

Miguel Aguilera Lizarraga

# Interaction dynamics and autonomy in cognitive systems

Departamento  
Informática e Ingeniería de Sistemas

Director/es

González Bedia, Manuel Andrés  
Serón Arbeloa, Francisco  
Barandiarán Fernández, Xavier Eugenio

<http://zaguan.unizar.es/collection/Tesis>

© Universidad de Zaragoza  
Servicio de Publicaciones

ISSN 2254-7606



**Universidad**  
Zaragoza

Tesis Doctoral

# INTERACTION DYNAMICS AND AUTONOMY IN COGNITIVE SYSTEMS

Autor

Miguel Aguilera Lizarraga

Director/es

González Bedia, Manuel Andrés  
Serón Arbeloa, Francisco  
Barandiarán Fernández, Xavier Eugenio

**UNIVERSIDAD DE ZARAGOZA**  
Informática e Ingeniería de Sistemas

2015



# Interaction Dynamics and Autonomy in Cognitive Systems

A DISSERTATION PRESENTED

BY

MIGUEL AGUILERA

TO

THE DEPARTMENT OF COMPUTER SCIENCE AND SYSTEMS ENGINEERING

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY.

SUPERVISORS:

MANUEL G. BEDIA

XABIER E. BARANDIARAN

FRANCISCO SERÓN

UNIVERSITY OF ZARAGOZA

ZARAGOZA, SPAIN

SEPTEMBER 2015





COPYLEFT ©2015 – MIGUEL AGUILERA

INTERACTION DYNAMICS AND AUTONOMY IN COGNITIVE SYSTEMS v.i.o.

[HTTP://PHDTHESIS.MAGUILERA.NET/](http://phdthesis.maguilera.net/)

CREATIVE COMMONS ATTRIBUTION-SHAREALIKE 4.0 INTERNATIONAL LICENSE.

YOU ARE FREE TO COPY AND REDISTRIBUTE THE MATERIAL IN ANY MEDIUM OR FORMAT, REMIX, TRANSFORM, AND BUILD UPON THE MATERIAL FOR ANY PURPOSE, EVEN COMMERCIALY. THE LICENSOR CANNOT REVOKE THESE FREEDOMS AS LONG AS YOU FOLLOW THE LICENSE TERMS. UNDER THE FOLLOWING TERMS: A) ATTRIBUTION – YOU MUST GIVE APPROPRIATE CREDIT, PROVIDE A LINK TO THE LICENSE, AND INDICATE IF CHANGES WERE MADE. YOU MAY DO SO IN ANY REASONABLE MANNER, BUT NOT IN ANY WAY THAT SUGGESTS THE LICENSOR ENDORSES YOU OR YOUR USE. B) SHAREALIKE – IF YOU REMIX, TRANSFORM, OR BUILD UPON THE MATERIAL, YOU MUST DISTRIBUTE YOUR CONTRIBUTIONS UNDER THE SAME LICENSE AS THE ORIGINAL. NO ADDITIONAL RESTRICTIONS — YOU MAY NOT APPLY LEGAL TERMS OR TECHNOLOGICAL MEASURES THAT LEGALLY RESTRICT OTHERS FROM DOING ANYTHING THE LICENSE PERMITS.

THE COMPLETE TERMS OF THE LICENSE CAN BE FOUND AT:

[HTTP://CREATIVECOMMONS.ORG/LICENSES/BY-SA/4.0/LEGALCODE](http://creativecommons.org/licenses/by-sa/4.0/legalcode)

Thesis supervisors:  
Manuel G. Bedia  
Xabier E. Barandiaran  
Francisco Serón

Author:  
Miguel Aguilera

## Interaction Dynamics and Autonomy in Cognitive Systems

### ABSTRACT

The concept of autonomy is of crucial importance for understanding life and cognition. Whereas cellular and organismic autonomy is based in the self-production of the material infrastructure sustaining the existence of living beings as such, we are interested in how biological autonomy can be extended into forms of autonomous agency, where autonomy as a form of organization is extended into the behaviour of an agent in interaction with its environment (and not its material self-production). In this thesis, we focus on the development of operational models of sensorimotor agency, exploring the construction of a domain of interactions creating a dynamical interface between agent and environment. We present two main contributions to the study of autonomous agency:

First, we contribute to the development of a modelling route for testing, comparing and validating hypotheses about neurocognitive autonomy. Through the design and analysis specific neurodynamical models embedded in robotic agents, we explore how an agent is constituted in a sensorimotor space as an autonomous entity able to adaptively sustain its own organization. Using two simulation models and different dynamical analysis and measurement of complex patterns in their behaviour we are able to tackle some theoretical obstacles preventing the understanding of autonomy in the sensorimotor domain, and to generate new predictions about the nature of autonomous agency in the neurocognitive domain.

Second, we explore the extension of sensorimotor forms of autonomy into the social realm. We analyse two cases from an experimental perspective: the constitution of a collective subject in a sensorimotor social interactive task, and the emergence of an autonomous social identity in a large-scale technology-mediated social system. Through the analysis of coordination mechanisms and emergent complex patterns, we are able to gather experimental evidence indicating that in some cases social autonomy might emerge based on mechanisms of coordinated sensorimotor activity, constituting forms of collective autonomous agency.



Directores de tesis:  
Manuel G. Bedia  
Xabier E. Barandiaran  
Francisco Serón

Autor:  
Miguel Aguilera

## Dinámicas de Interacción y Autonomía en Sistemas Cognitivos

### RESUMEN

El concepto de autonomía es de crucial importancia para entender la vida y la cognición. Mientras que formas de autonomía celular y orgánicas están basadas en la auto-producción de la infraestructura material que sostiene la existencia de un ser vivo como tal, en este trabajo nos interesamos en cómo la autonomía biológica puede extenderse a formas de agencialidad autónoma, en las que la autonomía como forma de organización se extiende al comportamiento de un sistema en interacción con su entorno (y no únicamente su auto-producción material). Esta tesis se centra en el desarrollo de modelos operacionales de agencialidad sensoriomotora, explorando la construcción de un dominio de interacción capaz de crear una frontera dinámica entre agente y entorno. Presentamos dos contribuciones principales al estudio de la agencialidad autónoma:

En primer lugar, contribuimos al desarrollo de un enfoque basado en modelos para el ensayo, comparación y validación de hipótesis sobre autonomía neurocognitiva. A través del diseño y análisis de modelos neurodinámicos específicos embebidos en agentes robóticos, exploramos cómo un agente se constituye en el espacio sensoriomotor como una entidad autónoma capaz de sostener adaptativamente su propia organización. Usando dos modelos simulados, diferentes análisis dinámicos y medidas de patrones complejos en su comportamiento, somos capaces de abordar algunos obstáculos en la definición de la autonomía en el dominio sensoriomotor, así como generar nuevas predicciones sobre la naturaleza de la agencialidad autónoma en el dominio neurocognitivo.

En segundo lugar, exploramos la extensión de formas de autonomía sensoriomotora en el campo de lo social. Analizamos dos casos desde una perspectiva experimental: la constitución de un sujeto colectivo en una tarea de interacción social sensoriomotora, y la emergencia de una identidad social autónoma en un sistema social a gran escala construido sobre el uso de herramientas tecnológicas. A través del análisis de mecanismos de coordinación y de patrones complejos emergentes, somos capaces de reunir evidencia experimental indicando que en algunos casos una autonomía social puede emerger basada en mecanismos de actividad sensoriomotora coordinada, constituyendo formas de agencialidad autónoma colectiva.

## CONCLUSIONES

En los dos ámbitos, la cognición sensoriomotora y la interacción social, en primer lugar caracterizamos la constitución de un sujeto en un espacio fundamentalmente interactivo: a través de un agente robótico en una tarea sensoriomotora simulada por ordenador en el primer caso, y en un entorno digital que conecta a dos agentes humanos en el segundo. Posteriormente, exploramos cómo un agente autónomo puede emerger a partir del espacio de interacción en ambos casos.

En el caso de un agente cognitivo analizamos en detalle cómo un agente autónomo se constituye a partir de la interacción circular entre sus patrones neurodinámicos, la plasticidad sináptica que regula su estructura neuronal y los patrones sensoriomotores que emergen de su interacción con el mundo. La agencialidad autónoma surge a través de la autoorganización entre diferentes escalas de comportamiento, en un proceso de ruptura de la simetría en la relación circular entre estos tres elementos, en los que la asimetría en la interacción distingue al entorno, cuya complejidad de estímulos genera una multiplicidad de respuestas neurodinámicas; y un agente, cuya reorganización interna le permite actuar de forma coherente y unitaria sobre este entorno.

En el caso de la interacción social, estudiamos cómo formas de organización y movilización social facilitadas por el uso de tecnologías de la información pueden permitir formas de agencialidad autónoma similares a las analizadas en agentes neurocognitivos. Observamos cómo 'identidades multitudinarias' se configuran mediante patrones de coordinación entre diferentes escalas en una estructura distribuida e internamente múltiple.

Desde un punto de vista metodológico, mostramos el potencial de indicadores de complejidad (medidas de teoría de la información, sincronización metastable, exponentes de criticalidad...) como indicadores de aspectos cuantitativos y cualitativos de la autonomía. El estudio de la interacción (entre elementos, entre escalas temporales y espaciales, etc.) destaca la importancia de la influencia y modulación entre niveles de actividad – micro, meso y macroscópicos – en un sistema autónomo. La autonomía no es sólo un proceso de clausura y aislamiento respecto al entorno, sino que la agencialidad autónoma es un proceso que se despliega a diferentes escalas, que son coordinadas a través de influencias causales asimétricas, dando lugar al agente, al mundo sobre el que éste actúa, y al dominio de interacción que conecta a ambos. Más allá de estos indicadores, mostramos cómo aspectos conceptuales relacionados con la metodología empleada para aplicar estas medidas en nuestros modelos y análisis experimentales – como la clausura operacional de una red y la precariedad de sus interdependencias, o la comparación entre agentes situados en su entorno y agentes pasivamente acoplados a él – nos proporcionan una rica perspectiva para analizar las propiedades específicas de la organización autónoma. Aunque los indicadores de complejidad no son suficientes por sí mismos para detectar formas de organización autónoma, nos ayudan a probar y delimitar algunos de los aspectos conceptuales relacionados con ella.

# Publications

## JOURNAL ARTICLES

- Aguilera, M., Barandiaran, X. E., Bedia, M. G., and Seron, F. (2015). Self-Organized Criticality, Plasticity and Sensorimotor Coupling. Explorations with a Neurorobotic Model in a Behavioural Preference Task. *PLoS ONE*, 10(2):e0117465. DOI:10.1371/journal.pone.0117465.
- Monterde, A., Calleja-López, A., Aguilera, M., Barandiaran, X. E., and Postill, J. (2015). Multitudinous identities: a qualitative and network analysis of the 15M collective identity. *Information, Communication & Society*, 18(8):930–950. DOI: 10.1080/1369118X.2015.1043315.
- Aguilera, M., Bedia, M. G., Barandiaran, X. E. & Serón, F. (2014). Intermittent animal behaviour: the adjustment deployment dilemma. *Artificial Life*, 20(4). DOI: 10.1162/ARTL\_a\_00133.
- Bedia, M. G., Aguilera, M., Gómez, T., Larrode, D. G., and Seron, F. (2014). Quantifying long-range correlations and 1/f patterns in a minimal experiment of social interaction. *Frontiers in Psychology*, 5:1281. DOI: 10.3389/fpsyg.2014.01281.
- Aguilera, M., Bedia, M. G., Santos, B. A., and Barandiaran, X. E. (2013a). The situated HKB model: how sensorimotor spatial coupling can alter oscillatory brain dynamics. *Frontiers in Computational Neuroscience*, 7:117. DOI: 10.3389/fncom.2013.00117.
- Santos, B., Barandiaran, X., Husbands, P., Aguilera, M., and Bedia, M. (2012). Sensorimotor coordination and metastability in a situated HKB model. *Connection Science*, 24(4):143–161. DOI: 10.1080/09540091.2013.770821.

## BOOK CHAPTERS

- Aguilera, M. (2015) La conciencia red de la multitud conectada: la mente colectiva del 15M. In Toret, J. (Ed.). *Tecnopolítica y 15M. La potencia de las multitudes conectadas*, 144-161. Barcelona: Editorial UOC. ISBN: 978-84-9064-458-4.
- Barandiaran, X. E. & Aguilera, M. (2015). Neurociencia y tecnopolítica: hacia un marco analógico para comprender la mente colectiva del 15M. En J. Toret (Ed.), *Tecnopolítica y 15M. La potencia de las multitudes conectadas*, 163-211. Barcelona: Editorial UOC. ISBN: 978-84-9064-458-4.

## CONFERENCE PROCEEDINGS

- Aguilera, M., Morer, I., Barandiaran, X. E., and Bedia, M. G. (2013b). Quantifying political self-organization in social media. fractal patterns in the Spanish 15m movement on twitter. In *Advances in Artificial Life, ECAL*, volume 12:395–402.
- Izquierdo, E.J., Aguilera, M. and Beer, R.D. (2013). Analysis of ultrastability in small dynamical recurrent neural networks. In *Advances in Artificial Life, ECAL*, volume 12:51–58.
- Castillo, L., Isaza, . Bedia, M.G., Aguilera, M. Correa, J. (2012). Grid Computing and CBR Deployment: Monitoring Principles for a Suitable Engagement. 9TH Distributed Computing and Artificial Intelligence. *Advances in Intelligent and Soft Computing, 2012*, Volume 51/2012, 355-361.
- Aguilera, M., Bedia, M. G., Barandiaran, X. E. & Serón, F. (2011). The adjustment-deployment dilemma in organism's behaviour: theoretical characterization and minimal model. *Proceedings of the IEEE Symposium on Artificial Life*.
- Castillo, L. F., Bedia, M. G., Aguilera, M., & Uribe, A. L. (2011). A proposal for improving retrieval processes in case-based reasoning systems able to act in dynamic environments. Proceedings of the International Symposium on Distributed Computing and Artificial Intelligence 2011.
- Castillo, L. F., Bedia, M. G., Aguilera, M., Uribe A. L., Manrique, M. & Isaza, G. (2011). Case-based reasoning and real-time systems: exploiting successfully poorer solutions. *Proceedings of the 6th Colombian Computing Congress*.
- Castillo, L. F., Bedia, M.G., Aguilera, M. & Uribe, A. L. (2011). Grid computing and CBR Deployment: Monitoring Principles for a Suitable Engagement. *Proceedings of the 2011 International Conference on Grid Computing and Applications*.

# Contents

o	INTRODUCTION	17
I	Conceptual and Methodological Review	27
1	AN INTRODUCTION TO AUTONOMY	29
1.1	Persistence and Identity in Living Beings . . . . .	31
1.2	Metastability, Soft-Assemblies & Critical Self-Organization . . . . .	48
1.3	Autonomy as an Organizational Principle of Life: from cells to brains to society . . . . .	62
1.4	Exploring Autonomous Agency . . . . .	67
II	Sensorimotor Constitution of Neurocognitive Autonomy	71
2	CONSTITUTIVE INTERACTION: THE SITUATED-HKB MODEL	73
2.1	Situatedness and Embodiment in Cognitive Science . . . . .	74
2.2	The HKB model . . . . .	87
2.3	The Situated HKB Model . . . . .	88
2.4	Three-Dimensional Description of the Situated-HKB Model . . . . .	89
2.5	Comparison between Situated, Passively-Coupled and Decoupled HKB systems . . . . .	96
2.6	Discussion . . . . .	106
3	SITUATED AUTONOMY: ORGANIZATIONAL HOMEOSTASIS IN THE SENSORIMO- TOR LOOP	111
3.1	Decision-Making, Habits and Autonomy . . . . .	112
3.2	Neurocognitive Autonomy . . . . .	115
3.3	A Situated Dynamic Core . . . . .	119
3.4	Model of a Neurodynamic Controller with Relational Homeostasis Embedded in a Robotic Agent . . . . .	124
3.5	Behavioural Analysis of the Evolved Agent . . . . .	129
3.6	Dynamics and Information Flows in the System . . . . .	133
3.7	Discussion . . . . .	149
4	SITUATED AUTONOMY: COMPLEXITY AND SELF-ORGANIZED CRITICALITY	153
4.1	Experimental Setup . . . . .	154
4.2	Metastable Organization . . . . .	155
4.3	Complex Systems Tools for Characterizing Autonomous Organization . . . . .	160

4.4	Integration and Segregation among the System's Components: TSE complexity . . . . .	163
4.5	Frequency Domain of Metastability: 1/f noise . . . . .	167
4.6	Discussion . . . . .	177
III	Interactive Autonomy in Social Processes	185
5	SOCIAL AGENCY: INTERACTION-DOMINANT DYNAMICS IN SOCIAL RECIPROCITY	187
5.1	Direct Perception and Social Interaction . . . . .	188
5.2	Constitutive Interaction in a Minimal Social Setting . . . . .	190
5.3	Theoretical Framework . . . . .	193
5.4	Experimental Setup and Methods of Analysis . . . . .	196
5.5	Preliminary Analysis: short-scale vs. multiscale descriptions . . . . .	198
5.6	Fractal Dynamics in the Structure of Social Interaction . . . . .	201
5.7	Discussion . . . . .	207
6	THE COLLECTIVE MIND: LARGE-SCALE SELF-ORGANIZATION IN SOCIAL SYSTEMS	211
6.1	Social Agency . . . . .	212
6.2	The 15M Movement and its Evolution . . . . .	220
6.3	The 15M as a Coordinative Structure . . . . .	222
6.4	Delimitation and Characterization of the 15M Collective Identity . . . . .	223
6.5	Self-Synchronizing Agency . . . . .	227
6.6	Multiscale Organization and Embodied Action . . . . .	237
6.7	Discussion . . . . .	243
IV	Conclusion	249
7	SUMMARY AND CONCLUSION	251
7.1	Recapitulation . . . . .	251
7.2	Conclusion . . . . .	257
	REFERENCES	293

# Acknowledgements

As with most scientific research, the work presented here is the product of many collective efforts.

I must thank all the people around ISAAC research group for the discussions and suggestions that have shaped some of the ideas in this thesis. I am grateful to Tomás Gómez for valuable discussions about conceptual and methodological notions about collective agency. Also, Tomás did part of the analysis of Chapter 5 and David Larrode developed the experimental crossing platform we used. Thanks to Carlos Alquézar for his enthusiasm in pushing forward ISAAC as an inspiring intellectual environment and for multiple discussions about neurodynamic organization. Thanks to all the other people that have participated in discussions or projects in ISAAC seminars and discussions during this time.

This thesis project has benefited greatly from the research visits I had the opportunity to enjoy during my work. I would like to thank Randall Beer for inviting me to visit his lab at Indiana University, as well as for his attention, discussions and comments about my work while I was there. I am grateful to Eduardo Izquierdo, Eran Agmon and the other people at Randall Beer's laboratory for discussions, suggestions and comments about my work. Also, thanks to Takashi Ikegami for inviting me to visit his lab at the University of Tokyo, as well as for valuable insights into research of massive social autonomy. Thanks to Federico da Rold for many discussions during my time there, and the members of the Ikegami Lab for their interesting seminars and feedback.

This thesis as it would never have been possible if I were not myself immersed in the social and political events that have transformed Spain during the last years. Thanks to the acts of hundreds of thousands of anonymous citizens, I have had the opportunity to witness from a privileged position the functioning of self-organizing autonomy at a massive scale and to participate in the collective constructions of a better future. I have to thank everyone inside and around DatAnalysis15M research network for embarking in the collective enterprise of scientifically and politically understanding the complex social processes that were changing our lives. Thanks to Pablo Aragón, Eunete, Óscar, David, Arnau, Toret, Pablo Rey, Emanuele, Ignacio, Luce, Juanli, Ariadna, Xabier, Antonio and Alberto for making it possible. I shall especially thank Javier Toret for introducing me to this research project and for his passionate and enthusiastic work in understanding and weaving the organizational infrastructure sustaining the revolution of connected multitudes. As well, thanks to Arnau Monterde, Antonio Calleja and Xabier Barandiaran for their conceptual contributions and the development of the notion of *multitudinous identities* and much of the discussion addressed in Chapter 6. Also, I have to thank Ignacio and Emanuele for designing with me our incursion into the analysis of network movements which eventually led me to DatAnalysis15M and the research about social agency presented here.

I am also appreciative the several comments and suggestions made by anonymous referees on my journal and conference publications.

Many people have provided emotional support and intellectual inspiration at different points during this intense endeavour. Thanks to Raúl, Guillermo, Emanuele, Elena P., Hilfe, David, Dani G., Pablo, Elena G., Pato, Lorién, Marta... for boosting my academic and political imagination and

for building common affinity spaces and life projects. Thanks to my old flatmates and class friends (Miguel, Manu, Jose, Jorge) and lifelong friends (David, Nacho, Luisja) for their company and encouragement. I have to thank the communities around Pantera Rossa, Nociones Comunes Zaragoza, Partido X, Zaragoza en Común, Podemos Aragón for providing an important network of care, collective intelligence and shared projects that have indirectly influenced and enriched my work. I also have to thank my family, my parents and brother, for their unconditional support since I dived into the academic world. Overall, I owe a huge amount of support to Lara. She has been the most caring and intelligent partner and a source of inspiration and encouragement in developing the work presented here.

