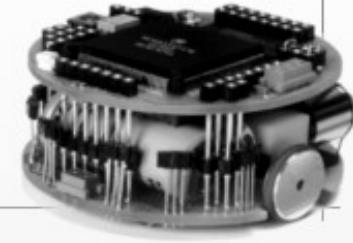


- Vida Mínima protocélular = metabolismo + membrana activa
- *Metabolismo*: Una red de reacciones químicas que produce los componentes de la propia red (en concreto los catalizadores)
  - *Membrana activa*: se produce una membrana que separa la red del entorno y gestiona el flujo termodinámico para su propio automantenimiento.





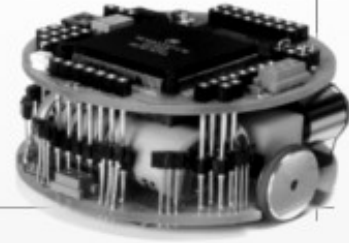
# La Vida como Autonomía Metabólica



- **Identidad:** cohesión automantenida
- **Normatividad:** procesos se requieren unos a otros, condiciones objetivas de automantenimiento, un proceso *tiene* que ocurrir en cierto modo o el sistema desaparece



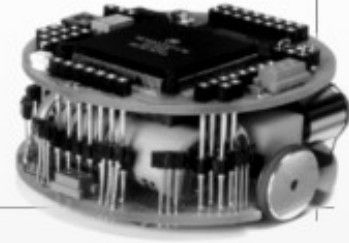
# Agencialidad y Conducta Adaptativa



- La evolución constriñe/diseña los mecanismos de la conducta para que el sistema interactúe con su entorno para sobrevivir
- La normatividad y la identidad son externas al dominio de la conducta.
  - El sistema es metabólicamente autónomo y define su propia norma
  - Pero está determinado por factores externos a la propia conducta
- En casos más complejos que la conducta puramente reactiva, p.e. aprendizaje reforzado, los objetivos están definidos desde fuera (placer/dolor).

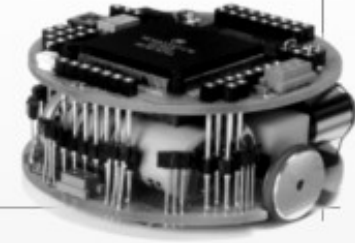


# Vida Mental: Autonomía de la Conducta



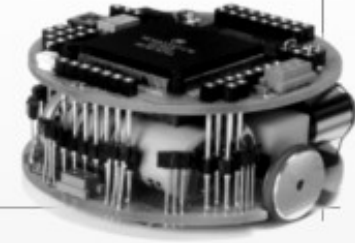
- El significado cognitivo no es “derivado” del significado biológico:
  - Un baso de agua no es significativo porque nuestro cuerpo se encuentre bajo un déficit de H<sub>2</sub>O, la sed se puede estimular artificialmente y el baso de agua nos resulta igualmente significativo
- **La mente tiene vida propia**
- Espacio de sub-determinación bajo régimen de subordinación
  - Ejemplo académico
- La vida orgánica hace posible un dominio (el de la conducta sostenida por el SN) que adquiere progresivamente mayores grados de libertad





- Tim Smithers (1997): “Designing and building autonomous agents thus becomes the problem of designing and building processes that can support and maintain this kind of identity formation through interaction: *processes that, through interaction, are continuously forming the laws of interaction that can sustain and maintain the interaction needed to form them.* In other words, we need interaction processes that can support the self-construction and maintenance of interaction processes through interaction, in essentially the same way that the material and energy interaction processes of single cells can be understood as being involved in the continual forming of the mechanisms that support this interaction. Such systems will thus be self-law making as well as self-regulating, in essentially the same way as we can understand biological systems and autonomous city states”
- Pero esto es una definición circular!!!
- El problema no es cómo salir de una definición circular... sino cómo entrar en ella de forma efectiva (Heidegger).

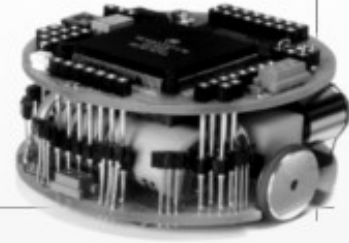




- ¿Cuales podrían ser los componentes de una organización cognitiva para poder hacer una analogía entre la vida orgánica y la vida mental?



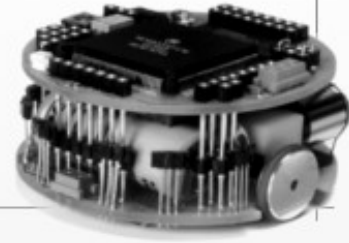
# Hábitos



- Aristotle, Lamarck, Bichat, James, etc.
- “Animals are bundles of habits” William James 1890
- Definición provisional: un hábito es un patrón automantenido de conducta.
- Automantenido: La estabilidad de la conducta esta acoplada a la estabilidad de los mecanismos que la generan. El hábito “llama” a realizar una conducta, la conducta refuerza el hábito.
- Estructuras sensomotoras plásticas: Piaget/Arbib (schema), Merleau-Ponty (forma dinámica), en neurociencia (atractores neurodinámicos)



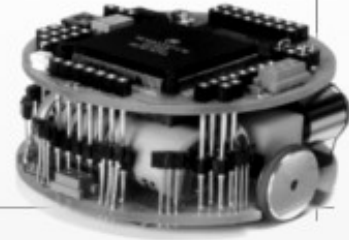
# Hábitos



- Una versión plástica de la arquitectura de subsunción de Brooks: la estabilidad del circuito depende de la estabilidad de la conducta que genera.
- Podemos entrar de forma efectiva en la circularidad de Smithers: “processes [habits] that, through interaction [through their exercise], are continuously forming [reinforcing] the laws [the mechanisms] of interaction that can sustain and maintain the interaction needed to form them”.



# Propiedades de los hábitos

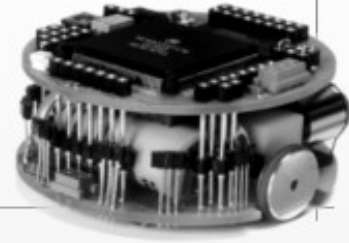


- a) el funcionamiento de un hábito se puede modelar en un marco completamente operacional/dinámico,
- b) los hábitos no presuponen una distinción o prioridad causal entre percepción y acción, al tiempo que integran a las dos
- c) los hábitos son estructuras inherentemente situadas o enactivas en el sentido de que atraviesan cerebro, cuerpo y entorno
- d) los hábitos son plásticos y maleables (a diferencia de las connotaciones de rigidez que acompañan a las nociones de arco reflejo o instinto)
- e) los hábitos proveen un sentido concreto de automantenimiento en el nivel de la conducta (son causa y efecto de su ocurrencia, formas de acción auto-reforzadas)
- f) los hábitos pueden componerse o anidarse a diferentes escalas





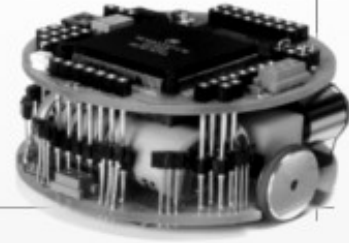
# Una red de hábitos organizada y adaptativa



- Sin embargo un hábito aislado parece llevarnos al extremo opuesto al que queríamos llegar: son como una posesión de la identidad (y la voluntad) del sujeto cognitivo.
- Pero qué pasaría si concebimos un *conjunto organizado de hábitos*? Diferentes hábitos podrían requerirse mutuamente, creando relaciones de dependencia de estabilidad entre hábitos y entre ellos y las conductas que generan. (ej.: coffee & cigarettes)
- Y una red adaptativa y automantenida de hábitos? El conjunto de hábitos interactúa continuamente para mantenerse a sí mismo como un todo, sacrificando hábitos particulares aquí y allí para preservar el conjunto.