A special mention is deserved to my thesis supervisors. Manuel Bedia introduced me to the world of complex adaptive systems and, since then, he has been an outstanding supervisor. Manuel has supervised with extraordinary care and dedication all the work actualized in this manuscript. In addition, he has persistently accompanied and supported me during this intellectual journey, often shedding light on the vast amount of possibilities of the ramifications for our research. Xabier Barandiaran has been a continuous source of philosophical inspiration and a sharp mentor unveiling the significance and potential of much of the work developed here. This thesis owes much to Xabier's creativity and scientific imagination, as well as his careful reviewing and assessment. Finally, Francisco Serón has been an inestimable support in fostering the material and intellectual conditions necessary to develop the work presented here, as well as a source of advice to help me persevere in the midst of the academic jungle.

Funding for this work has been provided by a FPU predoctoral fellowship from the Spanish Ministry of Education with reference AP-2010-6036. Part of the work was supported by the project TIN2011-24660 funded by the Spanish Ministry of Science and Innovation, and the project FCT-13-7848, funded by the Spanish Foundation for Science and Technology (FECYT). The EUCogIII network financed the author's research visit to Indiana University and the FPU programme financed the author's research visit to the University of Tokyo.





# O

## Introduction

### THE SELFLESS SELF: A SCIENTIFIC PERSPECTIVE ON AUTONOMY

Throughout the history of science, we can outline a tension between reductionist and holistic perspectives (Fang and Casadevall, 2011). That is, the contrast between the scientific approach of looking at specific material parts (we look inside an organism and discover tissues, cells, molecules), and the approach that retains the whole and yet struggles to explain its macroscopic properties. In the reductionist approach, the properties of the parts considered in isolation become simpler and amenable to scientific scrutiny, and yet the very object of study disappears: the organism, the self, the intention (Broad, 1925; Simon, 1996). For the holistic approach, the object of study is taken as a whole, though the behaviour of the mechanisms that built the whole are hindered under a veil of complexity (Weaver, 1991; Campbell, 1974).

This tension is particularly stark when it comes to the study of living beings. Cells, organisms, cognitive agents and communities, make themselves into entities that are different from the environment that surround them. This entity, the ‘self’ that becomes distinct from yet in continuous interaction with it, is on one hand a ‘virtual’ entity with no particular location and not directly reducible to the individual components that conform it, and on the other hand, this self emerges from a material substrate that allows the interaction of its different parts among themselves and, specially, with the environment surrounding them (since every living being needs to be in continuous material and energetic interaction with its environment to stay alive). Nonetheless, the system as a whole behaves as a unit that emerges in the eyes of an observer as something different from its surroundings, yet its components are in continuous interaction with those of its environment. This is what Francisco Varela labelled as a *selfless self* (Varela, 1991), whose definition poses the problem of characterizing the constitution of an *autonomous* agent and its mode of existence. Autonomous agents *actively* sustain their organization, their persistence, in a way that often demands to bring the coordinated whole into the explanation. Their autonomy demands a continuous circulation between whole and parts. In a sense the task ahead is extremely challenging, it demands reconciliation of matter

and form, the mechanical and the ‘spiritual’, to fill the gap between the domain of existence that made gods or spirits necessary to explain our life and environments (identity, form, persistence) and the domain constituted by the opposing earthly tendencies (difference, matter, contingency). In the first macroscopic domain, the self appears as a closed coherent entity, while at the microscopic level, what we observe is fundamentally a set of interactions between parts of the system and between them and the environment.

As we will see in the next chapter the history of the research trying to tackle the nature of autonomy is itself complex, multilayered and non-linear. It is a history of scientific abstractions that covers from the generation of principles or systemic organization to forms of coordination and interaction among its material components. The conceptual progress necessary to bring autonomy into science required deep philosophical insights and innovations. Further progress required new mathematical tools and creative distinctions like those made by cybernetists (Pickering, 2010). Other developments were only possible through complex models and continuing progress demands sophisticated computer simulations (Di Paolo et al., 2000). During the last decades, a series of conceptual innovations including the framework of coordination dynamics, self-organized criticality or the ‘edge of chaos’, together with new methodological and analytical tools have boosted our scientific imagination (Sole and Goodwin, 2008) helping us to understand living systems and imagining new forms of organization beyond reductionist schemata of both functionalist/computationalist paradigms and simplistic holistic/aggregationist approaches (Pigliucci, 2014; Mazzocchi, 2012). However, there are still a good number of problems waiting to be resolved. For example, the properties that are frequently proposed to describe the internal organization of complex autonomous systems can be displayed by toy models (e.g. sandpile models, chaotic maps, etc.) that do not capture the richness of autonomous behaviour. Instead, living autonomy arises in open systems existing within a rich context, constantly changing, growing, evolving, and thereby autonomously changing the way in which they interact with their surrounding environment. (Holland, 1992; Ruiz-Mirazo et al., 2007; Sornette, 2006). The patterns that they exhibit are neither imposed from outside, nor do they arise internally, but are a consequence of the emergence of a domain of interaction between endogenous and exogenous self-organized flows. In this context, this research explores the idea that the explanatory power does not lie that much in the system’s internal invariants but in emergent relational invariants capturing the nature of system-environment interaction and the modulation of such interaction. It is the constitution of an agent-environment interaction domain where the dichotomy between holistic and reductionist approaches can be overcome. In this thesis we focus on how interactive self-organized invariants emerge from and are regulated by agent-environment interaction dynamics themselves. We explore how some of these patterns may apply to a wide range of autonomous behaviour: from sensorimotor coupling in minimal robotics to collective decision making in massive human social networks.

The constitution of agent-environment interaction domains has been explored for some time in metabolic and cellular autonomy, with the emergence of a membrane allowing the

regulation of material and energetic flows between an autonomous organization and the environment Ruiz-Mirazo et al. (2004). Nevertheless, this is not clear at other levels of autonomy. When the relations with the environment are not based in material self-production but in sensorimotor interaction between agential beings and their surroundings (Varela et al., 1993; Kelso, 1995), it is not clear how autonomous organizations are constructed and what are their boundaries. We define these latter case of autonomy built over sensorimotor exchanges as *interactive autonomy*. We study two particular domains of interactive autonomy: the behavioural neurocognitive domain and the social domain. In the former, we make use of simulated robotic models inspired by theoretical neuroscience to characterize in detail the sensorimotor constitution of autonomous mental or behavioural organization. For the latter, we explore in two experimental social scenarios the possibilities of an extension of our model of sensorimotor interactive autonomy to intersubjective interactions and collective entities.

This thesis presents a novelty respect previous work around the concept of autonomy: we propose to study autonomy through the role of interaction dynamics between an autonomous agent and its environment or between more than one autonomous agents, proposing particular forms of autonomy that are build upon this interactive domain. This view allow us to propose a notion of identity (be it sensorimotor or collective) which is constructed directly from an interactive and systemic analysis, without the need to refer to more abstract or anthropocentric categories to account to psychological or social identity.

Since in this thesis we cover a wide range of topics (from neurodynamic and sensorimotor autonomous agency to the constitution of social and collective identities), before entering upon the subject with a theoretical framework about autonomy and interaction, we present a summary of the structure of the thesis to give the reader a global view of the thesis.

## SUMMARY OF CHAPTERS

This thesis is divided in three parts. The first one is focused in the conceptual and methodological aspects of our approach into the study of autonomy. The second part is centered on sensorimotor autonomy, exploring neurodynamic and behavioural models through the use of computer simulations. The third part focuses on forms of social autonomy, analysing experimental data to advance towards models of autonomy in the social realm<sup>1</sup>.

## PART I. CONCEPTUAL AND METHODOLOGICAL REVIEW

Part I of this thesis reviews the theoretical foundations that will guide the following chapters, tracing an evolution of the concepts, framework, analytic and modelling tools and epistemological choices that precede the contemporary account of autonomy, its characterization and application to different experimental domains.

---

<sup>1</sup>For the summary of chapters, we will not make use of references to the literature to facilitate the flow of the text. Appropriate references to the literature relevant to each part of the thesis can be found in the chapters.

## CHAPTER I. AN INTRODUCTION TO AUTONOMY

In the first chapter we review some theoretical notions around the concept of autonomy as an organizational principle of life (in its widest sense, from biological unicellular life, to mental and social life). In the first half of the chapter we introduce ideas that are historically connected to the development of modern views of autonomy: persistence, Spinoza's conatus, homeostasis, regulatory feedback loops, ultrastability and autopoiesis. These ideas help us conceptualize how certain forms of organization allow organisms to maintain their identity while interacting with a changing environment. We introduce autonomous systems as homeostatic systems that maintain their own organization as a system capable of reproducing and maintaining themselves and with the ability to differentiate themselves from their environment, albeit in a process of continuous coupling with the environment. In the case of the cell, a continuous flux of material exchange must take place for maintaining the far from equilibrium thermodynamics that make its own organization possible. In the case of higher organisms, we hypothesize that similarly a circular coupling between the nervous system and the stream of sensory modulation in a sensorimotor loop is necessary to constitute its neurocognitive identity.

During the second half of the chapter we present some key theoretical achievements in complex systems science that help getting a better grasp on the concept of autonomy. We analyse how in parallel to some conceptual advances reviewed in the first part of the chapter, different approaches to the study and design of complex systems have developed a formal framework that allow us to operationalize the generating mechanisms of the kind of processes (self-organized, emergent) underlying autonomous systems. First, we introduce the idea of self-synchronizing systems that, based on distributed coupling mechanisms, allow the emergence of coherent ensembles required for the self-organization of neural or social systems. Secondly, we present the notion of metastability, as it depicts coordination modes which are able to integrate several functional parts and yet also produce individual rhythms in a cooperative and coordinated manner. Finally, the idea of metastability is connected to some recent concepts from physics, specifically the idea of self-organized criticality, which characterizes a system where a metastable state spreads throughout multiple scales of the system behaviour. Self-organized criticality is a property of (classes of) dynamical systems which are able to drive themselves into the scale-invariant dynamics of phase transition states. This allows a system to display a self-regulated metastable configuration at multiple scales.

The chapter next introduces the central role of interaction as a constitutive element of autonomous organization and we extend it to different levels of autonomy. From forms of cellular and organismic autonomy in which the regulation of the flux of energy and matter through the cell is performed through the self-production of the material infrastructure of the living entity, to neurodynamic forms of autonomy in which the material self-constitution is substituted by the modulation of a sensorimotor stream that constitutes the identity of the agent in the perception/motion space. As well, we hypothesize about the possibility of forms of social autonomy, created upon forms of collective agency in interactive communi-

ties. Then, we extend these ideas to a shared inter-personal space of social interaction through the idea of collective autonomy. We detail then our hypothesis that autonomy may arise as a general phenomenon underlying different modes of life at several scales (cellular, organismic, sensorimotor, social). To conclude, we present a preliminary description of autonomy covering the different scales of interaction, finding a transition between forms of autonomy based in the material self-production of the system and those based in the dynamical closure of the system-environment interactions. We propose that for the latter case further conceptual clarifications are needed in order to formulate a clear definition of sensorimotor and social forms of autonomy.

## PART II. SENSORIMOTOR CONSTITUTION OF NEUROCOGNITIVE AUTONOMY

In Part II we focus on neurocognitive autonomy from a situated and embodied perspective. Analysing current neuroscientific experimental evidence and theoretical developments we depict a formulation of autonomy that is not limited to the internal self-regulation of an agent but also captures the construction of an asymmetry in the agent-environment interaction domain.

### CHAPTER 2. CONSTITUTIVE INTERACTION: THE SITUATED-HKB MODEL

In the first part of Chapter 2 we present an analysis of the role of sensorimotor interaction as a constitutive element of autonomy. We review the recent history of cognitive science and how internalist and ‘computationalist’ perspectives have dominated the second half of the 20th century<sup>2</sup>, and how during the last decades this dominance started to be questioned by dynamical embodied approaches to cognition. We highlight how cognitive science (and cognitive neuroscience in particular) is witnessing an increasing success of dynamical systems models, which are often displacing computational and representational conceptions of cognitive functioning in favour of interactive and dynamical views. We review two central contributions from dynamical systems approaches to cognitive science. First, the impact of the idea that some aspects of cognitive mechanisms (neural or otherwise) are better modelled and understood in terms of dynamical systems instead of symbolic representational algorithms. We review the evolution of dynamical models in cognitive science and we focus especially in advances in oscillatory and coordination dynamics, a field that has witnessed an increased popularity in recent decades. As a second central contribution from the dynamical system approach to cognition, we review situated/embodied perspectives of cognitive science, claiming that agent-environment coupling shapes brain dynamics in a manner that is essential to behavioural or cognitive functionality. We find that in this case some situated/embodied claims have not been adopted with the same emphasis than dynamical systems approaches, and existing theoretical frameworks and models in neuroscience are mostly developed without taking into account sensorimotor dynamics.

---

<sup>2</sup> That of computational functionalism and symbolic Artificial Intelligence, assuming that intelligent behaviour is causally explained by computations of symbolic mental representations

In the second part of the chapter, we analyse the implications of the lack of emphasis in situated/embodied approaches for analysing autonomous behaviour. Concretely, we propose a simple model as a thought experiment on how coordination dynamics under a closed sensorimotor loop might induce qualitatively different patterns of neural oscillations compared to those found in isolated systems. We take as a departure point the Haken-Kelso-Bunz (HKB) model; a generic model of dynamic coordination between two oscillatory components, which has proven useful for a vast range of applications in cognitive science and whose dynamical properties are well understood. In order to explore the properties of this model under closed sensorimotor conditions we present what we call the *situated HKB* model: a robotic model that performs a gradient climbing task and whose ‘brain’ is modelled by the HKB equation. We solve the differential equations that define the agent-environment coupling for increasing values of the agent’s sensitivity (sensor gain), finding different behavioural strategies. These results are compared with two different models: a decoupled HKB with no sensory input and a passively-coupled HKB that is also decoupled but receives a structured input generated by a situated agent. We precisely quantify and qualitatively describe how the properties of the system, when studied in coupled conditions, radically change in a manner that cannot be deduced from the decoupled HKB models alone. We also present the notion of a neurodynamic signature as the dynamic pattern that correlates with a specific behaviour and we show how only a situated agent can display this signature compared to an agent that simply receives the exact same sensory input. Finally, we discuss the implications of our model to understand the role of the environment as a constitutive part of cognitive dynamics in autonomous agents.

### CHAPTER 3. SITUATED AUTONOMY: ORGANIZATIONAL HOMEOSTASIS IN THE SENSORIMOTOR LOOP

In Chapter 3 we begin by reviewing current theories in neuroscience and psychology related to autonomous behaviour: from the organization of neural dynamics in the form of neural assemblies to the information integration theory of consciousness. These theories provide valuable insights for imagining the kind of neural self-organization that make possible the existence of a great diversity of patterns and different specialized functional parts, with the emergence of an integrated and coherent behaviour. Nevertheless, we find that these theories generally lack convincing explanations of how the sensorimotor loop comes into place or have a limited role for it (just modulating brain dynamics, or having a linear input-output relationship with it).

We review different models of autonomous behaviour in synthetic agents, suggesting an innovation in the design of their neurodynamical controllers using relational homeostatic mechanisms in which the variables that are subject to homeostatic regulation are not component variables but relational or organizational variables. As an implementation of this idea, we propose an operational model of plastic oscillatory mechanisms for exploring the role of sensorimotor adaptive behaviour into the emergence and reconfiguration of adaptive behaviour. We evolve a robotic agent with three plastic oscillatory units performing a behavioural pref-

erence task.

We analyse the behaviour of the different groups of variables of the agent under different situations, finding that the coordination of sensorimotor variables, the emergence of oscillatory patterns and the activation of synaptic plasticity in the neural controller is necessary for the generation of the agent's behaviour. Moreover, by analysing the information flows between different levels of description of the system at different timescales using transfer entropy measures, our research shows that autonomous behaviour is sustained through a double regulatory loop: one loop coordinates the activation of plastic mechanisms with the emergence and maintenance of oscillatory phase-locked patterns, and the other loop coordinates the relation of the neural controller and the stream of sensorimotor interactions. These loops are both highly asymmetrical, displaying information flows in one direction at fast scales (from plastic mechanisms to oscillatory patterns and from sensorimotor streams to oscillatory patterns) and the opposite direction in slow scales (from oscillatory patterns to plastic mechanisms and from plastic mechanisms to sensorimotor streams) exhibiting a self-organized system with two embedded emergent levels: one of plastic emergent neurodynamic patterns and the other of brain-body-environment coupling. We finish by proposing that the asymmetry of these two embedded levels might be a fundamental aspect of the interaction domain of autonomous agents and their environment.

#### CHAPTER 4. SITUATED AUTONOMY: COMPLEXITY, METASTABILITY AND SELF-ORGANIZED CRITICALITY

In Chapter 4 we continue the analysis of the agent in Chapter 3 from a different perspective. Whereas in the previous chapter the agent was characterized from the point of view of the information flows of its macroscopic patterns, in this chapter we characterize the interaction among different temporal and spatial scales. Using tools from different complex systems approaches to cognitive science, we investigate informational complexity and long-range temporal correlations in the agent. Previously, some approaches analysing indices of complex dynamics had been criticized for not being able to present mechanistic models explaining how these behavioural dynamics are connected to actual mechanisms in the brain. Here we aim to connect the emergence of different indices of complexity with the mechanisms generating them.

With our model we aim to explore the role of the sensorimotor coupling and synaptic plasticity to connect macroscopic behaviour with neural self-organized dynamics. We proceed by comparing the behaviour of different agents: regular situated agents, 'passively-coupled' agents, agents without synaptic plasticity, and agents that are both passively-coupled and without plasticity. First, we explore the flows of informational complexity in the interactions between different scales of the system components. This is undertaken by measuring TSE complexity (a measure of the levels of simultaneous integration and segregation of information in the system) across different subsystems including the agent's neural oscillators, its synaptic plasticity and its sensorimotor activity. We observe that when agents have synaptic

plasticity, complexity of information flows is enhanced for the sets of components that include sensorimotor variables. In the case of agents without synaptic plasticity, the values of complexity are the same for situated and passively-coupled agents.

Second, we measure long-range temporal correlations in the macroscopic behaviour of the agent in the same conditions above, finding that robust  $1/f$  noise and multifractality are only present simultaneously in internal and sensorimotor variables when the agent is situated in its environment and plastic mechanisms are active. We find that the emergence of  $1/f$  noise at different internal and external variables depends on the cross-relations between internal plasticity and sensorimotor coupling, suggesting that the system as a whole is preserved as a precarious organization dependent of the continuous interactions between its internal and sensorimotor dynamical landscapes.

We hypothesize that indices of TSE complexity and  $1/f$  noise are the result of: (1) non-linear interaction dynamics capable of generating stable collective patterns, (2) internal plastic mechanisms that allow self-sustained criticality of homeostatic, neurodynamic regimes and (3) strong sensorimotor coupling with the environment that induces transient metastable regimes.

Finally, we discuss the implications of our results for understanding the agent-world symmetry breaking and the precarious operational closure necessary for autonomous agency within the domain of sensorimotor interactions. We suggest that the modelling results above allow us to operationalize the notion of a dynamic core as a means to capture the idea of an organizational structure that maximizes both integration and segregation of its components, not in an agent's brain but in the structure of sensorimotor loops the agent is engaged in, allowing the emergence of an autonomous self in the sensorimotor space.

### PART III. INTERACTIVE AUTONOMY IN SOCIAL PROCESSES

Whereas in Part II of this thesis we have explored the constitution of a cognitive agent, its world and the important role of continuous agent-environment interaction, in Part III we change our focus to social autonomy modelling cases where agents interact together in a shared environment. We determine the extent to which the model of neurocognitive interactive autonomy depicted in previous chapters is useful for understanding autonomy and agency in the social realm.

#### CHAPTER 5. SOCIAL AGENCY: INTERACTION-DOMINANT DYNAMICS IN SOCIAL RECIPROCITY

This chapter analyses the constitution of processes of agent-agent interaction as a minimal case from where to analyse the possibility of collective instances of interactive autonomy. We start by reviewing current theories stressing the primacy of embodied interaction over detached social cognition, against traditional emphasis on an individual's mental processes as the basis of social awareness. Further, we review experiments that exploit technological interfaces that have been indispensable for providing support for an interactive approach to social develop-

ment. A minimal interactive setup is proposed for investigating social interaction: the perceptual crossing experiment, which recreates a minimal social environment enabling researchers to systematically investigate the real-time self-organizing dynamics of dyadic interaction by mediating embodied interactions of pairs of adults over a minimal human-computer interface.

We explore whether some critical aspects of social interaction in this minimal setup may not have been observed by previous studies. We consider alternative indicators that could complete, or even lead us to rethink, the current interpretation of the results obtained from both experimental and simulated modelling in the fields of social interactions and minimal perceptual crossing. In particular, we discuss that previous experiments have been analytically constrained to a short-term dynamic type response from the participating agents. Additionally, we propose considering these experiments from a multiscale framework based on the use and analysis of long-range temporal correlations and fractal dynamics. We reveal evidence supporting the idea that social interactions are deployed along many scales of activity and propose that a fractal structure may be better suited to characterize the type of social interaction patterns generated in a social engagement. Finally, the results of the experiment are evaluated in terms of the conceptual framework developed in the previous chapters, generalizing the notion of sensorimotor autonomy to the idea of an extended mode of autonomous organization in a shared social sensorimotor space.

## CHAPTER 6. THE COLLECTIVE MIND: LARGE-SCALE SELF-ORGANIZATION IN MASSIVE SOCIAL NETWORKS

In Chapter 6 we extend the ideas in the previous chapter and deal with the idea of autonomy in large-scale social networks. Specifically, we analyse whether the shift that new ICT and social media has promoted in collective patterns of social and political organization has the capacity to generate powerful forms of social and political autonomy. We analyse recent literature about social organization in the so called ‘networked movements’ (e.g. the Arab Spring, the 15M movement, Occupy Wall Street) and describe how new communication networks and practices have opened the door for radically distinct forms of self-organization based on distributed participation, where centralized hierarchical structures are substituted by a network of interactions that continuously generates, negotiates and maintains a collective identity.

Also, we review how the emergence of the networked movements from 2011 has opened the debate around the way in which new communication technologies and practices make possible innovative forms of collective identities in social movements. We briefly review the literature on ‘collective identities’, and review the tension between different contemporary positions from individual connective action to macroscopic aggregative dynamics. In the midst of this tension, we propose the notion of ‘multitudinous identity’ (a concept based on Hardt & Negri’s 2004 notion of ‘multitude’) to depict the specific forms of identity displayed by networked movements, characterized by a deep social transversality and heterogeneity, as well as a transient and distributed leadership composition driven by action initiatives. In order to char-

acterize the properties of multitudinous identities, we perform a series of analysis connected to the concepts presented in previous chapters.

We analyse the relationship between internal plastic reconfiguration of activation patterns in Twitter networks and moments of embodied action in the form of camps and rallies. We consider Twitter activity during the camps in the initial moments of the 15M movement. First, we use transfer entropy measures to analyse how the system maintains its operational closure while continuously changing its structural configuration. Then, we use phase-locking statistics to analyse how different patterns of synchronization at different scales are connected with different interactions between online communication networks and collective offline behaviour. We depict how synchronization at different frequency bands mediate between massive collective activity at different spatial scales. Furthermore, we characterize, in connection with previous chapters, informational complexity and fractal indices displayed by different moments of mobilization, showing how the presence of multiscale synchronization is coordinated with the emergence of interactive complexity among spatial and temporal scales at moments of strong embodiment of the protests. Finally, we conclude with a reflection about large-scale social autonomy in the framework of new sociotechnical practices, evaluating recent advances and possible future developments.

#### PART IV. CONCLUSION

In Chapter 7 we present an extended summary of the contributions of the different chapters of this thesis. Finally, we present a general view of the significance and scope of these contributions in current research into complex adaptive systems, as well as further research lines that may continue the current work.

## Part I

# Conceptual and Methodological Review



# 1

## An Introduction to Autonomy

**ABSTRACT:** Contemporary notions of autonomy (as a form of organization that defines the principles of operation of a system while constituting its identity) rely on a long history of theoretical and scientific approximation to the holistic properties of organisms. We make a historical and systematic approach to the concepts that shape contemporary theories of autonomy (persistence, adaptivity, self-organization) together with models able to operationalize some of these concepts at different levels (self-synchronization, metastability, self-organized criticality). Finally, we review how the idea of autonomy might operate at different scales of living activity, and propose a definition of autonomy that will shape the work presented in this thesis.

**CONTRIBUTIONS:**

- We perform a historical review of notions related to the notion of autonomy, both conceptual and mechanistic, especially from the cybernetic movement to modern views from complex systems and neuroscience.
- We provide a first tentative definition of autonomy upon which we will build in the next chapters.

Autonomy is nowadays seen as a pervasive property of living beings (from unicellular and organismic to neurocognitive and social life; Rosslénbroich, 2014; Bich and Damiano, 2012; Ruiz-Mirazo and Moreno, 2011). Autonomous beings are as those able to display forms of coherence and identity in which all their components (metabolites, cells, living tissue, organisms, communities, etc.) are coordinated as a unitary system displaying a robust constitution against external perturbations and, at the same time, are able to flexibly adapt to different forms of interaction with their surroundings in which they actively stand out as a differentiated entity with respect to them. How does this happen? What kind of structures and organization support these forms of being and acting? What differentiates autonomous living beings from massive computational systems or from complex physical phenomena like for example, tornadoes or earthquakes?

Somewhat similar to any modern computer, living systems coordinate the activity of a myriad of heterogeneous parallel processes. Our bodies have countless parallel regulatory loops coordinating all kinds of metabolic, organismic, neural processes. Nevertheless, unlike any computer, all these processes do not run independently but there is a coordinated macroscopic driving force behind them. Like in a tornado, a macroscopic process emerges becoming a reference point for the activity of the microscopic processes that constitute it. Furthermore, in the case of living entities we attribute certain properties to this emergent self – like intentionality, free will, or even consciousness – that actively modulates the interaction with its environment, allowing us to characterize such entities as autonomous.

Modern formulations of the concept of autonomy are the result of a long history of conceptual development and mathematical contributions that made possible to operationalize and sharpen more basic and relational notions required to shed light about how autonomous systems arise, maintain themselves and adapt to their environment, as well as what relationships are established between the different scales and domains of existence of the system. These organizational principles should not only help to explain how a system can operate as an integrated, coherent and robust unit, and at the same time display complex, ever-changing, adaptive behavioural patterns. The goal is to conceptualize certain forms of organization that allow organisms to maintain their identity (as a form of coherence in their own organization) against a complex and changing environment. We present a historical review of theoretical notions that precede current views of autonomy, together with complex systems descriptions aiming to operationalize some of these notions and make them plausible in terms of formal models.

In the first part of this chapter we study concepts that have led to a modern notion of autonomy: persistence, identity and self-organization. These concepts try to capture constitutive properties of living beings and eventually have evolved into an operational definition of autonomy. We follow a historical conceptual thread from (1) notions of persistence where the robustness of the system is imprinted in its own structure (e.g. homeostatic systems, negative feedback loops), (2) systems whose persistence resides in the self-regulation of its components in a region of viability (e.g. Ashby's Ultrastable System), to (3) systems where its persistence resides in the constitution and maintenance of its own organization in interaction with its environment (e.g. autopoietic systems). In the second part, we analyse the development of operational models of forms of organizational dynamics capable of sustaining modes of coherence or coordination proper of autonomous systems. We start with (1) modes of coherence that arise as the consequence of some structural coordinative property (e.g. spontaneous coordination), following with (2) modes of flexible coordination where the system is able to enter and exit regions of transient attraction (metastable and soft-assembled systems) and finish by conceptualizing systems where (3) transient coordination is extended to the different scales comprising the structure of the system (e.g. systems displaying self-organized criticality). Finally, we provide a provisional definition of autonomy based on the notions of individuality, interactive asymmetry and normativity, and analyse its domains of application at different

scales of life.

## I.I PERSISTENCE AND IDENTITY IN LIVING BEINGS

The first question that a study of autonomous systems faces concerns identity and how it can be preserved and adapted over time. By identity we will refer to the property allowing something to be, in a loose sense, ‘nearly identical to itself’, i.e. maintaining a set of invariances despite being subject to transformations, independently of external influences. Living beings are able to maintain its identity despite internal transformations or environmental perturbations, presenting certain type of persistence in their being. How does this identity emerge and how it is maintained against external or internal fluctuations? As we will show, modern notions of autonomous are rooted in the answers to the problem of identity and persistence.

### I.I.I PERSISTENCE AND STABILITY

One of the simplest forms of identity involves mechanisms that allow an agent to maintain its current state or its current existence as it is. One of the basic properties of living beings is to present certain type of persistence in their being. This idea was already a central theme in Spinoza’s philosophy with the idea of *conatus*.

#### CONATUS

According to Spinoza, ‘each thing, as far as it can by its own power, strives to persevere in its being’ (Spinoza, 1883, part 3, prop. 6. Originally published in 1677). Moreover, this ‘striving to persevere’ is not some action that an agent performs in addition to its other activities, instead, it is ‘nothing but the actual essence of the thing’ (Spinoza, 1883, part 3, prop. 7). The concept of conatus had been used before by Descartes in a more mechanistic sense to describe the tendency of bodies to move when they collide with each other (Descartes, 1984, Part 2, art. 37-40. Originally published in 1644). These tendencies were described in terms of centripetal (*conatus a centro*) or centrifugal (*conatus recedendi*) forces, and as a generalization of the principle of inertia (*conatus se movendi*, or ‘conatus of self-preservation’), which were already developed and experimentally demonstrated by Galileo, and were formalized by Newton after the death of Descartes. Spinoza applies the idea of conatus to all things as a resistance to self-destruction. In particular he applies it to the human body and mind for describing its will and appetites. As opposed to most philosophers of his time, Spinoza rejected a dualistic separation between body and mind. Consequently, for Spinoza, the substrate of the conatus is material and subject to the laws of the universe, which are inescapable. Spinoza claims that the principle of conatus explains the ‘self-evident’ truth that ‘nothing can be destroyed except by an external cause’ (Spinoza, 1883, part 3, prop. 4). For Spinoza, this resistance to self-destruction is materialized in terms of a striving to continue to exist. Consequently, the conatus acts as a force or a loose principle of inertia related to an agent’s current being, in confrontation with the possibility of self-destruction. A hungry animal strives for food and a thirsty one for water. This way,

‘striving for self-preservation’ resolves an organism’s internal tension (threatening the continuation of its existence) by creating an interactional domain (tendencies and inertias) leading to the satisfaction of such tensions. So, although Spinoza somewhat conflates *explanandum* and *explanans* (how the organisms makes possible its own persistence remains unresolved), in contrast to previous dualist conceptions of mind and body, the notion of conatus proposes an interactional mind/brain continuum. Spinoza put the persistence and self-preservation of living systems at the core of philosophical theorizing.

## HOMEOSTASIS

Although the concept of conatus influenced other concepts, for example, Schopenhauer *Will to Live*, Nietzsche’s *Will to Power* or Henry Bergson’s concept of *élan vital* or vital impulse (a concept which implies a fundamental driving force behind all life), here we want to focus on a different line of thought that instead of postulating abstract forces, tries to bring these principles down to the biological mechanisms underlying those impulses of living beings to self-preserve. A fundamental step in the search of these concepts was the development of the concept of homeostasis.

A precursor of the concept of homeostasis was developed by Claude Bernard, who, by 1860, had proposed the concept of a ‘milieu intérieur’ as the environment where biological processes occur. The constancy of this milieu intérieur was a prerequisite to preserve physiological functions within an optimal range. For Bernard ‘[t]he stability of the internal environment [the milieu intérieur] is the condition for the free and independent life’ (Bernard, 1974). This idea would be later developed by Walter Cannon who, in 1926, coined the term *homeostasis* for describing the ability of living beings to maintain their own stability. Cannon described living beings as open systems in relation to their environment. Changes in their environment, he proposed, excite reactions in the organism’s system provoking internal disturbances, that are automatically maintained within some narrow limits keeping them constant, i.e., at *equilibrium* (Cannon, 1929, pp.400-401). The coordinated physiological reactions that maintain the steady states conforming this equilibrium are the mechanisms required for homeostasis. States that are kept constant include the material supplies for cellular needs (water, sodium chloride, calcium, oxygen, etc.) and environmental factors affecting cellular activity (osmotic pressure, temperature, ion concentration, etc.). Also, cross-relations between homeostatic states exist (e.g. osmotic pressure in body fluids which is dependent on the proportions of water, salts and proteins).

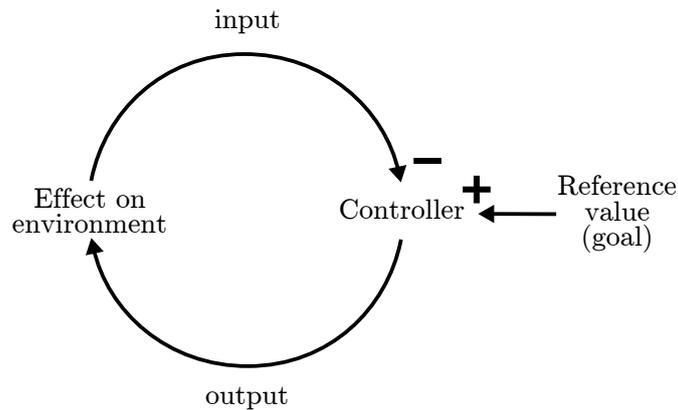
However, Cannon did not stop at the cellular level, and explained homeostatic functions of states such as hunger and thirst (Cannon, 1929, p.417) in terms of previous experiences with food and drink appetites inviting the renewal of those experiences. If the needs for food or water are not met, thirst and hunger appear as physiological responses to propel the organism to eat or drink and bring its state to equilibrium. The big question here is what is the relation between homeostatic regulation at the level of the whole animal and physiological lower level of homeostasis? How does homeostasis scale up and how is it implemented?

## CYBERNETICS FEEDBACK MECHANISMS

In the early 1940s, Norbert Wiener started a fruitful collaboration with Julian Bigelow and Arturo Rosenblueth that led to some fundamental ideas of cybernetics. Interestingly, Rosenblueth had been a former collaborator of Cannon. Cannon had also been a friend of Wiener's father and an important influence for Wiener. This collaboration started during World War II, within the efforts of the American National Defense Research Committee (which employed Wiener and Bigelow) to discover a better way to direct and control the firing of antiaircraft artillery. The problem was, at the time, solved manually by crews tracking the position of planes using binoculars, human computers calculating their projected location and gunners rotating and aiming the antiaircraft turrets. Wiener and Bigelow's goal was to develop an automatic radar-guided antiaircraft tracking device using new British microwave radars. Although their contributions ultimately failed to produce a practical device to be used in war, they developed an ingenious new approach to antiaircraft fire prediction, using new statistical methods of extrapolation, interpolation, and filtering, that were published by (Wiener, 1949, written in 1942) providing some basis of the cybernetic revolution some years later. Wiener's report redefined the whole concept of control, which had been traditionally associated with power engineering, by defining it as a part of communication engineering, putting the transmission of messages as the central element – understanding that a telephone conversation is as much of a message as the signals controlling a motor. It is important to recall that, at the time, the concept of information, so widespread and ubiquitous today, represented a great change of perspective. Wiener stated that communication operations are essentially no different from computational operations, adding statistical computation of time series as a central element of his approach. In this scenario, the time series of a message is not considered by itself, but as part of a 'ensemble of time series' judged 'not by their effect in a particular case, but by their average effect'.

Inspired by their collaboration in antiaircraft fire prediction and control, Wiener, Bigelow and Rosenblueth plunged into addressing physiological and neurological feedback, extending the 'cybernetic vision' to biological, physiological, and social systems. The contact of Wiener with what he later called 'the tragic insolence of the military mind' (Wiener, 1988, p.xxvii) ended in 1943, driving him into putting a great effort into elaborating cybernetics as a civilian discipline and bringing his ideas to broad communities of physiologists, physicians, and social scientists (Heims, 1993). Soon afterwards, Rosenblueth, Bigelow and Wiener coauthored one of the founding papers of cybernetics, entitled 'Behavior, Purpose and Teleology' (Rosenblueth et al., 1943). The paper studies the way that mechanical, biological, and electronic systems can generate purposeful behaviour, and states that intrinsic, purposeful behaviour requires negative feedback loops. These ideas instigated the creation of the Teleological Society and later the Macy conferences, which, in the words of the members of the Teleological Society, popularized 'the study of purpose [...] and how purpose can be imitated by mechanical and electrical means' (cited in Aspray, 1990, 315).

Wiener, Bigelow and Rosenblueth contributed with both a general description of how



**Figure 1.1: Negative feedback.** Organization of a system controlled by a negative feedback loop in interaction with its environment.

mechanical or biological systems can generate persistent or homeostatic-like behaviour – via negative feedback loops (Figure 1.1) – and introduced the concept of intrinsic purpose or teleological behaviour as a more general description for this kind of processes. Wiener proposed that the concept of homeostasis is implemented in a great number of cases in biological systems in which negative feedback loops are not only present in physiological phenomena but are ‘absolutely essential for the continuation of life’ (Wiener, 1948, p.114).

Along with many other brilliant cyberneticists, the work of Wiener, Bigelow and Rosenblueth propelled a progressive scientific and technological research program which would shape the basis of the research on intelligence and the mind in the second half of the 20th century. This research program aims at describing mental activity in terms of the operation of purely physical mechanisms. Cognitive activity is therefore described in terms of feedback mechanisms, algorithms and nonlinear dynamical systems, which allow machines to be built displaying ‘purposeful’ behaviour. In this context, purposeful behaviour generally denoted that which ‘may be interpreted as directed to the attainment of a goal’ (Rosenblueth et al., 1943). A paradigmatic example is a negative feedback loop as a mechanism producing goal-directed behaviour, where there is no difference if it is embedded in a human being or a mechanical system. Although Rosenblueth et al. (1943) acknowledge that the statement *may be interpreted* (by an external observer) is tremendously vague and operationally meaningless, they argue that the concept of purpose is useful and should be retained<sup>1</sup>. This loose definition of purpose, equating purpose and the mechanisms producing apparently purposeful behaviour, was shared by other prominent cyberneticists. For example, McCulloch and Pitts, other founding fathers of the cybernetic revolution, wrote a paper which became a manifesto for both symbolic and connectionist AI equating intentionality with propositional logic im-

<sup>1</sup>For example, they argue that the basis of the concept of purpose is the awareness of voluntary activity: when we perform an action we do not command certain muscles to contract in a certain sequence; we merely trip the purpose and the reaction follows automatically.

plemented by a neural network (McCulloch and Pitts, 1943). In the classical cybernetic framework, it seems that intelligence is reduced to physical processes creating a mere ‘as-if’ appearance of a subject acting for authentic reasons. This tension was evaded by assuming that intelligence is located in these physical mechanisms (e.g. feedback mechanisms) themselves (Froese, 2010).

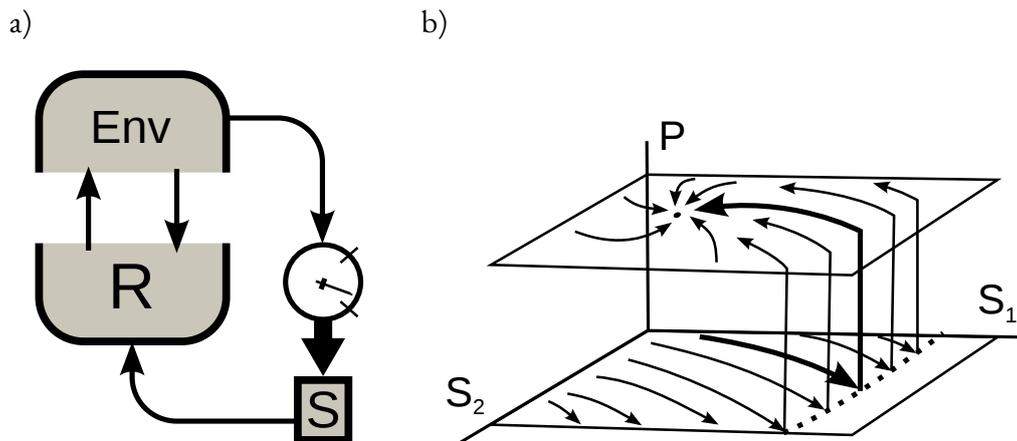
### 1.1.2 ADAPTIVE BEHAVIOUR

#### ULTRASTABILITY

In 1946 the British cyberneticist Ross W. Ashby designed what he described as an ‘isomorphism making machine’ to develop his concept of an ultrastable system (Ashby, 1946). Along with many of his cyberneticist colleagues, he believed that the concept of survival (and therefore being alive) could be captured by the generic systemic notion of stability. However, Ashby went further tackling the issue of adaptive behaviour beyond the mere regulation of feedback loops. For Ashby, ‘a form of behaviour is adaptive if it maintains the essential variables within physiological limits’ (Ashby, 1954, p.58). As a model of adaptive behaviour, he developed the notion of an ultrastable system, which he described as:

Two systems of continuous variables (that we called ‘environment’ and ‘reacting part’) interact, so that a primary feedback (through the complex sensory and motor channels) exists between them. Another feedback, working intermittently and at a much slower order of speed, goes from the environment to certain continuous variables which in their turn affect some step-mechanisms, the effect being that the step-mechanisms change value when and only when these variables pass outside given limits. The step-mechanisms affect the reacting part; by acting as parameters to it they determine how it shall react to the environment (Ashby, 1954, p.98).