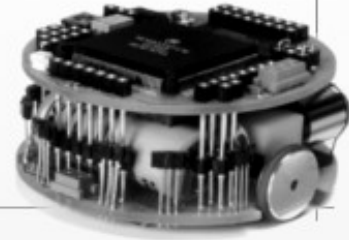
# Vida orgánica – Vida mental



- Si la vida orgánica se definía como una red automantenida de reacciones químicas alejadas del equilibrio....
- La vida mental puede definirse como una red automantenida de hábitos o estructuras neurodinámicas
- La VIDA MENTAL aparece cuando la conservación adaptativa de la organización neurodinámica de la conducta deviene el objetivo último de la regulación sensomotora.



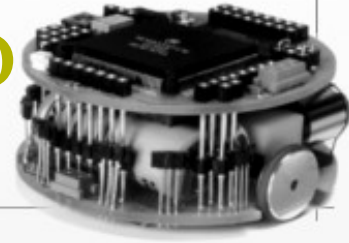
# Autonomía de la Conducta: Identidad y Normatividad



- ➔ La mente es fuente de una nueva identidad, que sobreviene a la biológica pero no es reducible a ésta.
  - Podemos estar completamente vivos (orgánicamente) y estar en coma.
  - Incluso doble personalidad (o hemisferios separados) y un sólo cuerpo
- ➔ La conducta se convierte en autónoma: se dota a sí misma de sus normas, aquellas necesarias para mantener su organización.
  - Coherencia interna como fuente de normatividad. La propia organización de la conducta, sus condiciones de estabilidad global.
  - Principios de integración y acomodación de Piaget

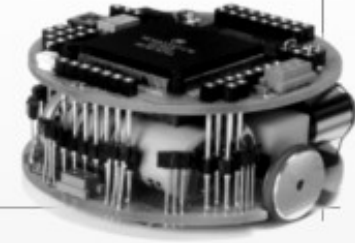


# Intencionalidad y significado



- Una conducta es intencional cuando se desencadena para recuperar una estabilidad amenazada, para compensar la desviación de una norma.
- Algo es significativo cuando se integra en el proceso de estabilización de una organización: principio de integración y acomodación (Piaget 1969)
- No puede haber significado si el proceso de definición normativa depende de un observador/diseñador externo

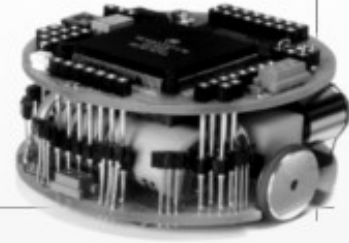




**INGENIERO:** Muy bien, muy bonito (o no!)... pero volviendo a mi departamento... ¿yo qué hago con esto? ¿cómo construyo un robot que tenga hábitos?



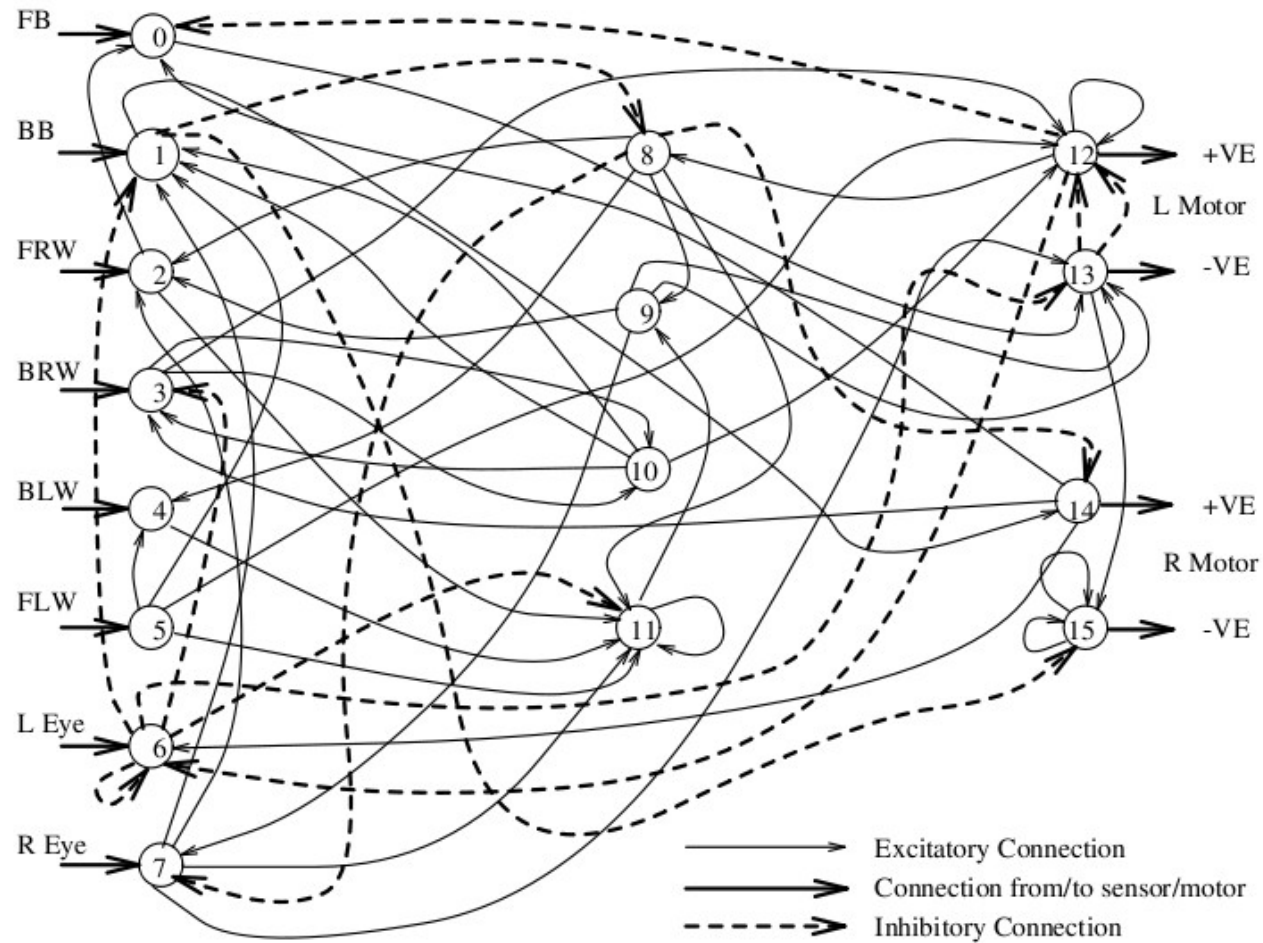
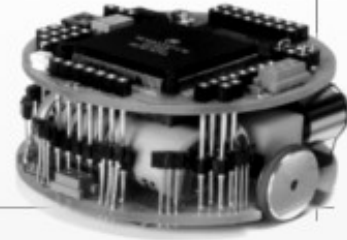
# Robótica Evolutiva I