As shown in Figure 1.2.a, a reactive system (R) in continuous interaction with its environment (Env), could maintain under control certain essential variables (S) by triggering random changes in the configuration of R when S crosses certain bounds. Ashby described S as the essential variables of an organism which should be kept under control for its correct functioning (e.g. blood temperature, pH). For Ashby, the mechanism for keeping those variables under control was a step-mechanism triggering random internal parametrical changes which would eventually return these essential variables to their viability zone. If these changes create new dynamics which are able to return S to the established bounds, the organism would have adapted to the perturbation that provoked S to get out of bounds in the first place (see Figure 1.2.b) According to Ashby, adaptation processes of an organism in an environment are based on the existence of feedback loops at two different levels which define the ultrastable system: (1) The continuous interaction between an agent and its environment, and (2) the



**Figure 1.2: Ultrastability.** a) Diagram of an ultrastable system, where Env is the environment, R is an organism's reacting system and S are the organism essential variables, triggering changes in R when they are out of bounds. b) Example of adaptive behaviour in the ultrastable system, where  $S_1$  and  $S_2$  represent the phase space of the essential variables, and P the parametrical configuration of the agent. Both schematics were adapted from Ashby (1954).

regulation of the essential variables of the organism through plastic changes in the agent's internal mechanisms. In parallel, some similar ideas were being proposed by Jean Piaget in the field of psychology<sup>2</sup>.

Ashby materialized its design of an ultrastable system in 1948 building the *Homeostat*, an adaptive ultrastable machine consisting of four interconnected Royal Air Force bomb control units with inputs, feedback, and magnetically driven, water-filled potentiometers (see Ashby, 1954, chapter 7). Each unit emits as an output a direct current proportional to the deviation of its magnet from the central position. These output currents act as inputs for other units influencing their magnets through a coil. Before each input current reaches its coil, it passes through a commutator (determining the polarity of entry to the coil) and a potentiometer (which determines what fraction of the input current reaches the coil). Commutators and potentiometers act as parameters of the system. The values of the parameters are provided by a random number generator each time a magnet diverges far from the central position. Ashby's machine displays behaviours such as habituation, reinforcement or learning through its ability to automatically adapt its configuration to stabilize the effects of any disturbances from its environment. A schema of the organization of Ashby's implementation of an ultra-

<sup>2</sup>Piaget published 'The Psychology of Intelligence', trying to explain the genesis of higher human capacities in terms of the ability to organize and re-organize the interaction with the environment in growing forms of complexity. Thus, Piaget proposes a framework to understand biology and psychology in which the focus is in the self-regulation of the development of structures that increase their degrees of (dynamic) equilibrium. This process is mediated by two complementary processes (similar to Ashby's double feedback loop): (1) assimilation, or the modification of incoming input according to pre-existing structures, and (2) accommodation, the active modification of the structure itself as a response to environmental perturbations.

stable system in the Homeostat is depicted in Figure 1.3, in which local components have an extra layer of regulation.

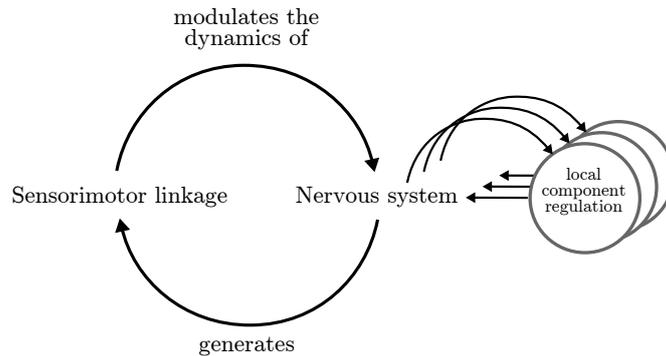
The concept of ultrastability and its materialization into the Homeostat challenged some ideas of the cybernetic framework and opened new perspectives on persistence and purpose in living beings (Froese, 2010). Intentionality is no longer assumed to be contained within the structure of the mechanisms of the system (feedback loops or logical functions). Instead, it is the regulation of some states of the system which makes the rest of the system act intelligently. As Wiener recognized,

Ashby's brilliant idea of the unpurposeful random mechanism which seeks for its own purpose through a process of learning is [...] one of the great philosophical contributions of the present day [...] Not only can we build purpose into machines, but in an overwhelming majority of cases a machine designed to avoid certain pitfalls of breakdown will look for purposes which it can fulfill (Wiener, 1988, p.38. Originally written in 1950).

Despite Wiener's optimistic view, the key contribution of Ashby's work was to show that behaviour that may appear intelligent to an external observer can potentially be accounted for by quite simple and unintelligent mechanisms driven by random step functions, as long as they were connected with the maintenance of the right viability regions. Thus, Ashby's model challenges reductionistic views assuming that purpose is embedded in feedback loops since they contain a representation of a goal-directed behaviour. Instead, one may explain the functioning of the Homeostat without the need of representation at all, since it does not represent the world in any formal sense. It can learn to adjust its parameters for interacting with the world to adapt to various disturbances and maintain certain desirable internal states. No wonder that Ashby's ideas were subject to fierce criticism by some cyberneticists present at the ninth Macy Conference (the only one in which Ashby took part). Wiesner, Pitts, and particularly Bigelow sustained critical pressure, questioning (among other aspects as the use of randomness in his model) whether the Homeostat *learns* when it finds an equilibrium state (Dupuy, 2009, p.148-155; Husbands and Holland, 2012, p.10-11). Ashby's design served no purpose and was designed to carry out no task but to maintain certain internal invariances, somehow plunging into a crisis part of the previous cybernetic hypothesis, and influencing the development of a second wave of cybernetic research that would propose new perspectives (Glanville, 2002; Froese, 2010).

#### SELF-ORGANIZATION

Ashby's work allows an interesting perspective about the self-organization of a system. An ultrastable system is to display what appears to be purposeful behaviour without a previous design specifying how to perform this behaviour. The Homeostat was not designed to perform any particular behaviour but it was able to modify its own organization to perform a



**Figure 1.3: Ultrastable regulation of the system components.** Organization of an ultrastable system as Ashby implemented it in his Homeostat. A layer of local regulation is added on top of the neural controller, in which local units trigger changes in its connectivity when their state is not maintained under some bounds.

variety of behaviours while maintaining certain variables under control. Although it can be argued whether Ashby Homeostat can account for autonomous behaviour, his model poses the idea that purpose of an agent does not lie in its mechanisms but in the interactions between its component that allow rearranging the organization of the agent. Ashby tried to coin some of these ideas under the term of ‘self-organization’, which he introduced into contemporary science (Ashby, 1962, originally written in 1947), to be further taken-up by other cyberneticists for example, Heinz von Foerster, who organized a conference on ‘The Principles of Self-Organization’ at the University of Illinois’ Allerton Park in June, 1960, leading to a series of conferences on Self-Organizing Systems. Norbert Wiener also took-up the idea in the second edition of his ‘Cybernetics: or Control and Communication in the Animal and the Machine’ (MIT Press 1961).

Self-organization is typically defined as a process where some form of overall order or coordination arises out of the local interactions between the system components that may be initially disordered. Self-organization is spontaneous, i.e. it is not controlled by any specific part of the system or any external system but emerges from the distributed interaction between the components of the system. Typically, self-organized systems are robust to external fluctuations and able to survive and sustain themselves in the face of perturbations from outside the system, at least to some extent. As homeostasis is usually modelled in terms of negative feedback loops correcting deviations of the system, self-organization is usually modelled in terms of systems with positive feedback loops that amplify random fluctuations in the system (Heylighen, 2001).

The term ‘self-organizing’ was originally introduced by Immanuel Kant in his Critique of Judgement, where he argued that teleology is a meaningful concept only if there exists such an entity whose parts or ‘organs’ are simultaneously ends and means:

For a body, therefore, which is to be judged as a natural end in itself and in

accordance with its internal possibility, it is required that its parts reciprocally produce each other, as far as both their form and maintain their combination is concerned, and thus produce a whole out of their own causality [...] In such a product of nature each part is conceived as if it exists only through all the others, thus as if existing for the sake of the others and on account of the whole [...] as an organ that produces the other parts (consequently each produces the others reciprocally) [...] only then and on that account can such a product, as an organized and self-organizing being, be called a natural end [...] An organized being is thus not a mere machine, for that has only a motive power, while the organized being possesses in itself a formative power, and indeed one that it communicates to the matter, which does not have it (it organizes the latter): thus it has a self-propagating formative power, which cannot be explained through the capacity for movement alone (that is, mechanism) (Kant, 1987, p.245-246).

Interestingly, when Kant defines self-organization he describes a particular kind of interaction between the parts and the whole. This interaction addresses a formative self-propagating organization communicated by the whole to the parts, though it is absent in the parts alone.

Kant uses the example of a watch, in which there is one part that makes other parts move, but it is not the cause that produces the others. As opposed to machines, organized beings not only have *motive* forces (the ability of one thing to move another) but a *formative* force that propagates itself producing patterns organizing the components of the system. These patterns would also include deviations to cope with different circumstances, guaranteeing the self-preservation of the organized being (Kant, 1987, p.374-375). For Kant, organisms are systems in which parts are formed for and from the others, constituting a dynamical unit that guides the activity of the organism as a whole. This is perhaps the first elaboration of the emergence of a self, which arises from the recursive processes that constitute them, which will be one of the guiding notions of the work presented here. This notion of self-organization was very influential on the organicist school of thinking that expanded from German idealism to some biological and ethological interpretations, but it was not until Ashby that it came to be reformulated in scientific formal terms.

Ashby gives two possible descriptions of self-organization (Ashby, 1962). A first type would be purely structural, producing coordination between the parts of the system and presenting certain resilience to perturbation (Ashby characterizes these systems as 'self-connecting', since they explain the transition from unorganized to organized states). The second meaning addresses what Ashby defines as 'changing from a bad organization to a good one', and he exemplifies it with the example of a child whose brain organization makes him fire-seeking, but that after being burnt becomes fire-avoiding (Ashby, 1962, p.115). In the latter normative sense, Ashby argues that no machine can be described to be self-organizing by itself. If we define  $S$  as the set of states of the machine, and  $f$  as the driving function that maps states of  $S$  into  $S$ , a self-organizing machine will be such where  $f$  is a function of  $S$ . This change would just convert the first machine into a new machine. For Ashby, the only way to conceive a self-

organizing system without being 'self-contradictory' is to conceive a machine which is coupled to an environment  $\alpha$  (Ashby, 1962, p.296). Thus, if changes in  $f$  are originally generated by changes in  $\alpha$ , we have a machine  $S$  that is self-organizing within the whole  $S + \alpha$ . That is, according to Ashby a self-organizing system can only be so by being coupled to another system. In this way, Ashby conceives self-organization and closure as a process taking place in an agent-environment interactive domain.

Although greatly inspiring, Ashby's description and implementation of adaptive behaviour and self-organization in terms of the stability of a system falls short of characterizing a truly autonomous behaviour. For example, Ashby's Homeostat would stay indefinitely trapped in a particular stable state if it is not perturbed by its environment. Oddly, in the implementation of the ultrastable system through the Homeostat, the essential variables of the system are collapsed into the components of the system that interact with the environment, so the viability region of the system is defined only in terms only of the stability of individual components. This description of the system reduced to just one scale of description limits the possibilities of Ashby's powerful idea of ultrastability. Nevertheless, Ashby's work inspired conceptual developments that would be of great importance for thinking autonomous behaviour. Moreover, the concept of self-organization would later become a common place in scientific literature with the rise of complex systems science and the development of non-equilibrium thermodynamics. As we will see later, the study of far-from-equilibrium physical systems and their role in life by authors as Prigogine or Kauffman will expand the notion of self-organization to richer possibilities.

### 1.1.3 IDENTITY, CLOSURE, SELF-REFERENCE AND CIRCULARITY

Although the concept of homeostasis has been a milestone inspiring dynamical characterizations of systems stability, the idea soon provoked calls to go further than the mere tendency toward equilibrium. One of the defendants of such position was Ludwig von Bertalanffy, one of the founders of general systems theory, an interdisciplinary discipline that tries to find 'principles which apply to systems in general, whatever the nature of their component elements or the relations or "forces" between them' (Bertalanffy, 1950). Bertalanffy argued that the equilibrium principle of homeostasis 'misses the fact that psychological and behavioral activities are more than relaxation of tensions' (Bertalanffy, 1969, p.108), and claims that homeostasis cannot explain certain aspects of life (e.g. play, exploratory activity, creativity or self-realization; but also many aspects of human culture), and some disturbances of living beings for example, psychosis-like episodes in sensory-deprivation experiments.

### EQUIFINALITY

To solve this problem, Bertalanffy proposes that it is fundamental to understand the difference between closed and open systems and think about organisms in terms of the later. Biological systems, which are in a perpetual state of change of their components (e.g. chemical compounds, cells, but also organisms, or species in ecological systems) constitute an example

of open systems. For Bertalanffy, a major difference between closed system and open system is equifinality, i.e., a property of systems whereby the final state may be reached from different initial conditions and in different ways. In closed systems, equifinality cannot be found since (according to the second law of thermodynamics) they always tend to a state of maximum entropy, i.e. a time-independent state of equilibrium where the composition remains constant. Conversely, open systems, which continuously exchange matter with their environment, under some constraints, can attain a stationary state in which through the inflow and outflow of materials the system reaches a *steady state*. Thus, under the right set of constraints, a system may reach the same steady state independently of the initial configuration of its elements. An example used by Bertalanffy is the development of sea urchin larvae (Bertalanffy, 1950). Normally, a larva develops from a normal germ of the sea urchin, but if the germ is cut in half, or if we fuse two germs into one, a similar larva will develop from the resulting (half or double) germ. A similar process takes place in the development of twins in humans and other species. For Bertalanffy, this is an example of equifinality, systems with completely different starting conditions will result in the same steady state (exemplified by a living larva; Bertalanffy, 1950). Bertalanffy explains that these phenomena cannot be explained in terms of the microlevel physico-chemical forces, but in terms of the properties that set up the equilibrium steady state, depending on the constants of reactions and the inflows/outflows of matter in the system. That is, equilibrium does not depend on the stability of the components of the system alone, but in the stability of the system as a whole.

Bertalanffy discusses the differences between inert and living systems in terms of the kind of *finality* in those systems (Bertalanffy, 1950). Specifically, he characterized certain modes of finality as *dynamic teleologies*, stressing the difference between:

1. Directiveness based upon structure: where the structure of the system itself leads the process in a certain way. This may be said of human-made machines aiming some performance, but also of certain biological processes of homeostasis.
2. Equifinality: where regulations of the system cannot be based upon predetermined structures or mechanisms, but they are based in the configuration of steady states in open systems and determined by phenomena of production of order in far-from-equilibrium thermodynamics.

For the second types of processes there is a special relation between the components and the whole of a system: that of circular causation.

#### CIRCULARITY AND OPERATIONAL CLOSURE

Apart from notions of ultrastability, one of Ashby's most influential contributions was the idea of closure applied to adaptive organisms. According to Ashby, the ultrastable system solves the problem of how an organism changes from the unadapted to the adapted condition, but this is only a first step towards understanding living entities. The exigences of adaptivity

in real world situations demands time and efficiency of the adaptation (Ashby, 1954, chapter 10). Here, the concept of closure is fundamental to understand the persistence or the survival of an organism. Ashby defines closure as:

When an operator acts on a set of operands it may happen that the set of transforms obtained contains no element that is not already present in the set of operands, i.e. the transformation creates no new element [...]. When this occurs, the set of operands is closed under the transformation (Ashby, 1962, p.11).

Ashby uses the example in which we can define a 'living mouse' as being a mouse in a state contained in a subset  $\{M_i\}_{i=1..N}$ . We can say that the mouse survives an operation  $C$  if the result of all possible  $C(M_i)$  is also a state contained in  $\{M_i\}$ . If we were to describe the states in  $\{M_i\}$  in terms of variables, these states would correspond to those in which certain essential variables are kept within the assigned homeostatic limits (Ashby, 1962, p.197). The property of operational closure guarantees that the identity (organization) of the living system is well defined and that it will persist throughout its operation.

This idea of closure in living beings was further developed by Humberto Maturana:

Living systems as they exist on earth today are characterized by exergonic metabolism, growth and internal molecular reproduction, all organized in a closed causal circular process that allows for evolutionary change in the way the circularity is maintained, but not for the loss of the circularity itself [...]. This circular organization constitutes a homeostatic system whose function is to produce and maintain this very same circular organization by determining that the components that specify it be those whose synthesis or maintenance it secures. Furthermore, this circular organization defines a living system as a unit of interactions and is essential for its maintenance as a unit; that which is not in it is external to it or does not exist. The circular organization in which the components that specify it are those whose synthesis or maintenance it secures in a manner such that the product of their functioning is the same functioning organization that produces them, is the living organization (Maturana, 1970, p.9).

Maturana reinterprets the concept of ultrastability (Froese and Stewart, 2010) to develop what he called the circular organization of the living, defining a living system as 'an homeostatic system whose homeostatic organization has its own organization as the variable that it maintains constant through the production and functioning of the components that specify it, and is defined as a unit of interactions by this very organization' (Maturana, 1970, p.48). Thus, living beings would be a kind of system that not only maintains a stable organization but also produces it. This organization defines the organism as a unit of interactions occurring in a cyclic manner. If those cycles are interrupted, the organism disintegrates (Maturana, 1970, p.10).

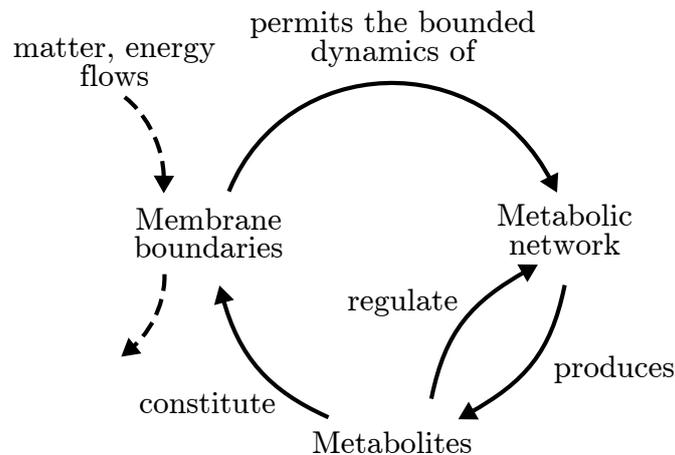
## AUTOPOIESIS AND AUTONOMY

The circularity of the living organization has been reformulated by Maturana and Varela's concept of autopoiesis to define the type of organization which constitutes living beings, steering away from a static conception of cybernetic feedback mechanisms (Maturana and Varela, 1980). They extend the previous definition of living beings stating that autopoietic organization is a sufficient condition for life. An autopoietic system is defined as a network of processes of production (transformation and destruction) of components which (1) through their interactions and transformations continuously realize and regenerate the network of processes (relations) that produced them; and (2) constitute the system as a concrete unity in the space in which they (the components) exist by specifying the topological domain of its realization as such a network (Maturana and Varela, 1980, p.78-79).

The paradigmatic example of autopoietic organization is cellular metabolism, where a network of chemical reactions produces molecules that '(i) through their interactions generate and participate recursively in the same network of relations which produced them, and (ii) realize the cell as a material unity' (Varela et al., 1974, p.188). The first condition acknowledges the importance of the material self-production of living beings, addressing the importance of materiality and embodiment within an environment. The second one rejects the static conception of living homeostatic systems, where the systems components are maintained constantly, favouring a view where what the system maintains constantly are 'certain relations between components otherwise in continuous flow or change' (Maturana and Varela, 1980, p.81). The organization of the system has to be captured by relational variables between the different dynamic processes that constitute the system. These relational variables are the ones that need to be regulated through homeostatic-like mechanisms in order to ensure the survival of the organism. A schematic of cellular autopoiesis is displayed in Figure 1.4.

Varela's idea of autopoiesis was modelled in a computer simulation of a two-dimensional cellular automaton (Varela et al., 1974) where substrate and catalyst elements were defined as moving randomly in the space of a quadratic grid, producing 'link' elements capable of bonding when they interact. The model is able to simulate a network of productions of components that effectively produces a boundary or membrane that distinguishes the system from the environment, being able to re-establish the boundary when it is broken by spontaneous decay of links.

For Varela, autopoiesis characterizes an instance of a form of organization endowing living system with a broader property: autonomy (Varela, 1981). Here, autonomy is defined as a general phenomenon applicable to other domains of interaction, such as neural or immune systems; whereas autopoiesis is only a particular case that applies to the mechanisms that produce the unitary character of living organisms in the physical space. Although the relation between autonomy and autopoiesis is not always clear in Maturana and Varela's literature, it is usually connected with the notion of closure and the affirmation that all autonomous systems are operationally closed. In Varela's words, autonomous systems are operationally closed implying that 'their organization is characterized by processes such that (i) the processes are



**Figure 1.4: Autopoietic organization.** A circular network of metabolic processes produces a membrane that encloses the system in space, regulating its interactions with the environment. Adapted from Varela (1997)

related as a network, so that they recursively depend on each other in the generation and realization of the processes themselves, and (ii) they constitute the system as a unity recognizable in the space (domain) in which the processes exist' (Varela, 1979, p.55). A good example of how autonomy may be extended to other domains is the work of Francisco Varela and colleagues conceptualizing and modelling autonomy in immune systems (Varela and Coutinho, 1991), where the autonomy of the immune system is characterized as fluctuation patterns of natural antibodies able to define an organism's molecular and cellular identity (and therefore aspects such as tolerance and self-nonsel self discrimination, memory and the evolution of the immune system). Another example is cognitive behaviour, which is the result of the autonomy at the level of the nervous system. The dynamic organization of neural networks defines a series of invariant patterns of sensorimotor correlations that define the organism as a mobile unit in space (Varela, 1997).

#### SELF-ORGANIZATION AND AUTONOMY IN FAR-FROM-EQUILIBRIUM PHYSICAL SYSTEMS

As Bertalanffy pointed out, circular organization is closely related to far-from equilibrium physical structures. Although Maturana and Varela sought to develop a general system's description of an autopoietic system, many researches have urged taking into account the physical dissipative structures that may sustain such processes (Moreno et al., 1994). The theory of dissipative structures was developed by Ilya Prigogine, with strong implications for the study of biological processes. For Prigogine, biological problems such as the origin of life are as much connected to physics as to biology. For example, he postulates non-equilibrium thermodynamics as a solution for the apparent contradiction between biological evolution, in which an irreversible increase of organization of more complex structures takes place, and the second law of thermodynamics, which postulates that the evolution of every system is directed

towards continuous disorganization (Prigogine and Nicolis, 1971).

Systems in thermodynamic equilibrium always tend to the 'most probable' state, corresponding to the state of maximum disorder. According to Boltzmann principles Boltzmann (1995), the resulting disorder is the product of a competition between energy and temperature at a molecular level, whose results are directly translated to the macroscopic level. Take for example the ferromagnetic Ising model (which will be described in section 1.2.3), in which the temperature acts as a control parameter in a one-to-one mapping between molecular and macroscopic behaviour. For system's in thermodynamic equilibrium, order is only possible at low temperatures. When the temperature raises energy flows fairly rapidly among the spin states of interacting atoms, which makes the creation of ordered structures very difficult. However, Prigogine described in depth a different kind of thermodynamical systems, those in *non-equilibrium* situations, in which order arises more easily. A recurrent example is that of Bénard cells. Consider a horizontal fluid layer heated from below. A temperature gradient is created through the fluid, opposing gravitational forces. When the gradient is large enough it reaches a critical value at which an internal convective motion emerges spontaneously. This convective motion forms very regular patterns such as hexagonal cells. Bénard cells are an example of a *dissipative system*, in which an external energy exchange (the temperature gradient) drives the system out of equilibrium allowing the system to give rise to regular flow patterns that would be highly unlikely to appear under thermodynamic equilibrium. Prigogine explains how the emergence of structured patterns in dissipative system is connected with the existence of specific nonlinear kinetic laws that amplify fluctuations at the molecular level, reaching a macroscopic level and stabilizing new, emergent structures (Prigogine and Nicolis, 1971). For Prigogine, the existence of different levels of organization in living beings is connected with a succession of instabilities amplified and stabilized through nonlinear dynamics.

With these developments Prigogine established the basis for a formal description of circularity in self-organization of physical and biological systems. Far-from-equilibrium thermodynamics introduce an interesting new perspective on how systems may be persistent. In equilibrium thermodynamic (conservative structures), systems may persist when they are assembled into spatially ordered forms (e.g. rocks, crystals or planetary systems). These systems will stay as they are indefinitely once created, since they are energetically conservative. The persistence of those systems is displayed in a form of cohesion that is maintained by the forces that hold the different parts together (e.g. chemical bonds or gravitational forces). These type of systems are not good candidates to support autonomous forms of organization (Barandiaran, 2008, p.93) since: (1) they tend to maximize their entropy levels and their level of complexity is defined by their boundary conditions (i.e. they are not capable of generating variable internal states by themselves), and (2) these systems can perform no work (i.e. useful directed release of energy), which according to some authors, for example Stuart Kauffman, will be a basic requirement for autonomy (Kauffman, 2002).

On the other hand, far-from-equilibrium systems generate dissipative structures which give rise to cohesive dynamic patterns under an energy gradient (e.g. whirls, hurricanes or

lasers). In these systems a great number of microscopic elements create a global, macroscopic ordered pattern in the presence of a flow of matter and energy. Interestingly, the emergence of ordered patterns is not only a consequence of the boundary conditions and the material properties of the elements, but the result of a process of interaction that creates a circular causality in which the generated macroscopic pattern contributes, in a non-trivial manner, to the maintenance of some cohesive or invariant relationships among microscopic elements. Nevertheless, complex forms of organization in living beings are too different from self-organizing systems that arise spontaneously in far-from-equilibrium conditions. Although the boundary conditions do not determinate the activity of the system, in standard dissipative structures, global emergent patterns only emerge when the system is under the appropriate boundary conditions. If those conditions (generally externally controlled) are not met, e.g. the energy flow sustaining the system disappears, the self-organizing system vanishes. This leads us to ask how autonomous forms of organization can come about in far-from equilibrium systems.

Kauffman defines autonomy as a necessary and sufficient condition for life. For him, an autonomous agent would be an agent defined as a 'self-reproducing system able to perform at least one thermodynamic work cycle' (Kauffman, 2002, p.49), where self-maintenance of a system is described as a work-constraint cycle – work being defined as 'constrained release of energy into relatively few degrees of freedom' (Kauffman, 2003, p.1094). To perform work, a system needs to build constraints but, at the same time, work is required to create those constraints. In this context, the self-maintenance of the system can be achieved as a coupling between: (1) exergonic reactions, which release energy in the form of work (e.g. diffusion, transport along a gradient), and (2) endergonic processes, which absorb energy in the form of work (e.g. synthesis of molecules, transport against a gradient). Therefore, a work-constraint cycle (Kauffman, 2002, 2003) arises when exergonic processes are constrained so that work is used to generate endergonic processes which generate those constraints, which in turn allow exergonic processes to occur. In Kauffman's words: 'work begets constraints beget work' (Kauffman, 2002).

On similar lines, Ruiz-Mirazo and Moreno (2004) claim that autonomous organization is only possible if it generates constraints that modulate the flows of energy so that those constraints are generated and contribute in this way to the recursive maintenance of the organization. They propose the notion of 'basic autonomy' as a property of self-constructing (cellular) systems in which endergonic-exergonic couplings involve: (1) the functional and organizational integration of inner constitutive processes and (2) the regulation of the boundary conditions through the control of the flow of matter and energy through the boundary. The constitution of a physical boundary in the form of a membrane is the requirement that avoids the problem of diffusion and dilution and allows the control of energy and material flows, allowing the system to display these two requirements for basic autonomy.

This poses a serious challenge to the cybernetic conceptions of negative feedback loops as the generators of purpose and self-maintenance. Imagine a system controlled by a simple negative feedback loop. It cannot generate anything by itself other than a state of stability

that will not change if it is not challenged from the outside. The system is not able to generate any work, or modify the boundary conditions that force the system to converge to a certain point. Even in Ashby's elaboration of an ultrastable system, without perturbations from the outside world, the system would end in a conservative state in which an immutable order is maintained. Thus, in order to describe far-from-equilibrium systems we need a different formal descriptions than the proposed by early cyberneticists. As we will see in the next section, nonlinear amplification and coordination dynamics are important aspects of the behaviour of far-from-equilibrium systems and are more suited to modelling them.

#### 1.1.4 TOWARDS AN OPERATIONALIZATION OF AUTONOMY

In this section we have described the historical development of some theoretical concepts useful for defining autonomous systems: conatus, homeostasis, ultrastability, equifinality, operational closure, circularity, autopoiesis or far-from-equilibrium systems. Specifically, we have reviewed three notions that allow us to think about the properties that define an autonomous system and its capability to display an emergent identity that differentiates it from its environment while at the same time it is able to compensate or adapt to perturbations from or interactions with it:

- Persistence of a system showing a tendency towards stability at certain states.
- Self-organized adaptation and reconfiguration of the system through regulatory mechanisms modulating agent-environment exchanges.
- Integration of persistence and adaptation through causal circularity between different levels of the system (physical substrate and emergent order parameters) through the regulation of interaction between system and environment.

At this point, we may provisionally rewrite Varela's description of an autonomous system as a system in which (1) integration of internal constitutive processes in a network of circular dependencies allowing the emergence of persistent high-order coherence patterns (2) the active regulation of processes of exchange with the environment through adaptive breaking mechanisms, and (3) the constitution of the system as a unity recognizable in the space (domain) in which it exists through its process of interaction with its environment. Barandiaran et al. (2009) describe an autonomous agent as a system which meets three conditions: individuality, interactional asymmetry and normativity, meaning that an autonomous system, to be so, must be able to constitute itself as an entity different from its environment, it must be an active source of activity in its environment, and it must regulate its activity in relation to certain norms that guarantee the continuation of its existence. Our provisional definition allows an agent to meet these conditions.

Furthermore, we have addressed some operationalizations of these concepts that took place in parallel with these theoretical developments: the feedback loop, the Homeostat (with

its second order plastic feedback loop) or the self-organization of a cellular membrane. Nevertheless, some definitions, for example, autopoietic or autonomous systems being homeostatic systems having ‘their own organization as the variable that they maintain constant’, or far-from-equilibrium systems with different levels of organization are still way too abstract and not easy to operationalize.

Together with the formal definition of concepts related to autonomy, during the last couple of decades we have observed an increasing interest in operationalizing concepts in cognitive science and neuroscience. Although some conceptual notions defined in this section were operationalized by its authors (e.g. Ashby’s Ultrastability and its implementation in the Homeostat, or in Varela’s models of autopoiesis and immune networks), recent advances in dynamic systems models allow defining through notions that help us to operationalize some concepts about autonomy. A parallel history of operational developments was underway, sometimes strongly intertwined with the former, and others branching out and opening new doors to both conceptual and operational innovations.

## 1.2 METASTABILITY, SOFT-ASSEMBLIES & CRITICAL SELF-ORGANIZATION

If in the last section we focused on conceptual notions around autonomy, here we proceed to the description of mechanisms with the objective of operationalizing these notions. We do not claim that a strong boundary exists between the development of the two (as we just saw the conceptual development of ideas as ultrastability and autopoiesis happened in parallel with their operationalizations), nevertheless, the objective of this section is to offer a perspective on other lines of thought that widen our understanding about autonomy in terms of mechanistic models reproducing modes of adaptive coherence and circular interaction between scales.

As we described in previous sections, one of the limitations of early cybernetic operationalizations of persistence and coherence in adaptive systems was the focus on the stability of particular components of the system. Without an interactive domain of description of the system, our perspective is limited to a single scale of operation of the system, escaping the possibility of a holistic description of a system.

In the 1950’s Arthur Iberall developed the concept of homeokinetics to apply the laws of thermodynamics to self-organizing living systems. Iberall was interested in using physics for overcoming a static definition of homeostasis, where the internal states of an organism are preserved at equilibrium despite changes in the external environment. In sharp contrast, the physical scheme outlined by Iberall tries to capture the essence of the dynamic regulations and interactions that constitute and preserve life. He studied how complex systems generate their own pressures for action, which emerge from the great number of internal degrees of freedom of the system (Iberall, 1977).

Starting from the idea that periodicities of cycles constitute the only form of dynamic stability for nonlinear systems whose processes degrade large amounts of free energy, the central feature of homeokinetics is the dynamic regulation of a system’s internal degrees of freedom by means of coupled ensembles of limit cycle oscillatory processes. Dynamical regulation of

a system involves switching among various marginally stable operating regimes, where different thermodynamical engines follow limit-cycle trajectories. Homeokinetic regulation consists in the modulation of these oscillatory processes, which are coupled or entrained constituting constellations of oscillators displaying a frequency spectrum that characterizes the overall, particular operating model. For Iberall, the nervous system constitutes a complex ensemble of thermodynamical engines (Llinás and Iberall, 1977). Iberall proposed a novel description of the organization of living beings:

An essential characteristic of a living system is its marginal instability [...] Both the external and internal environment constantly present the organism with an impulsive (vicissitudinous) input against a background of the slowly searched, changing milieu. As a result, the motor systems of the organism are plunged into intermittent search modes to satisfy all of its hungers [...] The motor-actuated living system unfolds its states, posture by moment. In each posture (the action of the body on the body), the system is temporarily locked into an orbital constellation of all of its oscillators. The psychological-physiological 'moment' emerges from the many oscillatory chains that enter into the constellation (Iberall and McCulloch, 1969).

Iberall's pioneering research opened up the road for the emergence of the modern field of coordination dynamics (Kelso, 1995). Disappointed with the representationalist turn that started to dominate cognitive science after the disintegration of cybernetics, Peter Kugler, Scott Kelso and Michael Turvey proposed to pick up Iberall's work and developed a frame of analysis where dissipative structures could be reduced to coordinative structures composed of far-from-equilibrium thermodynamical engines working in a periodic, limit-cycle fashion (Kugler et al., 1980). Their driving example was the coordination of sets of individual muscles creating ensembles of muscle collectives which are responsible for coordinated human movement (Kelso et al., 1981). For them, physical considerations of far-from-equilibrium thermodynamics could be dismissed as long as the properties of limit-cycle processes were taken into account. That is, description of dissipative systems should characterize (1) a tendency to a fixed amplitude and frequency robust to perturbations and (2) a tendency to mutually entrain or synchronize. Relying on the experimental evidence of the independence between timing and energetic considerations in muscle activity, they claim that self-organizing phenomena in animal behaviour can be modelled by coupled limit cycle oscillators dissociated from the thermodynamical system that generated them.

#### 1.2.1 COHERENCE IN SELF-SYNCHRONIZING SYSTEMS

The connection between far-from-equilibrium systems and coordinating structures opened the road to integrate a long line of operational models traditionally associated with the field of physics into a conceptual framework of life and autonomy. However, this exchange was also

happening spontaneously and the study of oscillations and coordination in nature had been there for some time. The idea is that the stability or the coherence of a system can emerge, not in terms of the individual dynamics of the components (which can be continuously changing) but in the relationships of coordination of those components in self-synchronized systems (Strogatz, 2004).

One of the first scientific reports on self-synchronization was made in the 17th century by Christiaan Huygens. In 1656, Huygens invented the pendulum clock. By determining the mathematical formula that related pendulum length to time (99.38 cm for a one second movement), Huygens discovered how to compute precisely the internal timing of a pendulum-driven machine. Before Huygens, clocks were much more inaccurate being based on springs, which can create an oscillatory behaviour when connected to a bouncing weight or to the axle of a turning wheel, which presented the problem of how to have the clock running at a constant rate as the spring ran down. More interestingly for us, Huygens also discovered the phenomena of entrainment. In a letter to his father dated 26 February 1665 (Huygens et al., 1888, p.243), Huygens describes how being confined in his room by a minor illness, he stared at two of his pendulum clocks mounted next to each other on the same support. He discovered a striking phenomenon, the two pendulums were synchronized swinging in opposite directions. He tried to disturb them, but within half an hour the pendulums regained synchrony. In contrast, when the two pendulums were placed in opposite sides of the room they lost synchrony. Huygens referred to this synchronized tendency as ‘an odd kind of sympathy’, and hypothesized that the pendulums must have influenced each other somehow through small vibrations in their support or through movements in the air (this hypothesis was to be confirmed by modern mathematical models c.f. Oliveira and Melo, 2015). This was the start of the study of coupled oscillations.

It was not until more recently when coordinated oscillatory activity was found in the human body, and particularly in the brain. In 1929, Hans Berger published a paper with his findings about the brain’s electrical activity. Initially motivated by the possibility of electromagnetic forces in the brain being the carrier of some form of telepathic activity, he performed numerous experiments analysing the behaviour of neural activity through a galvanometer, discriminating the activity coming from the brain from artefacts consequential of blood pressure changes or scalp skin. In his paper, he concluded that the most prominent electrical activity could be recorded from the occipital part of the skull when the subject’s eyes were closed. Although this electrical activity was too small to support the possibility of any type of telepathic activity, Berger’s paper provided a groundbreaking discovery: the recorded activity presented patterns distinguishing two types of waves. The first waves were large-amplitude rhythms or ‘alpha’ rhythms about 10 waves per second, induced when the subject was calm and with her eyes closed. The second type were faster smaller amplitude waves or ‘beta’ waves, present when the eyes were open. Berger’s discovery opened a path that, especially during the last decades, has promoted the study of brain oscillations which has brought considerable progress in the understanding of a variety of cognitive phenomena. The self-organization, emergent

patterns, phase-synchrony and oscillatory rhythms are now common issues in the study of brain dynamics (Buzsaki, 2006).

A couple of decades later, Norbert Wiener published his book *Cybernetics* (Wiener, 1948), popularizing a term that would gather under a unified framework some of the most important scientific advances of his time. The framework proposed by Wiener allowed thinking about problems of communication and control, no matter the system under analysis. This facilitated designing mathematical tools that could be applied to understand the functioning of machines, neural networks, animals or societies alike.

At that time, although the evidence for oscillatory brains in the wave was solid, nobody really knew why the brain's signals should oscillate. Wiener hypothesized that alpha oscillations in the brain work as a master clock for the brain. However, evidence at the time showed that individual neurons did not make for very efficient clocks; they were 'too imprecise'. Instead, Wiener thought that this clock may be a *collective* clock. Somewhere in the brain, groups comprising millions of neurons oscillating at different frequencies may spontaneously synchronize by pulling on one another's frequency through some kind of control mechanism. Wiener suggested that a very short-range form of radiation might be present in neurons, producing a resonance phenomenon pulling together large masses of neurons into a synchronized alpha rhythm (Wiener and Schadé, 1965, p.7). Beyond brain activity, Wiener also pointed out that mass synchronization was ubiquitous in nature: chirping crickets, croaking frogs or flashing fireflies also produce a 'natural rhythm', and urged biologists to conduct experiments following this research line. Wiener even proposed a signature of what the resulting spectrum of this synchronized activity might look like, modifying a Gaussian distribution of neuron frequencies into a narrow peak of synchronized frequencies, with two dips at the extremes of frequencies falling too far away (Strogatz, 2004, p.45). However, Wiener could not find convincing evidence of this signature.

Some years later Arthur Winfree discovered a much more fruitful approach to give a response to Wiener's intuition. Instead of thinking about the frequencies of the oscillators and how these may be 'pulled together', Winfree focused in the oscillators themselves, proposing a model that described the oscillators in terms of their natural frequencies, but also characterizing the *interactions* between oscillators. Winfree proposed a general description of a family of relaxation oscillators, in which an 'influence function' varies depending on the oscillators' activity, and a 'sensitivity function' representing how an oscillator responds to the signals it receives (Winfree, 1967):

$$\dot{\theta}_i = \omega_i + \frac{\kappa}{N} \sum_{j=1}^N P(\theta_j) R(\theta_i) \quad (1.1)$$

where  $\theta_i$  represents the phase of oscillator  $i$ ,  $\omega_i$  is its natural frequency,  $\kappa$  is the coupling strength between oscillators,  $N$  represents the number of oscillators in the system, and functions  $P$  and  $R$  represent the influence and sensitivity functions of the oscillator, typically described through sinusoidal functions.

For simplicity, every oscillator was identical to others, except in the value of its natural frequency that, as in Wiener’s assumption, followed a Gaussian distribution. Moreover, connectivity across oscillators was uniform and all oscillators were equally affected by others. Winfree derived the set of differential equations describing his system. Since the differential equations were nonlinear, traditional methods for linear differential equation systems were not useful for solving the system. Instead, Winfree combined mathematical analysis together with experiments with electronic systems and computer simulations to analyse the behaviour of the system.

These experiments showed how the behaviour of the system was quite counter-intuitive and contradicted in part the predictions made by Wiener. First, Winfree discovered that depending on the influence and sensitivity functions, the system displayed different behaviours. For some configurations, the oscillators actively opposed synchronization, while for others, they synchronized spontaneously, independently of their initial configuration. In these cases, synchronization emerged cooperatively, in a process in which a few oscillators synchronized by chance, exerting a stronger influence over other oscillators in a process of positive feedback, expanding synchrony to a larger cluster of oscillators. The system was self-synchronizing.

Furthermore, a curious effect appeared. If the coupling factor  $\kappa$  was too weak, oscillators were unsynchronized. But increasing slightly the value of  $\kappa$  did not at first have an apparent effect on the collective amplitude of the oscillators; they were still out of synchrony. However, further increase of the value of  $\kappa$  reached a point in which the collective amplitude of the oscillators (i.e., their level of synchrony) raised abruptly (Winfree, 1967). Suddenly, the system changed its behaviour from uncoordinated activity to a synchronized pack after a certain threshold was reached. The behaviour of the system does not correspond to statistical aggregation as Wiener imagined but to nonlinear interaction. Winfree discovered a phase transition.