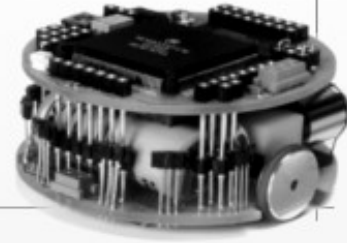
## Métodos



# CTRNN



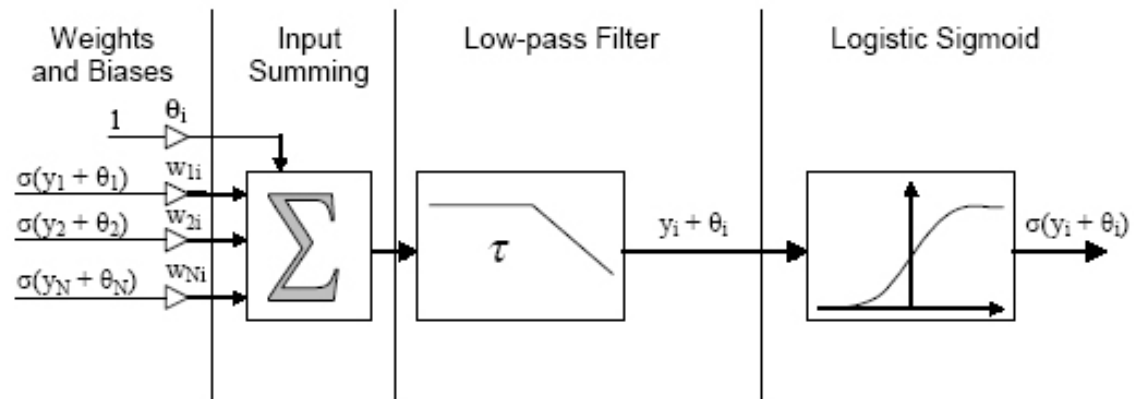
# CTRNN



- Continuous Time Recurrent Neural Network
- Tiempo continuo
- Totalmente conectadas (muchos bucles retroalimentados)
- Pueden aproximar cualquier sistema dinámico (Funahashi & Nakamura 1996)

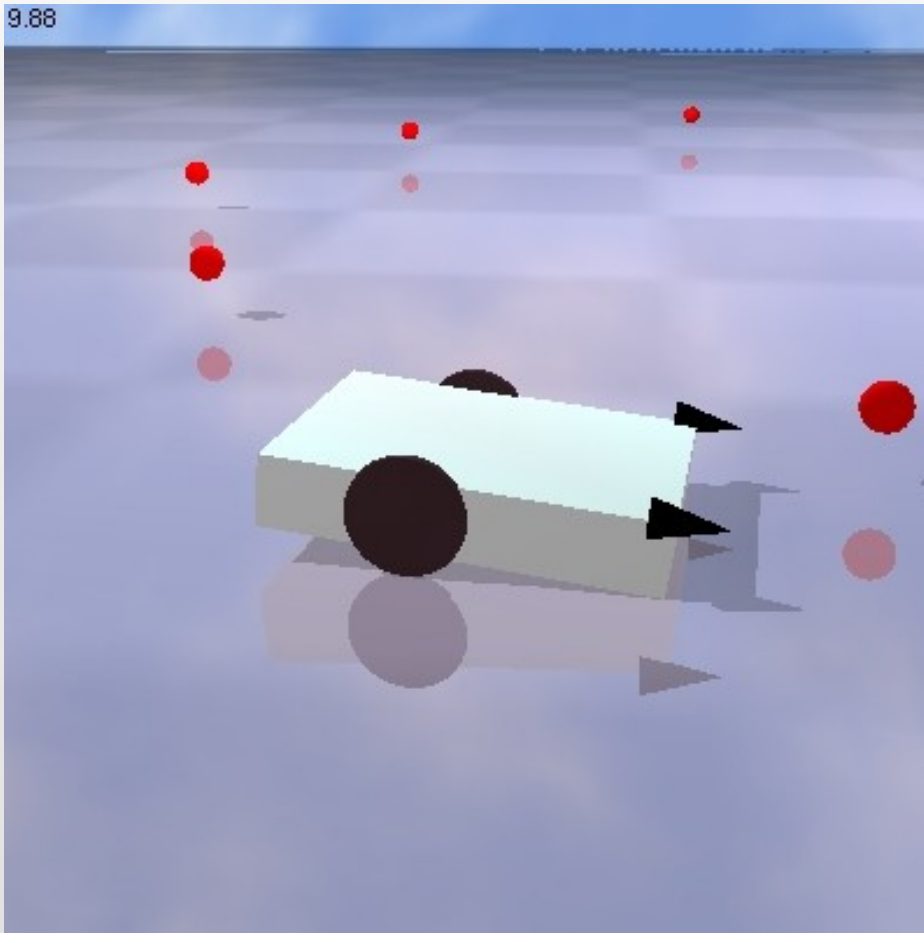
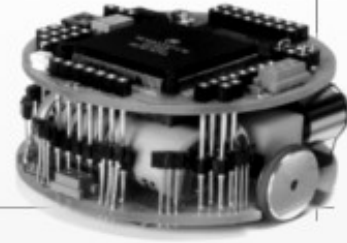
$$\tau_i \dot{y}_i = -y_i + \sum_{j=1}^n (w_{ij} z_j) + g_i \sum_{k=0}^5 s_{ki} I_k ;$$