In 1975, Yoshiki Kuramoto extended Winfree’s work proposing a different set of equations as a more tractable special case of his model (Kuramoto, 1975, 2003):

$$\dot{\theta}_i = \omega_i + \frac{\kappa}{N} \sum_{j=1}^N \sin(\theta_j - \theta_i) \quad (1.2)$$

where the influence and sensitive functions have been substituted by a simple sinusoidal function of the phase difference between oscillators. Kuramoto focused on this simpler version of the model and showed how to solve it exactly. This symmetrical rule materialized Wiener’s concept of frequency pulling. Another interesting contribution, instead of focusing on the collective amplitude between oscillators to quantify synchronization, was to propose a specific variable called the order parameter (ranging between 0 and 1) capturing precisely the collective coherence of the system. Kuramoto’s formula predicts the value of the threshold when the system goes from uncoordinated activity and synchronized clusters arise. Moreover, the order parameter allows describing a degree of ‘partial synchrony’, characterizing a state in which part of the oscillators in the system are synchronized and other parts are oscillating faster or

slower. The order parameter describes the state of this partially synchronized network as a value between total phase inhomogeneity and perfect synchrony.

Interestingly, Kuramoto model is rich enough to display a large variety of synchronization patterns and simple enough to adapt it flexibly to different contexts and analyse it mathematically, prompting an avalanche of literature exploring its possibilities (cf. Acebrón et al., 2005).

## SYNERGETICS

Like Kuramoto's model, many complex systems self-organize effectively reducing the dimensionality of their patterns. In many cases, the relaxation of many variables of the system is much faster than the slowest variables in the system, allowing an *adiabatic elimination* of these fastest variables. This is called by Haken the *slaving principle*, which allows reducing considerably the number of degrees of freedom in a system (Haken, 1978, p.191-199). The enslaving principle does not always completely dominate the system, and there are situations in which after a threshold is met the slaving principle can be transiently broken by, for example, chaotic motion (Haken, 1978, p.328-329).

Haken extends the notions of order parameter and collective variables from their traditional application in homogeneous physical systems to 'more complex' self-organizing systems. Order parameters typically addressed how individual atoms in a substance are constrained by other atoms. For example, a gas presents an order parameter equal to 0 since atoms move independently. As the gas is cooled down, there are temperature thresholds (critical points) where the substance goes through *phase transitions*, where the order parameter exhibits a discontinuous jump.

The use of order parameters proposed by synergetics allow us to operationalize the idea of reduction of the number of degrees of freedom that takes place in the self-organization of far-from-equilibrium systems, and the locking into an orbital constellation of oscillators that Iberall described (Iberall, 1977). It also generalizes phenomena like the nonlinear thresholds in large-scale synchronization we saw above.

### 1.2.2 FLEXIBLE COORDINATION

In the last section we have described basic mechanisms which can create tendencies in a system to joint its components together in a temporal coherence. However, this is not of much use if the system gets permanently trapped in a state of synchronization, or if it can only respond to the external perturbation of a synchronized state. Nevertheless, synchronizing mechanisms can also display modes which not only represent states of absolute coordination but also more flexible coordinating tendencies.

## RELATIVE COORDINATION

In the early 20th Century, Erich von Holst developed two principles for analysing movement coordination and locomotion in animals for example, lip fish and centipedes. These two principles characterize the coordinative properties of neural oscillations:

- A *maintenance tendency* of an oscillator to keep a steady rhythm. That is, each lipfish fin or each centipede leg was observed to have its preferred frequency of oscillation.
- A *magnet effect* that influences oscillators drawing them together to oscillate in synchrony.

A classic example for this phenomenon was devised by Kelso (Kelso, 1995, p.128) describing a parent and child walking hand in hand. Since the longer legs of the parent have a lower preferred frequency than the shorter legs of the child (i.e. they have different maintenance tendencies), there is a tendency for their movements to be out of phase. However, holding hands provokes the complimentary effect driving them to synchronize their movement (the magnet effect). The result of both tendencies is that the parent and child's strides are in phase for intermittent periods of time, drifting out of phase in some instants in which the child may perform an extra stride to cope with the larger strides of the parent.

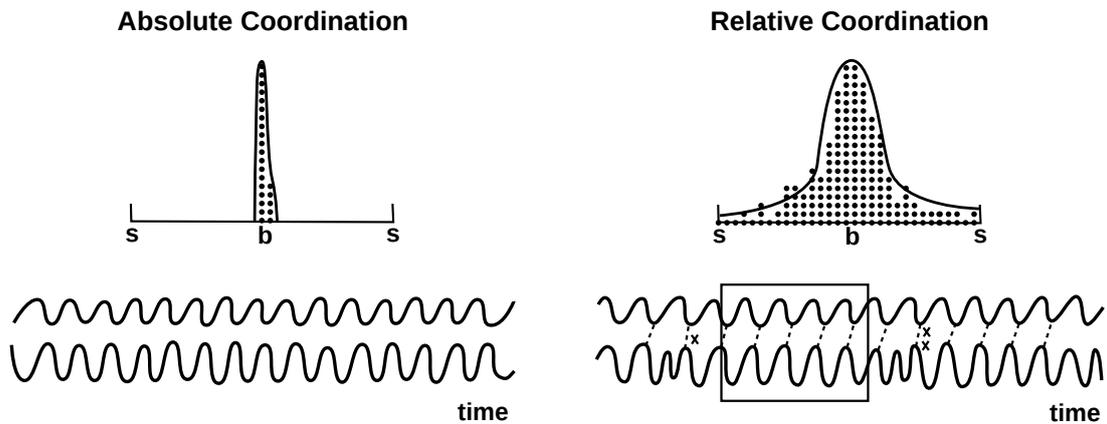
This idea was conceptualized by von Holst under the term of *relative coordination*. As we can see in his recordings of limb oscillations (Figure 1.5), relative coordination differs from absolute synchronization in the sense that the elements are coordinated in a much looser sense than required by strict phase-locking between their frequencies. Phase differences tend to particular stable relations, but this tendency is not absolute and the system is posed in a state where maintenance and magnet effects continuously compete but never completely dominate each other.

As a result of the combination of these opposing tendencies, an infinite number of variable couplings are possible: limbs may go in and out of phase depending on a changing balance between magnet and maintenance tendencies. As described in (Kelso, 1995, pp.74-79), animals with quadrupedal gaits may display a repertoire of different patterns depending on the intrinsic frequency of their movement and their coupling tendencies. The equilibrium between the two explains the transitions between jump, pace, bound and trot. Similarly, centipedes crawl by oscillating their legs in travelling waves from anterior to posterior. These waves vary phase relations among legs over time performing the centipede's locomotion. Moreover, von Holst discovered that amputating the legs of centipede leaving four or six, the resulting centipede engaged phase relations similar to those of four or six-legged creatures.

The idea of relative coordination depicts the behaviour of systems in which the behaviour cannot be explained from the individual behaviours of each element. Instead, the behaviour of the system emerges from the interaction between individual and collective tendencies in the system.

## METASTABILITY

Inspired by the development of synergetics, the notion of relative coordination has been operationalized under Scott Kelso's framework of Coordination Dynamics (Kelso, 1995), which combines experiments and formal theoretical models to study how the components of a system interact and produce coherent coordination patterns. Specifically, relative coordination



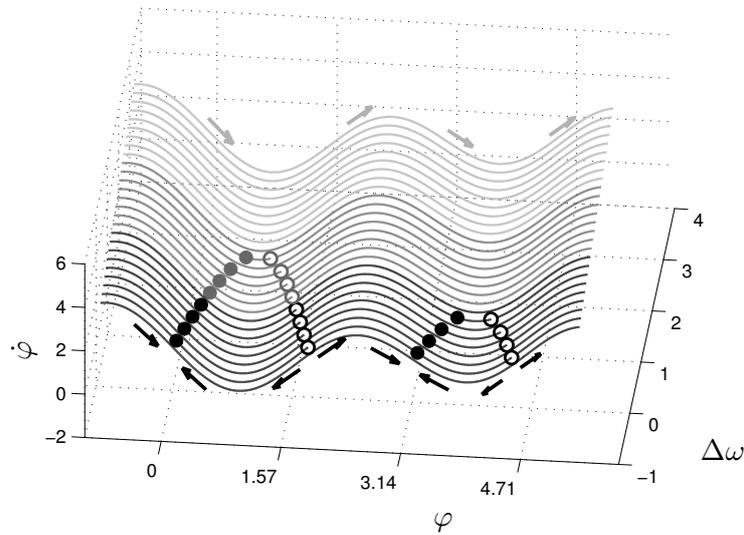
**Figure 1.5: Relative and absolute coordination.** Recordings of limb oscillations taken by von Holst under two different conditions. The signals come from pectoral (upper time series) and dorsal (lower time series) fin movements of a fish. Absolute coordination (left) and relative coordination (right). The top plots display the distribution of possible phase relations between the two signals under two conditions. The distance from S to S on the abscissa spans the phase interval 0 to  $2\pi$  rad. The bottom plots display the corresponding time series from which the phase relation is extracted. The left plot shows dominance of the magnet effect and the right plot shows relative coordination in the balance of magnet effect and maintenance tendency. Adapted from Kelso (1995).

has been typically exemplified under the concept of metastability in Kelso's most famous model: the *HKB model*. The HKB model describes a system composed of two coupled oscillators reduced to a single equation where the main variable is the relative phase between the two oscillators, and whose dynamics are shaped by the difference between the natural frequency of the oscillators and their coupling strength:

$$\dot{\varphi} = \Delta\omega - a \cdot \sin(\varphi) - 2b \cdot \sin(2\varphi) \quad (1.3)$$

The relative phase or phase difference,  $\varphi$ , represents the order parameter or collective variable that emerges from lower-level interactions of the two coupled oscillators,  $a$  and  $b$  are the coupling coefficients between the two oscillators, and  $\Delta\omega$  is the difference between their intrinsic frequencies. Despite its simplicity, this equation captures a wide range of self-organized phenomena. Different combinations of the control parameters  $a$ ,  $b$  (or rather  $b/a$ ) and  $\Delta\omega$  give rise to different collective behaviours. For example, when shifting the value of  $\Delta\omega$  while the values of  $a$  and  $b$  are held fixed, the system experiences phase transitions between three different modes of behaviour: monostable, bistable and metastable (Figure 1.6).

The monostable mode only has an in-phase synchronization stable point, whereas the bistable mode has two stable points corresponding to in-phase and anti-phase synchronization. Interestingly, the metastable region presents a surprising behaviour: the relative phase tends to in-phase synchronization and spends most of its time around this point. However, after some time the relative phase 'slips' and one of the oscillators advances a full cycle respect to the other until they find each other almost in-phase (Figure 1.7).



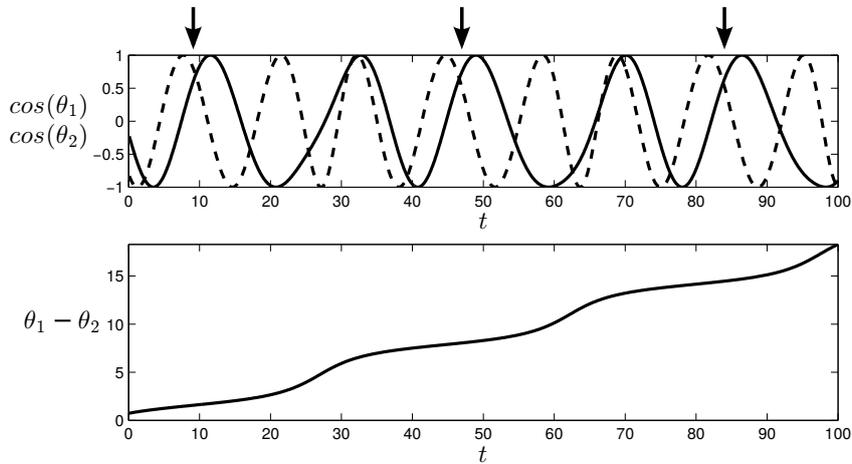
**Figure 1.6:** Phase space of the extended HKB equation for fixed values of the coupling coefficients  $a$  and  $b$ . The system exhibits three different types of phase space depending on the control parameter  $\Delta\omega$ , showing multistable (black), monostable (dark grey) or metastable (light grey) dynamics. Black and white dots represent respectively the attractors and repellers of the system for different values of  $\Delta\omega$ .

Metastable systems do not need to be defined as either being attracted or repelled by an attractor as two incompatible states, but they constantly dwell transiently near attractive states, switching flexibly and quickly between them. In this sense, metastability allows an operational description where attraction and repulsion influences coexist in a continuous scale of tendencies for dynamical groups to coordinate and segregate. The existence of a metastable continuum between attraction and repulsion seems to be supported by the presence of non-linear asynchronous couplings in the brain (Friston, 2000a,b), which support the ability of a system to create *dynamical instabilities* (i.e. metastability) that is fundamental to generate the diversity necessary for adaptive behaviour in the brain (Friston, 2000b).

A metastable regime is characterized by transient periods of stability, which allow the system to display coherent activity at the same time as it has some degrees of flexibility. Kelso claims that metastability 'is essential for flexibly entering and exiting coherent neural patterns and avoiding resonant mode-locked states [...] This is necessary if the mind is not to get stuck, or worse still, fly apart' (Kelso, 1995, p.225).

#### SOFT-ASSEMBLIES AND THE DYNAMIC CORE HYPOTHESIS

Some of the most relevant characteristics of biological systems is that, in certain conditions, they may exhibit behaviours generated by transient coalitions between their components. These coalitions, sometimes addressed as *soft assemblies* are functional groupings of a series of structural elements of the systems that are temporally forced by the influence of metastable



**Figure 1.7: Example of metastable dynamics.** Two variables  $\theta_1$  and  $\theta_2$  are coupled in a metastable mode. We can observe the oscillations of both signals displaying moments of synchronization and desynchronization (top) and the evolution of their phase difference presenting moments of stability and phase slippage (bottom). Moments of quasi-synchronization are indicated with an arrow.

couplings to act as a unified coherent entity (Kello and Van Orden, 2009). This view contrasts with ‘component-dominant’ dynamics or ‘hard-assembled’ systems, in which the resulting dynamics would be either a linear combination or a fixed function of the elements of the system.

The idea of soft-assemblies underlying coherent global behaviour is far from new, at least in neuroscience. Already half a century ago Donald Hebb (Hebb, 1952) developed the notion of a ‘cell-assembly’ as a ‘diffuse structure comprising cells [...] capable of acting briefly as a closed system, delivering facilitation to other such systems and usually having a specific motor facilitation’ (Hebb, 1952, p.xix). Such cell-assemblies act as transient ‘memories’ of a particular stimulus, in which repetition of the same excitation patterns provoke structural changes in the neural space associating neurons that are frequently activated together. For Hebb, neurons do not work in isolation and they cannot account by themselves for any percept or ability. In contrast, Hebb suggests that individual neurons may participate in different cell assemblies and be involved in multiple functions. This view is still alive in modern neuroscience. For example, hypothesis based on neural assemblies have been proposed for solving the so-called ‘binding problem’, i.e., how coherent representational states are established in the distributed neuronal systems subserving these functions (Engel et al., 1999).

Moreover, metastable modes of soft-assembling have been proposed as mechanisms that may mediate a compromise between large-scale integration in the brain and the ability to give rise to an extraordinary variety of complex, coordinated patterns (Kelso, 1995; Friston, 2000a; Werner, 2007a; Le Van Quyen, 2011). One of the most promising approaches to this problem was suggested by Francisco Varela (Varela, 1995), proposing that mental-cognitive states are mediated by a specific neural ensemble that emerges through transient phase locking of dis-

tributed neural regions. Every cognitive act corresponds to the emergence of one dominant ensemble which incorporates or discards cortical and subcortical regions into a unified and transitory whole. Later, Tononi and Edelman (Tononi and Edelman, 1998) used the term ‘dynamic core’ to describe this process, emphasizing both its integration and its ability to generate an extraordinary variety of complex patterns. Ever since the notion of a dynamic core has become relatively widespread in large-scale neuroscience studies (Varela et al., 2001a; Seth et al., 2006; Edelman, 2011; Ward, 2011).

Many of these views focus (as we will see in Chapter 4) on the interactive activity of the system at different spatial and temporal scales. If we want to take seriously the idea of extending metastable mechanisms to large-scale systems, we need an adequate framework to address how different scales and levels of description of the system might be interrelated maintaining an overall metastable state (Le Van Quyen, 2011). In the next section, we connect the concept of self-organized criticality with the idea of pervasive metastable mechanisms.

### 1.2.3 SELF-ORGANIZED CRITICALITY AND LONG-RANGE CORRELATIONS

Paradigmatic examples of the conceptual tools of the complex systems approach, with applications in several domains (in particular in cognitive science and neuroscience), are the concepts of self-organized criticality (SOC hereafter) and  $1/f$  noise (also known as scale-free or pink noise) (Bak et al., 1987; Jensen, 1998). The concept of SOC was initially proposed by Bak, Tang and Wiesenfeld (Bak et al., 1987) to define certain classes of dynamical systems which have a critical point as an attractor, displaying critical behaviour without any significant ‘tuning’ of the system from outside. Critical systems present very interesting properties: the most characteristic of which is the lack of a dominant scale of activity.

A popular example of criticality is the *Ising model* (Christensen and Moloney, 2005, Chapter 2), originally formulated to represent the behaviour of ferromagnetic elements. The model depicts the behaviour of a mesh of atoms with two dipoles oriented in one direction (the spin of the unit) interacting with their neighbours through their mutual attracting magnetic forces. If these forces are strong, the spin of all the units is aligned in the same direction, they are all aggregated and the complexity of the system is reduced to one dimension. However, if the attracting forces between the units are weak, each unit will face a random direction, and the system will be utterly disordered. Interestingly, at a point between order and disorder the critical state emerges, a permanent state of non-equilibrium where atoms’ spins are configured in transient, complex patterns in a process of continuous evolution (Figure 1.8.c). When the Ising mode is in this state between order and disorder it shows complex dynamical responses and their statistical properties have to be described by power laws. This point is a *critical point of phase transition* between uniform phases. Near this critical point the effect of thermal noise (independent spins) is balanced against neighbour interactions (interdependent spins), and this balance allows the emergence of complex transient patterns of aligned spins. More generally, critical systems typically display temporal and spatial scale invariance in the form of fractals and  $1/f$  noise, reflecting the process of propagation of long-range interactions

based on local effects. In contrast, SOC systems display a critical behaviour but, unlike the Ising model, their criticality does not depend on a fine tuning of their parameters. Instead, they are self-organized to operate under a critical regime, and they can do so spontaneously under a range of parametrical configurations. In SOC systems critical dynamics emerge from interaction-dominated threshold activity, in which many degrees of freedom interact, and the dynamics of the system is dominated by the mutual interaction between those degrees of freedom (Jensen, 1998). Remarkably, as critical systems do not present a specific scale of activity, they act as system displaying metastable behaviour at all scales. Moreover, SOC systems are self-organized to preserve this pervasive metastable configuration.

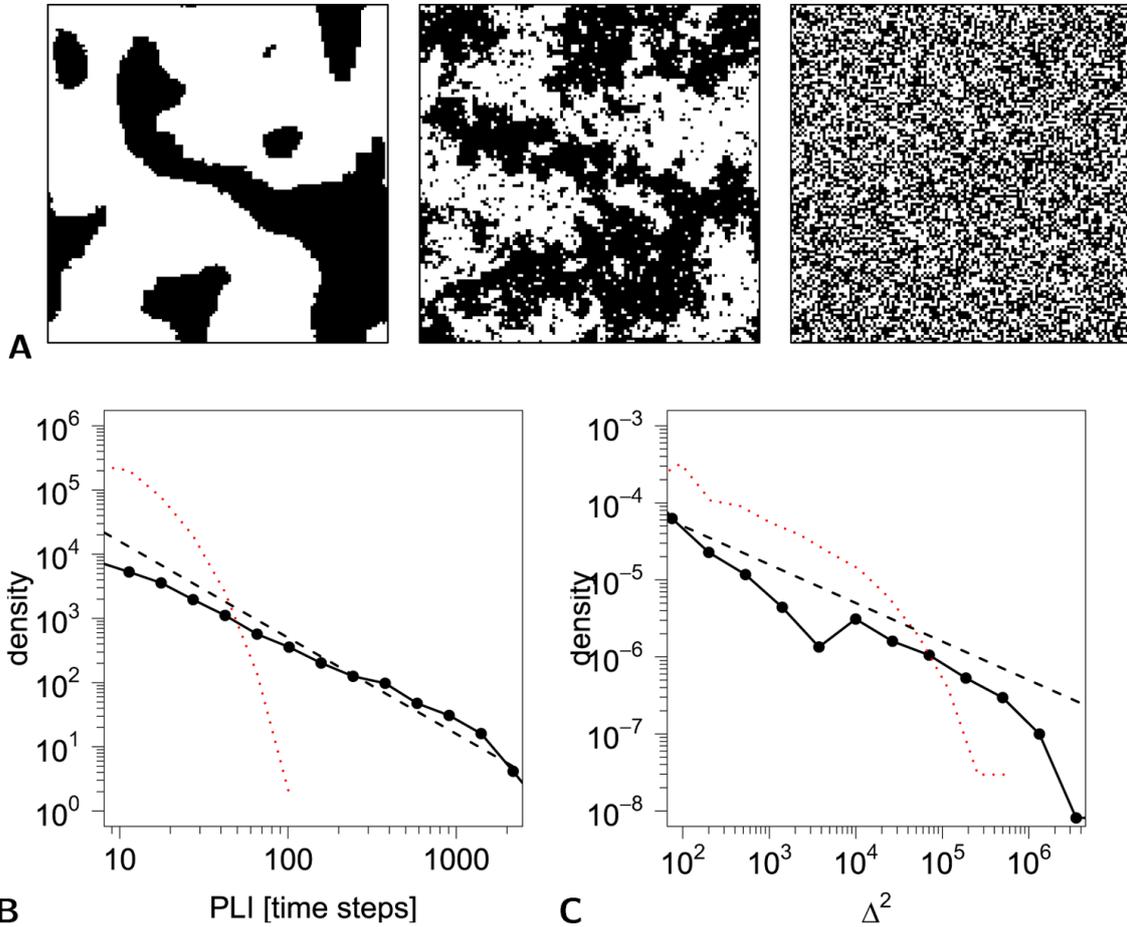
It has been shown that the brain stands in a continuous regime of SOC. Experimental evidence supports this hypothesis and also artificial neural network models. Criticality in the brain is suggested by power law scaling in degree distributions of functional brain networks or avalanche size distributions (Chialvo, 2006). Long-range correlations have been broadly found in EEG and MEG measures of brain activity in alpha, mu, and beta oscillations (Linkenkaer-Hansen et al., 2001). Scale free neocortical dynamics was also ascertained by Freeman (2005) in the EEG of rabbits. In addition, there are many instances of models of critically self-organized behaviour in artificial network models, which provide some interesting insights (Rybarsch and Bornholdt, 2012).