$$\text{where } z_j = \frac{1}{1 + \exp(-(y_j + b_j))}$$





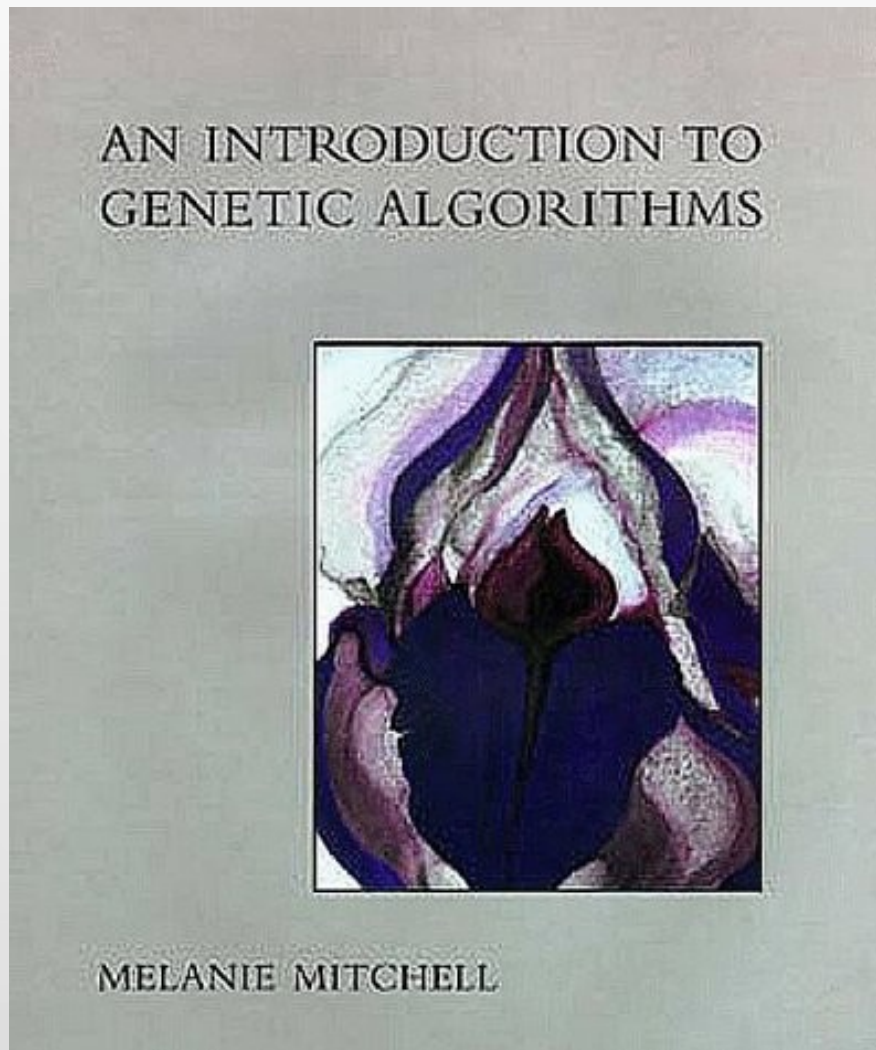
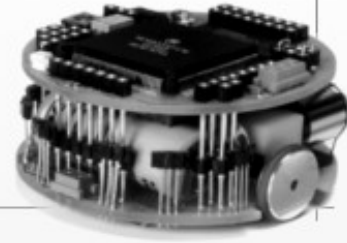
# Cuerpo y Entorno



- Simulación biomecánica del cuerpo y el entorno
  - Diferentes grados de realismo
- No sensores mágicos
- También en entornos físicos reales
  - Transferencia probada con envoltorio de ruido



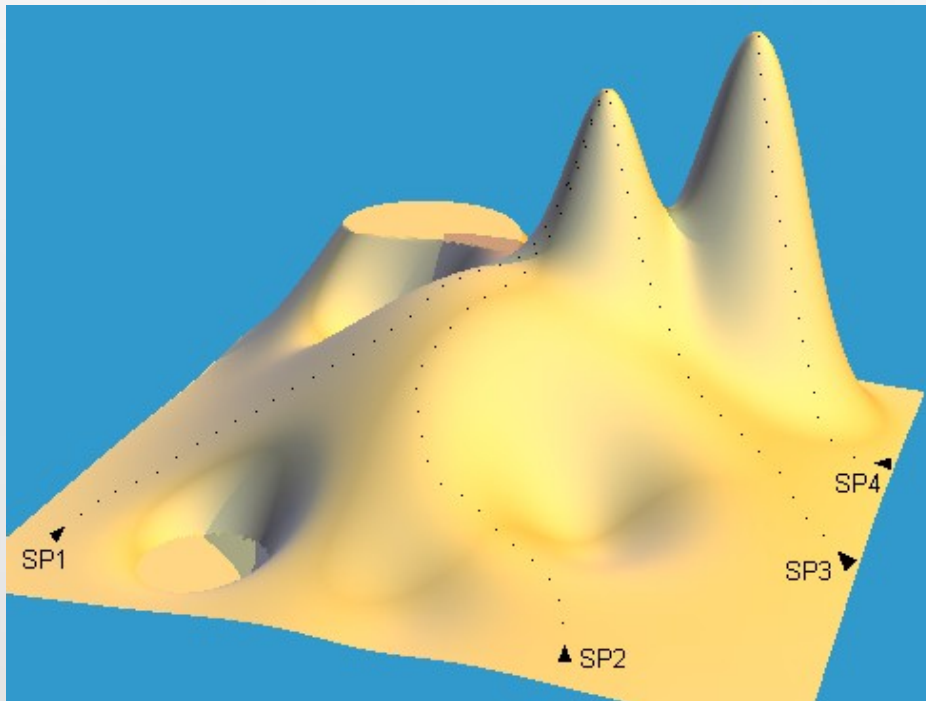
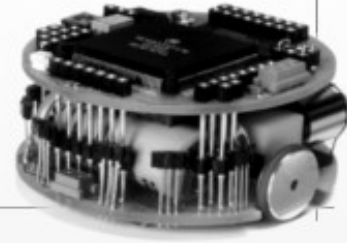
# Evolución Artificial



- Holland (1975) *Adaptation in natural and artificial systems*
- Algoritmos Genéticos:
  - 0. Crear población inicial de soluciones posibles
  - 1. Evaluar el *fitness* de cada individuo
  - 2. Reproducir la siguiente generación en proporción a la clasificación de fitness
  - 3. Mutar
  - 4. Volver a 1



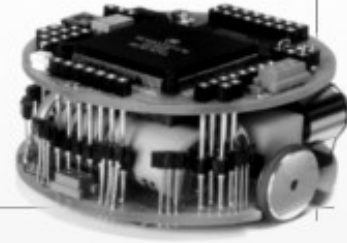
# Evolución Artificial



- Optimización por *Hill Climbing*
- Ejemplo: ala de un avión
  - x: anchura
  - y: longitud
  - z: fitness (resistencia y capacidad de vuelo y maniobra)



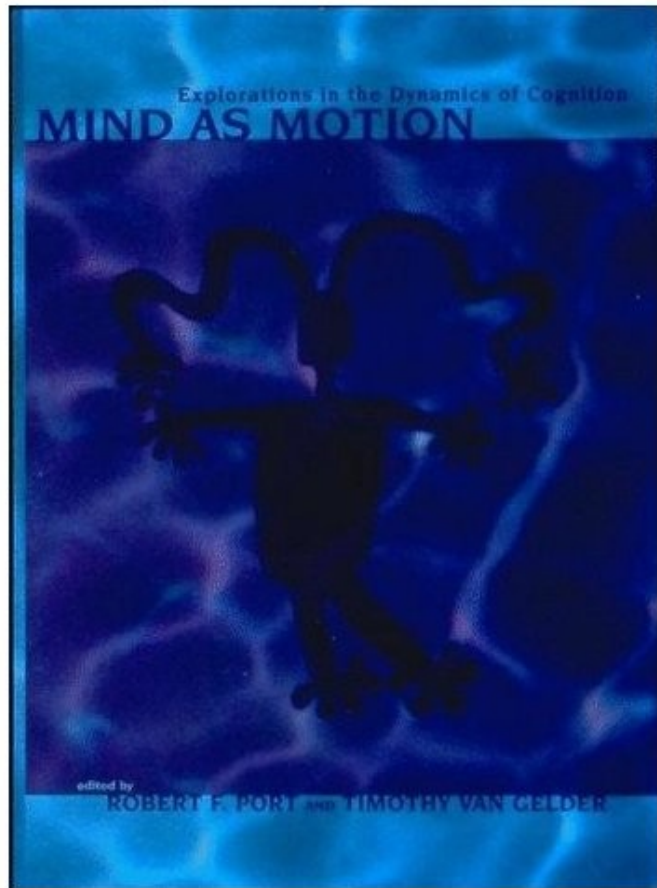
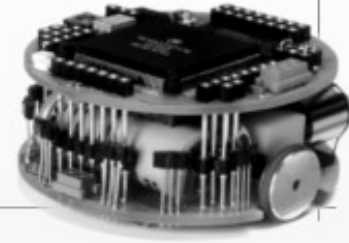
# Robótica Evolutiva II



## Principios



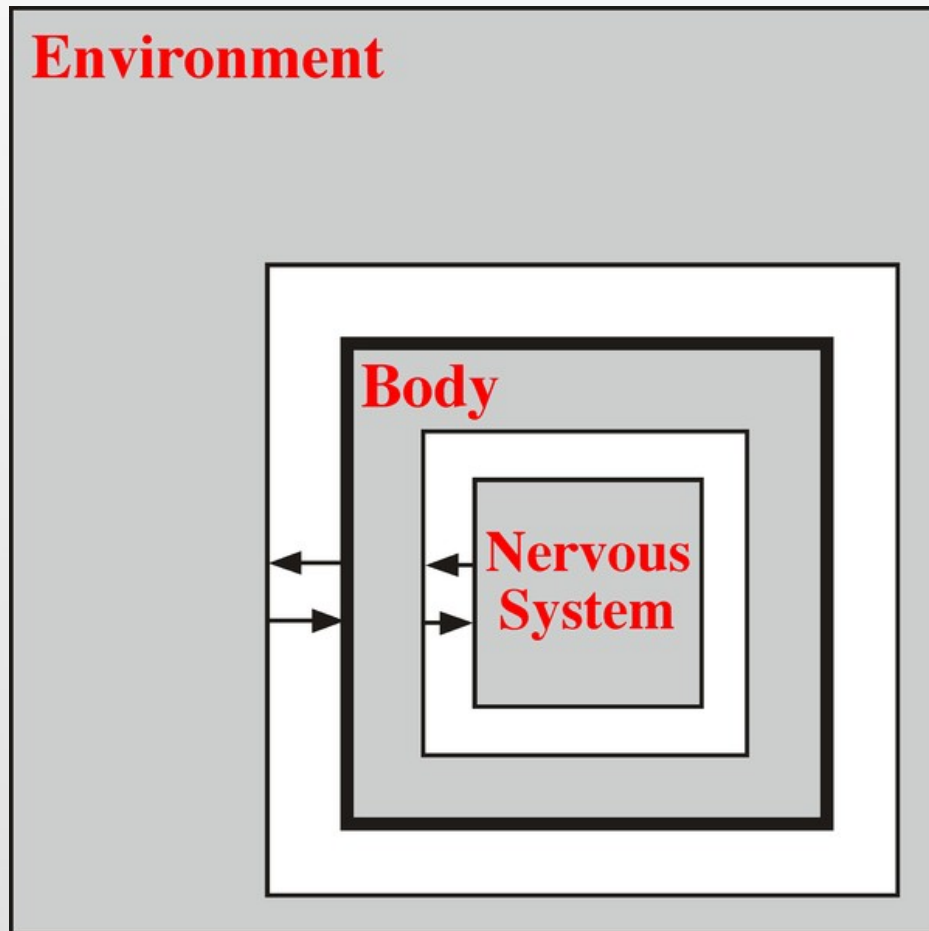
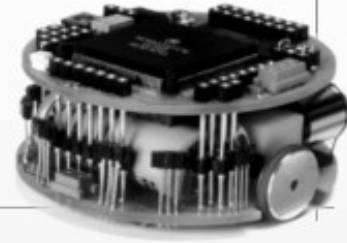
# Hipótesis Dinámica



- Port & van Gelder (1995)  
*Mind as Motion*
- Romper con el computacionalismo de estados discretos y simbólicos
- Los sistemas dinámicos permiten modelar de forma integrada diferentes niveles de realismo biológico



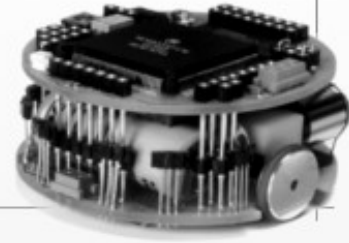
# El continuo cerebro-cuerpo-entorno



- La cognición emerge de la interacción recurrente entre el cerebro, el cuerpo y el entorno
- No hay privilegio causal (el cerebro como controlador)
- La biomecánica y el contexto conductual son parte integral de la mente



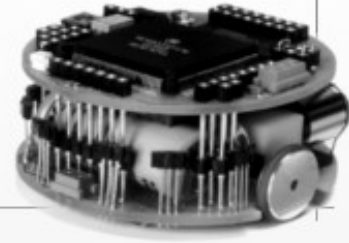
# Modelar sin prejuicios teóricos



- Harvey (1999) *Philosophy of the Mind using a Screwdriver*
- La mejor forma de testear una teoría sobre la cognición es construyendo un robot (modelo generativo)
- Toda reconstrucción implica preconcepciones teóricas
- La evolución artificial se encarga del proceso de diseño, liberando al teórico de preconcepciones
- Permite explorar aspectos teóricos de forma práctica y concreta



# Robótica Evolutiva III

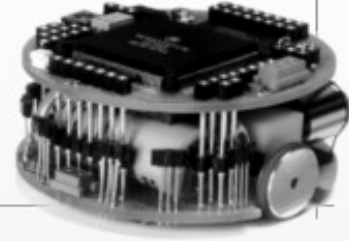


Adaptación a la inversión visual  
Di Paolo 2000





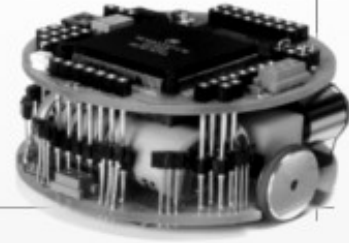
# Experimentos de rehabilitación



- Ivo Kohler 60's
- “Habits exist in all areas of human personality” (p. 137) and only when we undergo a strong process of rehabilitation “do we notice what habit is, and to what extent we consist of many and strong habits” (p. 138).
- La mayoría de los hábitos son transparentes para nosotros, sólo perturbando un hábito podemos conocer su estructura
- Inversión visual



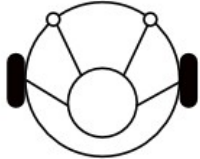
# Plasticidad homeostática



- Reglas de Hebb se activan cuando las neuronas sobrepasan un cierto umbral de activación
- Evidencia empírica (Turrigiano 1999), neuronas buscan estabilidad en su frecuencia de disparo
- Se selecciona para estabilidad interna y estabilidad conductual, de tal forma que la estabilidad de la conducta aparezca acoplada a la estabilidad de los parámetros sinápticos
- Cuando la estabilidad conductual se pierde induce inestabilidad sináptica hasta recuperar la estabilidad conductual (principio del Homeostat de Ashby)