SOC and long-range correlations have also been broadly found in cognitive science and psychology. Van Orden, Holden and Turvey (Van Orden et al., 2003, 2005) used long-range correlations in different tasks to gather evidence to defend the hypothesis that certain systems are not modular and decomposable but ‘softly assembled’ systems sustained by *interaction-dominant dynamics* (IDD hereafter) as opposed to *component-dominant dynamics* (Van Orden et al., 2003). That is, IDD systems do not consist in additive interactions of their components, but multiplicative interactions that imply coordination between the different timescales in the system.

## EXTENDED CRITICALITY

Although critically self-organized systems provide insights into the kinds of organization that may operate behind living beings, most models of SOC are far from resembling living beings. SOC models such as sand hips or earthquakes display SOC properties in perfectly ‘flat’ and homogeneous systems, without any differentiation between levels of organization or any modularity in the system. Actually, what distinguishes biological and cognitive systems from inert systems lies in the role played by mechanisms of regulatory control in the functioning of these systems (Christensen, 2006; Moreno et al., 2011). That is, certain parts of the system (components or substructures) influence or modulate others to functionally constrain them and channel their dynamics to ensure or improve the maintenance of the system as a whole.

A system permanently poised in a state of critical phase transition suggests a configuration of a system where the local states depend upon (are correlated to) the global situation. However, Giuseppe Longo suggests that while in physics critical phase transitions are pointwise



**Figure 1.8: Criticality in the Ising Model.** a) Example of instances of the Ising model binary  $128 \times 128$  lattices showing the configuration of spins after 2,000 timesteps at low temperature (left); critical temperature (middle); and high temperature (right). At hot temperature the spins are randomly configured, at low temperature they are close to an entirely ordered state, and at critical temperature they have a fractal configuration. (b) Probability distribution of phase lock interval (PLI) between pairs of processes at critical (black line) and at hot temperature (red line) plotted on a log-log scale. The black dashed line represents a power law with slope . (c) Probability distribution of lability of global synchronization at critical temperature (black line) and at hot temperature (red dotted line). For the cold Ising model the equilibrium state of the system is a monolithic lattice with either all spins up or down, resulting in an entirely static system for which the PLI distribution is a Dirac Delta peak at the duration of the time series. Extracted from [Kitzbichler et al. \(2009\)](#).

(around a critical value of the control parameter), living systems may behave more as coherent structures posed in continual extended critical transitions (Longo and Montevil, 2012). These transitions are extended in spacetime and with respect to all relevant control parameters (e.g. pressure, temperature, etc.), and coordinated through integratory and regulatory mechanisms of the organism's global structure.

An important aspect stressed by Longo is that extended criticality is extended in both space and time, in the sense that critical transitions are maintained and persist in their many dimensions, while preserving all the physical properties of pointwise criticality (long-range correlations, new coherence structures, symmetry changes, etc.). This requires an active maintenance of organization (what Longo addresses as anti-entropy), i.e. the formation and maintenance of a global coherence structure that underlies this extended critical transition.

#### 1.2.4 OPERATIONAL TOOLS FOR MODELLING AUTONOMY

For Varela autonomy can be understood in terms of two interdependent properties (Varela, 1997). The first one is that *organisms are fundamentally a process of constitution of identity*. Identity here refers to a coherence in the system that provides it with some *unitary quality*. This is created by an operational closure of the system, that is, a circular interdependence of the parts of the system in which the system is the cause of its own production. Therefore, identity emerges without the need of any central controller but from a web of interactions, as an emergent process. The second property is that this emergent identity determines a *point of reference for a domain of interactions*. This means that what is produced with the emergence of an identity is not a process bound to a closed system, but an emerging interactive level between the system and its environment, which is the source of the informational, intentional or semantic value of the living being.

In this section we have focused on the possibilities of coordinative structures for representing processes typical of far-from-equilibrium thermodynamics. Specifically, different forms of coordination dynamics are useful for operationalizing concepts we described in the previous section. We have seen how mass self-synchronizing systems can account for phenomena of emergent persistence and coherence in large scale distributed systems, reflected in the emergence of an order parameter that influences the components of the system into a network of circular independences (order parameters emerge from the joint state of the oscillatory units of the system, and at the same time order parameters channel and regulate its activity). Moreover, the pulling tendencies of the order parameter of synchronizing systems also allow describing metastable dynamics, in which moments of quasi-synchronization are combined with moments of phase scattering, providing a flexible mechanism to break moments of synchronization and find new configurations of the system. In this way, the control parameter serves as a point of reference of the system's activity (although it is not located in any specific location). Although in the models above we have not differentiated what is a system and what is an environment, we can easily imagine how in a metastable coordination between an autonomous system and its environment, we cannot understand the history of interaction between the

two of them in terms of their individual activity, but instead as a result of the multilevel co-regulation between their individual activity and the emergent order parameter between them. The dynamics that dominate the process is not located in any particular component of the system, but in their shared interactive space. Moreover, we have seen how these interaction-driven systems tend to display scale-free dynamics and self-organized criticality when different levels of activity are involved. In such a way, SOC arises as a signature of complex interaction-driven processes, extending the notion of metastability to multiple scales. This allows us to conceptualize autonomous organization like a nested systems of interrelated interactive layers.

### 1.3 AUTONOMY AS AN ORGANIZATIONAL PRINCIPLE OF LIFE: FROM CELLS TO BRAINS TO SOCIETY

Above we have introduced a series of conceptual and methodological tools to analyse autonomy as an abstract systems property. We have described autonomy as a property ascribed to living beings, but we have not analysed in detail to what domains of life autonomy is applicable (e.g. molecular, cellular, organismic or social). Moreover, some properties described above can be applied to self-organized physical and chemical processes that are not alive in any sense. Thus, what is the relation between the concepts described so far and autonomy at different scales? We now explore whether different modes of autonomy may emerge at different scales in terms of these properties.

#### DISSIPATIVE AUTONOMY?

As we saw before, far-from-equilibrium or dissipative systems typically display emergent self-organized patterns. These systems are able to escape stochasticity at a microscopic level to generate global coherent patterns of order. All living beings depend on dissipative structures too, but there are many examples of non-living dissipative structures in which self-organized patterns emerge under a flux of energy. Previously, we described the behaviour of Bénard cells, but many examples can be found in nature including earthquakes, auroral substorms, or solar flares, which also frequently display properties as self-organized criticality. Are dissipative systems autonomous? The emergence of order patterns may provide the system with some kind of coherence in which a web of interactions constitutes the system as a unity. Nevertheless, this emergent identity does not generate a domain of interactions between the system and the environment. As we will see below, living systems are able to constitute themselves as an identity able to regulate the loop of its own interaction with the environment, generating a history of interactions in which the system acts and adapts to its environment to guarantee its own conservation. Dissipative systems may be able to sustain themselves while a flux of energy or matter is present, but are not generally able to interact with the environment for maintaining the conditions of this loop.

## METABOLIC AND CELLULAR AUTONOMY

As we saw when analysing cellular autopoiesis, cells can create their own identity through processes of metabolic self-production of the network that produces the cell, and constituting the cell as a unit and a domain of interactions with its environment through the constitution of a permeable membrane that contains the cell. Autopoietic theory proposes that the autonomy of the cell should be understood in terms of regulatory interactions between its metabolic structure and its environment that makes possible the maintenance of its organizational invariance. Several models have analysed how autopoiesis, emergence of boundaries, cell-environment interaction and so forth may take place in chemical media (Luisi, 2003; Solé et al., 2007; Ruiz-Mirazo and Mavelli, 2008; Hanczyc and Ikegami, 2010).

Here, the system does not become a distinguishable entity because of its molecular composition or its internal configuration. In contrast, what is maintained in the system is its own organization which not only involves internal metabolic cycles but the cell-environment cycles that allow a continuous regeneration of materials while preventing the cell from dissolving in its chemical milieu through the formation and active maintenance of a cellular membrane. The existence of this invariant organization is constantly challenged by internal and external perturbations, which the system counteracts. It is precisely these breakdowns in autopoiesis (even if they are minor, like changes in chemical concentrations, or major, like disruptions of the membrane) that creates significance and value about the surrounding environment. It is the creation and regulation of the membrane (which allows the material and energetic flows between cell and environment that sustain the existence of the cell) allowing both the constitution of the system as a distinct unit from its surroundings and the active regulation by the cell of the interactions with its environment. This active regulation creates an asymmetry in these interactions in which the cell as a virtual entity constitutes the point of reference of these interactions.

## ORGANISMIC AUTONOMY

Integrity and self-preservation of the organism depend on regulatory cycles involving brain and body at multiple levels. Multicellular living beings are built over many cells that interact in a medium that generates the coordination that constitute collective organisms. Multicellular organisms actively regulate that tridimensional boundary with their environment, so the resulting system can operate as a single organism against its environment. The systemic regulation of the boundaries of the organism and its cellular structure allows the emergence of its ontogenetic development. Over multiple timescales, developmental processes are subject to processes of selection and differentiation (Elman et al., 1998; Gottlieb, 2007). The energetic and organizational maintenance of multicellular systems requires inter-cellular mechanisms of control of cell differentiation and cell division processes. This regulatory control allows organisms to exhibit a functionally differentiated but globally integrated developmental organization, which constitutes an ensemble of cells as a self-determining collective entity (Arnellos et al., 2013; Arnellos and Moreno, 2015).

Another level of multicellular organisms' autonomy is constructed by immune network. The central characteristic of the immune network is the definition of the organism's molecular identity (Varela and Coutinho, 1991). From the assertion of this identity an organism is able to derive aspects as tolerance and self-nonsel self discrimination. Natural tolerance is the result of a dynamic pattern of assertion of this identity: 'self is what the system has learned to "see" as self and to discriminate from all other antigens that did not take part in its constitution' (Varela and Coutinho, 1991, p.165).

In animals, organismic regulation also has an affective and conscious dimension. Proponents of *affective neuroscience* (Damasio, 2000; Panksepp, 1997, 1998) have emphasized the dependence of a *background consciousness* on the homeodynamic regulation of the whole body: different parts of the cortex, brain stem and hypothalamus process neuronal and humoral signals from the body and integrate them into a 'body landscape' that is constantly changing. This body landscape would include interoceptive and proprioceptive signals, as well as the state of the inner milieu (hormone concentrations, pH-value of the blood, glucose levels, etc.). Thus, brain and body influence each other on a constant circular feedback of organismic self-regulation. This interaction results in a background feeling of being alive, a basic self-affection which lends a sense of 'mindness' to all our experiences (Fuchs, 2011). Moods, emotion and every conscious state are ultimately rooted in a loop of homeodynamic regulation between brain and body that integrates the state of the organism as a whole.

#### SENSORIMOTOR AUTONOMY

When we move up to the autonomy of multicellular organisms, we find a quite different scenario compared with single-cell autonomy. Organisms do not depend anymore on a physical membrane to maintain a continuous flux of material exchange or to prevent being dissolved into its surroundings. Furthermore, in the case of organisms with nervous systems, the emergence of identity is defined not only in terms of the physical construction of the organism and the physiological coordination of multicellular tissues and organs, but a different level of identity emerges. The key point is that the development of a neural system is closely connected to motion being an integral part of the lifestyle of the organism (in order to correlate muscles, secretions and other effectors to sensory surfaces). Neurocognitive identities emerge not in the physico-chemical interactions between the agent and its surroundings, but rather in the space in which the movement of the agent is coupled to a stream of sensory modulations (Varela, 1997). This interaction creates an equivalent circular relation between a network of neural interactions and a series of sensory streams that continuously create invariant patterns of sensorimotor correlations. This gives rise to the agent as a behavioural mobile unit in space that is in continuous interaction with its environment.

This view of autonomy connects directly with sensorimotor theories of cognition and situated approaches. For these perspectives sensory input is much more than an information flow to be processed by the brain. Rather, situated approaches to cognitive behaviour claim that the agent-environment coupling shapes brain dynamics in a manner that is essential to be-

havioural or cognitive functionality (Steels, 1990; Clark, 1997; Chiel and Beer, 1997). In other words, macroscopic functional behaviour (e.g. intentional grasping or perception) emerges from microscopic sensorimotor dynamics (e.g. proprioceptive and visual feedback in grasping or saccadic movements in visual perception). Thus, cognitive behaviour is not the result of a linear computational sequence involving sensation → perceptive-categorization → planning → action-selection → motor-execution, but the result of recurrent sensorimotor and brain oscillatory coordination at multiple scales. The central role that sensorimotor dynamics play in cognitive phenomenology has been recently highlighted by Sensorimotor Contingency Theory (O'Regan and Noë, 2001), defending that what is constitutive of visual perceptual awareness (and, it could be argued, other cognitive processes) is not a specific internal state of an agent, but the structure of sensorimotor contingencies. For O'Regan and Noë, vision is 'a mode of skilful encounter with the environment, requiring knowledge of sensorimotor contingencies and also the ability to make use of that knowledge for the purpose of guiding action, thought, and (in humans) language use' (O'Regan and Noë, 2001, p.959-960). To perceive is to act in a specific manner that brings forth the structure of sensory changes in relation to the activity of the agent. To see or to perceive is something that is done and lies on the very sensorimotor coupled dynamics of an agent. In neuroscience, the 'radical embodiment' framework proposed by Thompson and Varela (2001) suggest that processes that are crucial for consciousness are not confined inside the brain but cut across brain-body-world divisions. According to this view, the brain is organized in resonant neural assemblies that are transiently integrated, each specific assembly underlying the operation of determinate cognitive acts. These assemblies crucially depend on the manner in which brain dynamics are embedded in the somatic and environmental context of the agent's life.

All of these perspectives share the idea that an emergent interactive level reshapes the domain of behaviour. This level is constituted by the interaction itself between the agent and the environment, through a stream of sensorimotor interactive loops. It is the constitution of the system as a distinct entity in space that constitutes a cognitive self acting as an autonomous system, creating a point of reference leading the action of an agent over its world and allowing the emergence of the autonomous agent.

#### SOCIAL AUTONOMY

A special case of the kind of autonomy above is the case when the 'environment' contains other autonomous agents, as in the case of social interaction. In such cases a process of interaction between individuals creates a process of coordinations of movements and utterances. Social interaction frequently generates collective coherent patterns in which the dimensionality of the system is reduced in a synergistic way, creating interpersonal patterns exhibiting circular regulation (Oullier and Kelso, 2009b; Riley et al., 2011). Some authors argue that the process of social interaction itself can take a form of autonomy. The underlying idea is that the process of interaction itself may become leading over the two interactors: the interaction process gains 'a life of its own' (Jaegher and Paolo, 2007; Fuchs and Jaegher, 2009). The

process of sense-making out of agent-world interaction necessary for autonomy is elaborated here as a *participatory sense-making*. This notion addresses different cases where the identity of the process generated is not defined in terms of an individual organism but is generated among two or more autonomous organisms. Along the same lines, Evan Thomson argues that in humans and other social creatures embodied cognition emerges from the dynamic co-determination of self and other, affording in some cases non-egocentric or self-transcendent modes of consciousness (what Thomson calls ‘interbeing’) (Thompson, 2001). Social autonomy arises when neural and somatic activity of the participating subjects is regulated by the ongoing embodied awareness and interaction between them.

Moreover, these forms of social autonomy can be applied to a transient dyadic interaction, but also to a large-scale social level. One example is Niklas Luhmann’s idea of social autopoiesis (Luhmann, 1986). According to some perspectives on social autopoiesis, different social entities as a political party or a family can be autopoietic since the rules that define them can act as a kind of boundary formed by the social structure. These social boundaries are open systems, which allow members to join or leave the structure. Also, as in biological autopoiesis the social structure that acts as a boundary emerges from a recursive network of the members of the structure, which is what allows the constitution of the social structure as a differentiated unity in respect to its environment. However, Maturana and Varela have argued at different moments that autopoiesis is not directly applicable to the biological realm (Mingers, 1992), and Varela is concerned about the effects of the authoritarian implications of a direct transposition of the concept of autopoiesis to the social realm (Protevi, 2009, p.12-13).

During the last decades, the idea of large-scale autonomous social organization has become even more attractive due to the possibility of digital communication tools. Howard Rheingold has coined the term *smart mobs*, to characterize large groups that cooperate in novel ways taking advantage of the communication and computing capabilities of widespread technological devices (Rheingold, 2007). A smart mob, in contrast with the usual connotations of a mob, would be able to behave in intelligent and efficient ways as a unitary system, thanks to the connectivity enabling people to coordinate as a coherent social entity. Others have stressed the role of information and communication technologies (ICTs) in the emergence of collective intelligence based in distributed, enhanced, real time coordination resulting in effective mobilization of skills (Levy, 1999), and by creating a collective pool of social knowledge by expanding the domain of human interactions (Flew, 2011). Manuel Castells has proposed that the development of interactive, horizontal networks of digital communications tools over the Internet and wireless communication networks has induced the rise of a new form of communication – mass self-communication (Castells, 2007). Castells argues that the autonomy and influence of insurgent politics and social movements is enhanced as historical shifts of the public sphere take place from the institutional realm to a new communication space.

The possibilities of a social autonomy requires the interaction between a community of individuals and the outside of the community to be mediated by the emergence of a collective self, which both determines interactively who belongs to the community and serves as a point

of reference for the action of a collective self as an autonomous entity.

#### 1.4 EXPLORING AUTONOMOUS AGENCY

Among the different modes of autonomous organization of the living, we observe that there is a transition from modes of autonomy based in the material and energetic constitution of living beings (in cellular and organismic autonomy) to the sensorimotor interactive constitution of individual or collective agents. For the first, the study of the role of material interfaces that allow the constitution of the agent as a material unity in front of its environment has been a central element of the characterization of autonomous organization. Nevertheless, the agent-environment boundary is much more fuzzy for the second group. Autonomy, with its emphasis on closure and system-environment distinction, seems to be challenged when the interactions between agent and environment are open and recurrent, which is characteristic of cognitive and social life.

As we saw above, cellular autopoiesis is a well characterized case of autonomy, and frequent attempts have been made to extrapolate cellular autonomy to other domains like cognitive or social autonomy. As we will see in the following chapters, these attempts have at best fallen short of obtaining a convincing and experimentally testable extrapolation from biological into neurocognitive or social autonomy. In contrast, we believe that in these domains the modes of autonomy in operation are fundamentally different from those present in a cell. They do not have much to do with the construction of a physical boundary of the system, but with the emergence and maintenance of embodied forms of agent-environment coordination. The distinction between biological forms of autonomy and those built over forms of open and recurrent interaction deserves to develop a clear framework to account for the latter. The contribution of this thesis aims to advance in models and analytical tools to define and delimit a notion of interactive autonomy rooted in the history and conceptual tradition of biological autonomy.

In section 1.1.4 we defined autonomous agency in terms of the conditions, proposed by Barandiaran et al. (2009), of individuality, interactional asymmetry and normativity. The notion of autonomous agency allows defining forms of autonomy that are not (only) based on its material self-production but in the interaction dynamics between an agent and its environment. Nevertheless, several questions are to be clarified. For example, what is the relation between the operational closure of the system (its individuality) and the (asymmetric) interactions with its environment? What is the role of internal and external coordination in the generation of identity? Imagine the case of a wave surfer. We know that she is an organismic autonomous system, since she is a healthy human being able to maintain her own biological organization. However, what characterizes her as an autonomous cognitive agent in the very act of surfing?