**A**

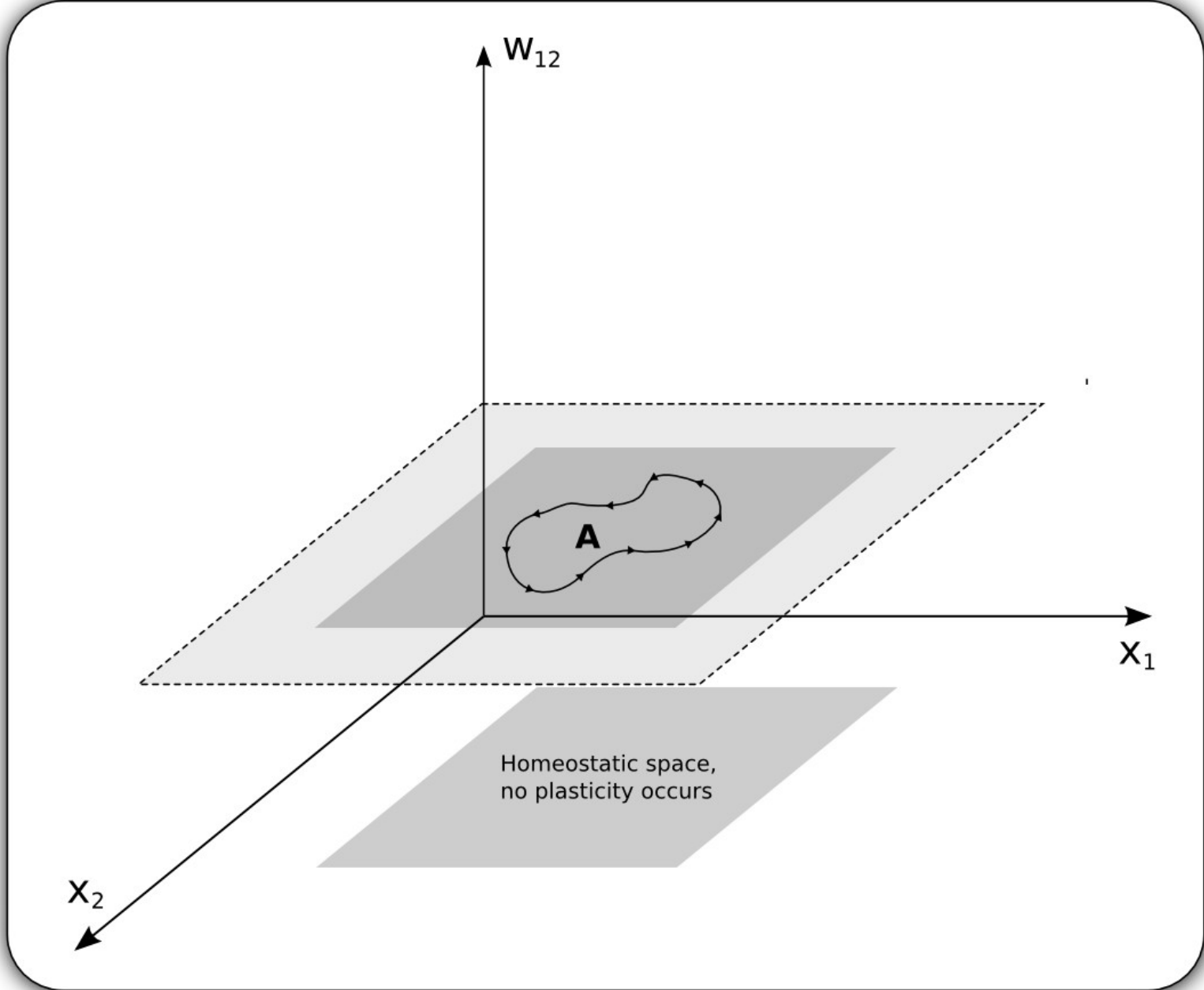
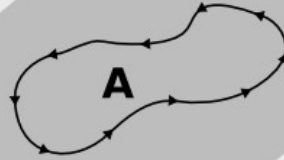


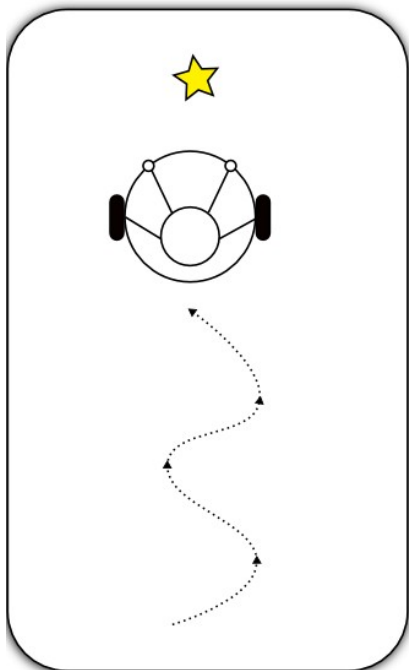
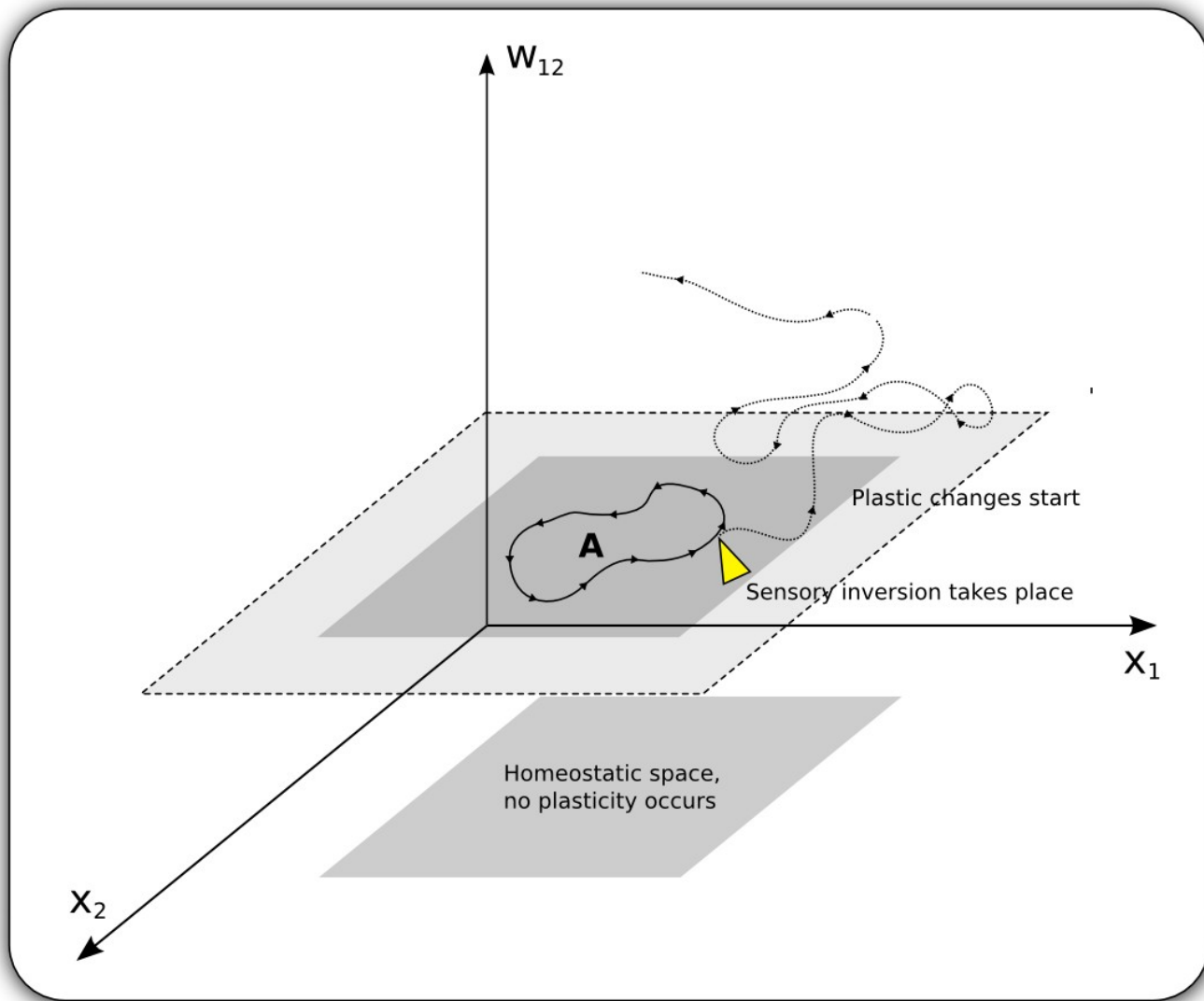
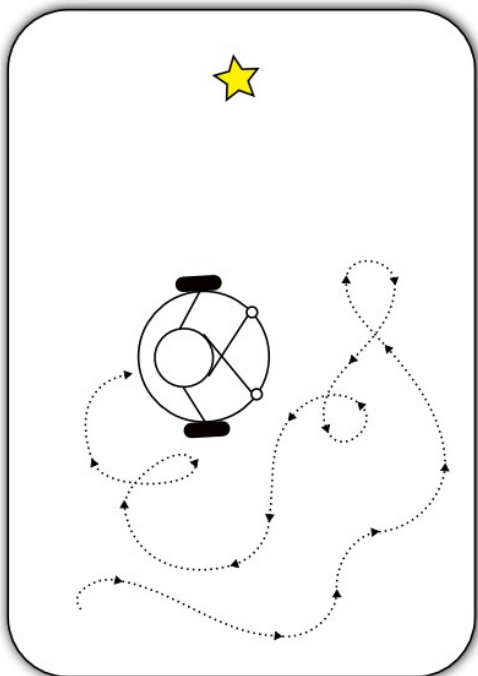
$W_{12}$

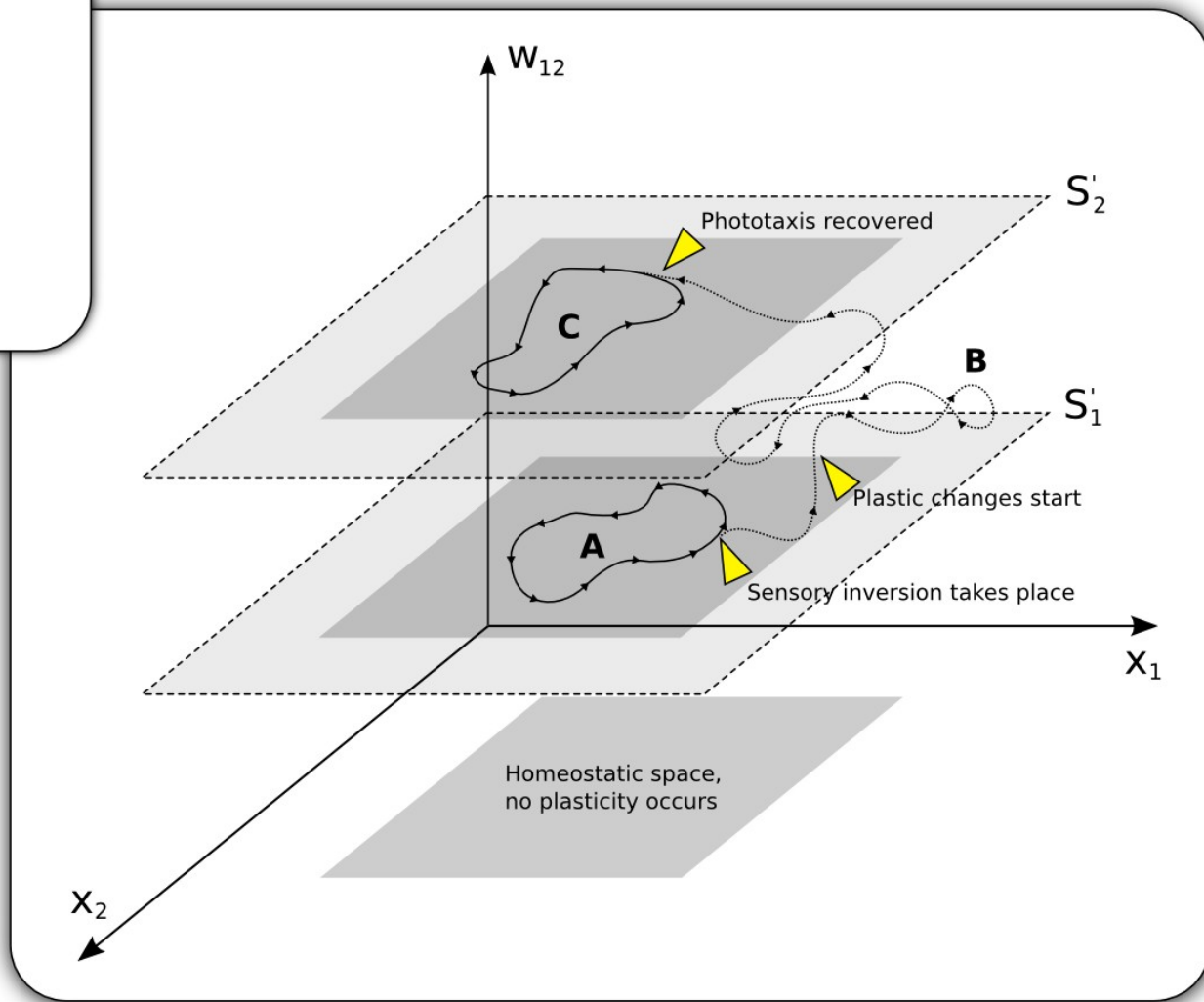
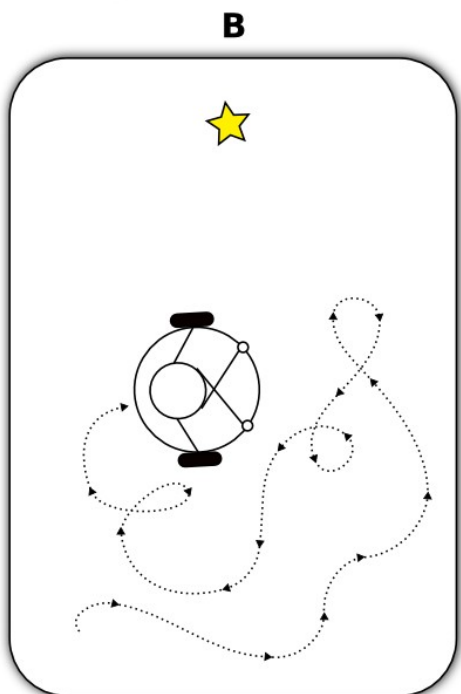
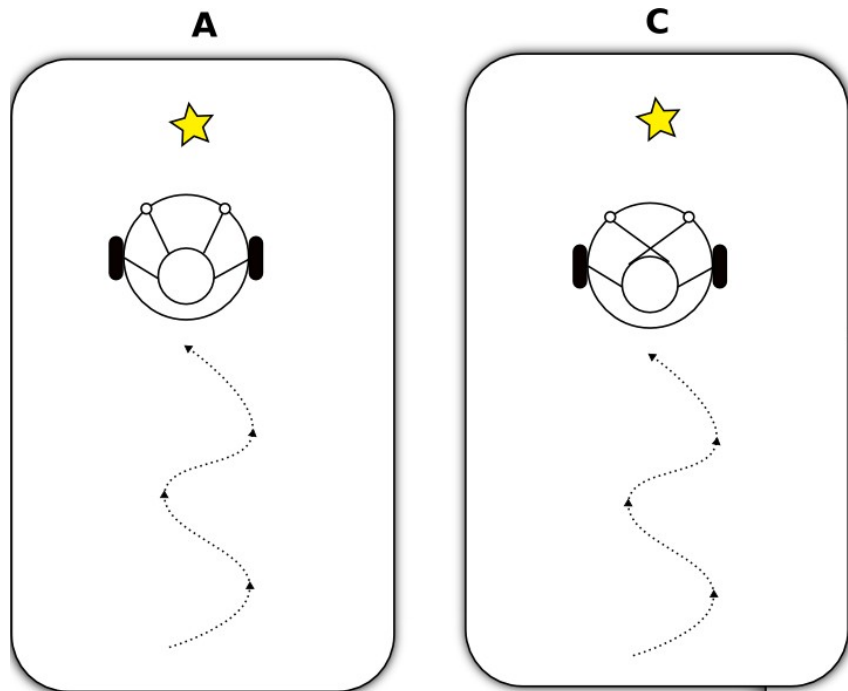
$x_1$

$x_2$

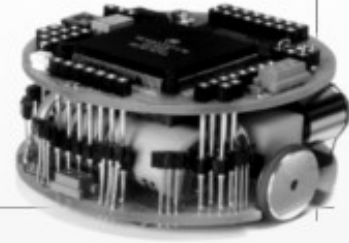
Homeostatic space,  
no plasticity occurs



**A****B**

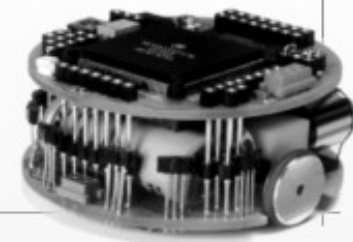


# Conclusiones



- Dos problemas fundamentales de la cognición (los problemas de la identidad y de la normatividad) pueden resultar peligrosamente “invisibles” para la ingeniería robótica
- Un robot genuinamente cognitivo requiere de autonomía en un sentido fuerte
- Pero no tiene porqué estar biológicamente vivo...
- El concepto de *vida mental* (el automantenimiento adaptativo de la organización de la conducta) es un camino a explorar y puede hacerse de forma rigurosa
- Hay que hacer robots que puedan volverse locos pero luchan por mantener su coherencia interna





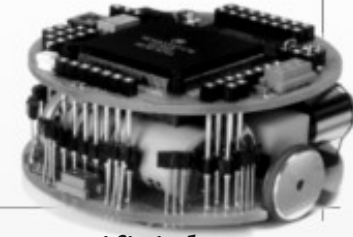
# *fin*

¡gracias!

*Copyright 2008 Xabier Barandiaran: eres libre de copiar, modificar y distribuir esta charla siempre y cuando mantengas esta nota, bajo los términos de la licencia GFDL*



# Referencias

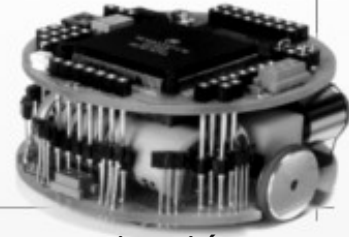


- Beer, R. D. (1995) A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence* 72:173—215.
- Di Paolo, E. (2000) Homeostatic adaptation to inversion of the visual field and other sensorimotor disruptions. In Meyer, J.-A., Berthoz, A., Floreano, D., Roitblat, H., and Wilson, S., editors, *From Animals to Animats 6: Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior*, pages 440—449. Harvard, MA: MIT Press.
- Di Paolo, E.A. (2003) Organismically-inspired robotics: homeostatic adaptation and natural teleology beyond the closed sensorimotor loop. In Murase, K., Asakura, T. (Eds.), *Dynamical Systems Approach to Embodiment and Sociality*. Advanced Knowledge International, Adelaide, pp. 19–42.
- Gánti, T. (1984/1991) *The Chemoton Theory. Vol I: Theoretical Foundations of Fluid Machines*. OMIKK, Budapest, 1984.
- James, W. (1890) *The Principles of Psychology Vol. I*, Henry Holt.
- Jonas, H. (1966/2001) *The phenomenon of life: Toward a philosophical biology*. Evanston, IL: Northwestern University Press.
- Kohler, I. (1964) The formation and transformation of the perceptual world. *Psychological Issues* 3: 1—173.





# Referencias



- Maturana, H.R. & Varela, F. (1973/1980) *De Máquinas y Seres Vivos: Una teoría sobre la organización biológica*, Santiago de Chile: Editorial Universitaria. English version: Maturana, H. & Varela, F. (1980) *Autopoiesis and Cognition: The realization of the living*. D. Reidel.
- Piaget, J. (1967 /1969) *Biología y Conocimiento. Ensayo sobre las relaciones entre las regulaciones orgánicas y los procesos cognoscitivos*. Siglo veintiuno editores.
- Port, R. & van Gelder, T. (1995) *Mind as motion: Explorations in the dynamics of cognition*. Cambridge, MA: MIT Press.
- Rosenblueth, A., Wiener, N., & Bigelow, J. (1943) Behavior, purpose and teleology. *Philosophy of Science* **10**(1):18—24.
- Ruiz-Mirazo, K. & Moreno, A. (2004) Basic Autonomy as a Fundamental Step in the Synthesis of Life. *Artificial Life*, **10**:235—259.
- Smithers, T. (1995) Are autonomous agents information processing systems? In Steels, L. and Brooks, R.A. (Eds.) *The artificial life route to artificial intelligence: Building situated embodied agents*. New Haven: Erlbaum.
- Smithers, T. (1997) Autonomy in Robots and Other Agents. *Brain and Cognition* **34**, 88—106.
- Turrigiano, G. G. (1999). Homeostatic plasticity in neuronal networks: The more things change, the more they stay the same. *Trends Neurosci.*, **22**:221–227.