As we described above, the emergence of high-order coherence patterns (e.g. in the form of self-synchronized processes), is a good candidate to define the individuality of a system, defining the operationally closed network that defines a system. Other processes may define

the individuality of a system under certain circumstances, but emergent order parameters is a characterization general enough and experimentally plausible enough to apply to a vast range of circumstances at neural and behavioural levels (Buzsaki, 2006; Strogatz, 2004). In the case of the surfer, when she rides a wave, her body is coordinated by several synchronized waves of movement that equilibrate her body, her arms, her legs, to move as a unit that is different to the sea and the wind, or other organisms around her. Furthermore, when she is surfing a wave, the surfboard, or even some mesoscopic flows of water in the wave synchronize with her movement, creating a transient wave-board-surfer dynamical unit. In these cases, the skilled surfer does not have to think how to control the surfboard or the wave as something ‘external’. Instead, she ‘sees through’ them, but they become ‘transparent’ (Dotov et al., 2010), that is, absorbed into the dynamics of her own activity. This absorption does not have anything to do with an external observer deciding which is the cognitive agent performing the action, but with the emergent order parameter modulating the wave-surfboard-surfer dynamics.

On the other hand, interactional asymmetry is somewhat more problematic. Although there are forms to characterize it through energetic considerations or in terms of statistical correlations, they are limited and contain challenges (Barandiaran et al., 2009, p.3-4). Barandiaran, Di Paolo and Rohde define interactional asymmetry as the condition in which ‘an agent as a whole drives itself, breaking the symmetry of its coupling with the environment, so as to modulate it from within’. However, there are many cases in which it is not clear how this asymmetry might work. In the case of the surfer on top of a wave, the dynamics of the wave probably propels the surfer much more than the surfer propels herself, and obviously the wave provides much more energy to the surfer’s movement than vice versa. Nevertheless, the surfer continuously equilibrates herself in order to modulate the effect of the wave over her, coupling and decoupling herself from the different flows that compose the wave, taking advantage of them to *control* its movement. Our intuition, which we will test in detail in chapters 3 and 4, is that what matters here is not a mere asymmetry, but an asymmetry in the circularity of the causal relations between surfer and waves. Probably, in a short timescale, it is the perturbations and flows of the wave which drive the dynamics of the surfer, which only has the option to be *seized* by the wave or lose its equilibrium and fall. However, in a longer timescale, a series of subtle movements, reequilibrations and reorientations of her movement allow her to ride the wave, i.e. to control it. We propose that the asymmetry between agent and environment is basically a matter of multiscale causal relationships, in which a necessary condition for agency is to be able to be reactive to the microscale perturbations from the environment, and at the same time be able to act as a macroscopic entity over the environment, modulating the effect of environmental perturbations in the long run.

Finally, normativity emerges from the precariousness of the system’s autonomous organization. That is, normativity is defined as the conditions of viability of the system. Actions are good or bad, appropriate or inappropriate depending if they contribute or not to supporting the system as an autonomous organization (Christensen and Bickhard, 2002). While viability of the system has been generally understood in terms of the physiological viability (e.g. home-

ostatic mechanisms maintaining temperature or nutrient levels), in forms of autonomy not based only in biological regulation, we face the question of how do we define the boundaries of viability of a process of autonomous interaction (Barandiaran and Egbert, 2014). In the case of the surfer, since the wave-board-surfer system is constituted in an interactional domain, the very existence of the system depends on maintaining a specific coupling between the surfer, her surfboard and the waves. The act of surfing itself defines some needs and norms required to its own viability. We can imagine how different events as failing to maintain coordination or the impact of a wave might end the system itself: making the surfer lose equilibrium and fall, breaking the surfboard, or even ending her life in the most extreme cases. The constitution of a wave-riding autonomous entity is what creates a domain of perspective by which actions might be adequate or not to ensure the survival of the action of surfing. The series of interactions between autonomous agent and embodiment can maintain its own organization or be the cause in the breakdowns of its vital, psychological or cultural requirements (Barandiaran et al., 2009, p.5).

Along the next chapters of this thesis, we will explore the mechanisms sustaining that each condition for autonomous agency is met and what is the relation between them. Using the tools described in this chapter, we will describe the coordination processes that constitute the operational closure of the system in a sensorimotor space, as well as the system environment asymmetric interaction for different cases. In the second part of this thesis we will present a robotic model of autonomous behaviour to explore this idea, characterizing in detail a model of interactive sensorimotor autonomy and its properties. In the third part of this thesis, we will assess whether autonomous agents can arise from processes of collective interaction in the social domain. Our objective is to contribute to the characterization of a particular type of autonomy, that of autonomous agency, and specifically to describe what kind of processes of interaction dynamics can make it possible.



## Part II

# Sensorimotor Constitution of Neurocognitive Autonomy



# 2

## Constitutive Interaction: the situated-HKB model

**ABSTRACT:** We address the constitution of neurocognitive identity and argue that it is built in a sensorimotor space. In contrast with much modern neuroscience, we analyse in detail what might be lost when a sensorimotor coupled system is treated in isolation.

**CONTRIBUTIONS:**

- We extend the popular HKB model embedding it into a minimal robotic model to explicitly address the issue of embodiment for a simple coordinative controller.
- We solve the dynamical system of the situated HKB model, and compare it with a passively-coupled HKB model, in order to reveal the quantitative and qualitative differences provided by situatedness.

**RELATED PUBLICATIONS:**

- Aguilera, M., Bedia, M. G., Santos, B. A., and Barandiaran, X. E. (2013a). The situated HKB model: how sensorimotor spatial coupling can alter oscillatory brain dynamics. *Frontiers in Computational Neuroscience*, 7:117.
- Santos, B., Barandiaran, X., Husbands, P., Aguilera, M., and Bedia, M. (2012). Sensorimotor coordination and metastability in a situated HKB model. *Connection Science*, 24(4):143–161.

Embodiment is becoming an increasingly popular idea in robotics, cognitive science and neuroscience. The idea that our brain is intimately entangled with our body seems like a more plausible hypothesis than it being confined into our skull sending orders to a merely reactive body. In this and the following chapter we present the idea of interaction between an agent and its environment as a constitutive element for the creation of the identity of an autonomous agent as a mobile unit in space. In Chapter 1 we proposed that neurocognitive identities scale up from metabolic and organismic forms of autonomy in a way that autonomy is not based on the maintenance of a network of physico-chemical interactions, but rather the

maintenance of a space of sensorimotor interaction in which the actions of the agent are coupled to a stream of sensorimotor modulations maintaining some organizational invariants. This interaction creates an equivalent circular relation between a network of neural interactions and a series of sensorimotor loops that continuously create invariant patterns of sensorimotor correlations.

Nevertheless, although many cognitive scientists and neuroscientists would agree that the brain is continuously modulated or biased by the body and the environment, the idea of brain-body-environment coupling as a constitutive element of neurocognitive identity is more controversial. Where does such resistance come from? We argue that the dominion of the computationalist paradigm during most of the second half of the 20th century still has a great influence in present-day science, and, that even now that embodied or dynamical explanations are much more common, many of these explanations ultimately depend on a representational and internalist view of cognitive processes. To support this claim we first review recent history of cognitive science, since the rise of computationalism to the emergence of new dynamical and embodied approaches. We argue that, even when dynamical approaches have been largely accepted in neuroscience, situated and embodied views are usually neglected in favour of representational relations with the environment. We present a minimal model, the *situated HKB model*, which illustrates what internalist approaches might be missing. The model presents different modes of engagement with the environment by varying a sensory control parameter. The model is simple enough to perform a full dynamical analysis and extract the underlying dynamical landscape governing each behavioural mode. Assuming that cognitive autonomy is built over an interactive domain instead of a purely internal one, we consider what consequences this implies. We explore this issue by comparing our robot situated in its environment with an exact copy of the robot that only receives an input signal recorded from the first robot. The model shows that after removing sensorimotor coupling from a simple robot, even when it is presented with an identical input, the internal neurodynamic configuration changes dramatically. The model shows how fine grained sensorimotor contingencies are strictly necessary or constitutive of functionally distinct neural signatures.

## 2.1 SITUATEDNESS AND EMBODIMENT IN COGNITIVE SCIENCE

### 2.1.1 DUALISM, EXPERIMENTAL PSYCHOLOGY AND PASSIVELY-COUPLED KITTENS

The idea that cognitive autonomy is constructed over a domain of sensorimotor interactions clashes radically with the pervasive dualist tradition in western thought. Although a dualistic conception of body and mind was already present in Plato and Aristotle, Stoic and Augustinian traditions, and scholastic philosophy, Descartes it is thought to have provided dualism with an unprecedented protagonism stating that the human mind is separated and operates independently from the body, being a completely different entity. The body is conceived as an automaton comparable with man-made machines, which can be understood in purely mechanistic terms of physical causes and effects. Descartes conceived that stimuli were able

to set up vibrations in sense organs, which in turn would tug tiny fibres through the nerves, opening small valves letting a cerebral-spinal fluid into the muscles, which fattened up and shortened producing movement. In *De Homine*, Descartes illustrates these mechanisms by means of the automatic reaction that is triggered when we touch fire, thus providing the first clear discussion of reflex action in the history of science (Clarac, 2008). Descartes imagines that animal behaviour can be entirely accounted for by such automatic, reflexive mechanisms as a hydraulic machine. Other faculties, such as the ability to think and reason, consciousness, (i.e. the ability to think and act independently of the physical causes affecting the body) are ascribed to the soul. For Descartes, the soul is informed of the vibrations in sense organs, and could gain control over the movement of the body by pushing the pineal gland and altering the flow of cerebral-spinal fluid to influence muscles.

This dualistic conception of the mind leads Kant to claim that a psychological science is not possible, since pure thought lacks spatial dimensions rendering impossible the application of mathematical laws. It was not until the second half of the 19th century that experimental psychology arose, contradicting Kant's claim, promoted by Wilhelm Wundt. Experimental approaches to psychology soon brought a quite different perception of how body and mind interact. One of the first to propose a different schema of interaction between mind and body was William James. He proposed that a reflex action was composed by three elements: a sensation followed by a mediation of the nervous system (spinal cord or brain), and finally by a response or action (James, 1879). With this schema he stressed the role played by the central nervous system as closely connected with the other two, and entirely purposed to lead to action. For James:

The structural unit of the nervous system is in fact a triad, neither of whose elements has any independent existence. The sensory impression exist only for the sake of awaking the central process of reflection, and the central process of reflection exists only for the sake of calling forth the final act. All action is thus *re*-action upon the outer world; and the middle stage of consideration or contemplation or thinking is only a place of transit (James, 1896, p.113).

The interrelation between perceptual and action processes, as opposed to one-way stimulus-response links, was further stressed by John Dewey. According to Dewey, a sensory stimulus is not an isolated trigger for the perception-action process. Instead,

We begin not with a sensory stimulus, but with a sensory-motor coordination [...] In a certain sense it is the movement which is primary, and the sensation which is secondary, the movement of the body, head, and eye muscles determining the quality of what is experienced. In other words, the real beginning is with the act of seeing; it is looking, and not a sensation of light (Dewey, 1896, p.358)

For Dewey, both sensory and motor process are coupled so '[a]ctions are not reactions to stimuli; they are actions into the stimuli' (Dewey, 1914).

Donald Hebb developed in 1949 a theory of brain organization, proposing the idea that neurons are able to group together in diffuse structures he labelled as 'cell assemblies' that result from simultaneous activation (frequently expressed by the dictum 'neurons that fire together wire together'). Hebb rejects the assumption of a sensory dominance of behaviour and introduced the interest in motor aspects of perception (Hebb, 1952, p.3). Hebb also rejects Gestalt notions of perception and learning dependent upon the 'pattern, or shape of the sensory excitation'. On the contrary, Hebb takes interest in the motor aspects of perception arguing 'that "simple" perceptions are in fact complex: that they are additive, that they depend on motor activity and that their apparent simplicity is only the result of a long learning process' (Hebb, 1952, p.17). For example, Hebb sees eye movements as activities that facilitated perception and enabled complex perceptions to take place. These movements *facilitate* the emergence of a particular cell assembly, thus allowing the succession of phase sequences of perceptual events (Hebb, 1952, Chapter 5).

Hebb's work inspired experimental research on the effects of exposure to the environment in the development of spatial perception and coordination, often involving rearing animals in conditions of sensory deprivation by placing them in restricted environments. For example, after constraining kittens to environments of diffused or patterned visual experiences under holders, they failed to learn a movement discrimination task in a patterned environment with an exercise pen, while normal kittens could perform the task correctly (Riesen and Aarons, 1959). However, Richard Held and Alan Hein claimed that in such experiments it can still be argued that rather than sensorimotor coordination, what was lacking is the variation in visual stimulation for the kittens under holders. Thus, they designed a clever experiment, in which two neonatal kittens were harnessed to a carousel (Held and Hein, 1963). One of the kittens was harnessed but could move the carousel as it walked. The other kitten could not move the carousel and was driven by the other kitten movement. Kittens raised in the dark were given their first experience in light at 8 to 12 weeks of age. The experience consisted of a short daily period of controlled visual stimulation in a patterned environment. They spent the remainder of the day in the dark. The environment presented a vertical pattern that the kittens experienced as they moved. Interestingly, the visual input of both kittens is similar, but the coordination between movement and visual stimulation is not. One kitten perceives the environment as a result of its own movement while the other just passively perceives a signal structured by the movements of the other. The second kitten's movement is correlated to the visual patterns it perceives but it is not really producing the variation in sensory stimulation: it is *passively-coupled*. After a number of hours of exposure to the experiment, the kittens are faced with a task consisting in discriminating a visual cliff (suggestive of the ability of depth perception), and only the kitten that was able to move when experiencing visual stimulation is capable of successfully accomplishing the task. For Held and Hein this proves that coordination between self-produced movement and its concurrent visual feedback is necessary for

the development of visually guided behaviour. The co-occurrence of visual stimulation and movement produced by other means (although having the exact same structure) is not sufficient.

The experiment developed by Held and Hein provides interesting insights into the emergence of cognitive autonomy and the role of sensorimotor coordination. Unfortunately, the dominating psychological current at the time, behaviourism, was not particularly interested in the internal nature of mental processes. The next dominating paradigm in cognitive science, as we will see below, would shift its focus towards an internalist perspective making difficult to take into account the role of sensorimotor processes. How could Held and Hein's experimental results be operationalized and modelled? For most of the computationalist models that dominated the imagination of cognitive scientists in subsequent decades (and up until now in many cases), there is no difference between perceiving a sensory pattern generated by the movement of the agent and passively perceiving the same input.

### 2.1.2 COMPUTATIONALISM AND INTERNALIST PERSPECTIVES

Early cognitive science was born in opposition to behaviourism and led by the idea that behaviour is mediated by something internal to the organism. One major landmark for the emergence of cognitive science was Noam Chomsky's review (Chomsky, 1959) of Skinner's 1957 book, *Verbal Behavior*. Chomsky points out that, contrary to Skinner's claims, language learning and use are too complex to be explained by stimulus-response mechanisms, even if these are chained in sequences. In contrast, he claims that mental structures (hierarchical architectures, internal rules and representations, generative algorithmic mechanisms that can produce grammatical structures) play an internal mediation between perception and action. Soon this argument was generalized in cognitive science from language to all complex cognitive behaviours (e.g. maze navigation and spatial representation), bringing representations at the centre of mental processes. At the time, research on perception suggested that our sensory access to the world was not very good (Marr, 1982), thus encouraging a view in which the brain has to deal with impoverished, probabilistic information and has to bring all its computational capacities to build accurate models of its changing environment (Wilson and Golonka, 2013). Since the world is not directly accessible in a reliable way, internally represented knowledge in the form of discrete symbols is a necessary mediator between sensory inputs and motor outputs. In parallel, the development of computers and the birth of Artificial Intelligence provided an empirical ground alongside experimental psychology. Quite often, the idea of representation was reduced to discrete symbolic representations that could easily be processed by the logical rules of computer programs. Thus, cognitive science was confined to identifying how the brain codifies, accesses and uses content in the form of discrete symbolic representations (Dietrich and Markman, 2003). This computationalist/cognitivist framework can be summed up in the following ideas: (1) representationalism – cognition requires the mediation of internal characterizations of the world – (2) computationalism – cognition consists in the manipulation of symbolic representations through a set of rules – and (3) functionalism

– particular cognitive mechanisms are defined in terms of their computational functions (i.e. the input-output relationships that computational rules establish). These assumptions conceptualize cognition as a process generated by a computational architecture (e.g. Fodor, 1983; Pinker, 1997; Carruthers, 2006).

However, the dominant position of the computationalist perspectives on cognition has been relaxed after the 1980s. The relative failure of symbolic Artificial Intelligence propelled new departures in the field, for example connectionist approaches to distributed parallel processing (Rumelhart and McClelland, 1986), which weakened some assumptions about the brain operating as a merely computational device. The moderate success of connectionist models inspired in neural structures allowed an approach to biology that, in tandem with the technological advances in the development of sophisticated tools of brain imagery, shifted the focus from abstract computational models to the research of brain functioning. However, this was only a first step. Advances in perception action research, especially the work of Gibson on direct perception (Gibson, 1966, 1986) revealed that our perception is not limited but allows a high quality online access to the world. Researchers on emotion (Damasio, 2000), perception (O'Regan and Noë, 2001), or action (Rizzolatti and Sinigaglia, 2007) concluded that cognitive processes may go beyond the brain and extend to the whole body and environment. A popular concept reflecting such embodiment was conceptualized by the notion of *extended cognition* (Clark and Chalmers, 1998), which included the body, tools and other minds as part of our cognitive processes.

However, although these changes widened the perspectives and possibilities in cognitive science and computational neuroscience, frequently they failed to imply a radical departure from the postulates of computationalism. Connectionist models allowed the explanation and reproduction of cognitive phenomena that had been ignored or left in the background (pattern recognition, associative memory, robustness respect to local disruptions...). Nevertheless, while the functionalist approach presents a subtle twist in connectionist models (also referred to as subsymbolic, since the symbols are distributed among the network parameters), connectionism keeps the core of computational assumptions: cognition is a computational and representational disembodied process decoupled from its biological grounding (Clark, 1989).

### 2.1.3 SITUATEDNESS: BRAINS, BODIES AND DYNAMICAL SYSTEMS

During the late 80s and early 90s a series of new perspectives in cognitive science arose from diverse disciplines (e.g. robotics, psychology, philosophy or artificial life) that started to bring into question some of the principles and methods of representational functionalism. One of the pioneers of such approaches was Rodney Brooks, who argued that Artificial Intelligence had been unnecessarily focused on the issue of representation (Brooks, 1991). Against this, he proposes a robotic approach based on loosely coupled modules able to produce behaviour without a central controller. Interestingly, rather than communicating to each other to a large degree, every module produces behaviour (avoid objects, explore, wander, and so forth) by being directly connected to the world through a perception and action loop (Brooks, 1985).

The different modules are coordinated in a decentralized manner by feeding its outputs as inputs of other modules, or via inhibition links between modules that block certain inputs or outputs when activated. The framework presented by Brooks differs from the previous connectionist framework at least in three important aspects: (1) a fundamental aspect of the robots is that they are continuously coupled to the world (whereas connectionist models at the time were mostly disembodied), (2) the system is not composed of homogeneous nodes whose particular connectivity instantiated a function. Instead, behaviour-centred modules are combined in a *subsumption architecture* and (3) while connectionist models were usually assumed to incarnate a distributed representation of their perceived world, Brooks claims that representations play no role in the behaviour of his robots and only appears in the mind of the observer.

The pioneering work of Brooks was followed by a wave of research and development in robotics that deepened criticism of computationalism and gave rise to the field of autonomous or situated robotics (Maes, 1991), that by opposing rule-based architectures, emphasizes instead direct coupling between sensing and action and distributed and decentralized dynamic interaction with the environment. Between 1992 and 1993, three research groups (at the EPFL in Lausanne, COGS at the University of Sussex and at the University of Southern California) reported promising results from experiments on artificial evolution of autonomous robots (Cliff et al., 1992; Lewis et al., 1992; Floreano, 1993). Their success triggered a wave of activity in labs around the world focused on the analysis of neural networks designed by artificial evolution in simulations of situated agents. The qualitative change responded to the fact that the network was not designed first following representational or functional criteria (and only after inserted in an interactive domain), instead it was designed through evolution in a recurrent process of agent-environment interaction.

The 90s also brought an intense emphasis in dynamical systems models and analyses. The change displacing computational and representational conceptions of cognition in favour of dynamic and interactive ones was not exactly new, since it can be traced back to early cybernetics (Ashby, 1954; Walter, 1963; Powers, 1973), neuroscience (Holst, 1973) phenomenology (Merleau-Ponty, 1942) and pragmatism (Dewey, 1896, 1922). However, during the 90s it propelled a strong paradigmatic shift among a variety of fields: autonomous robotics (Brooks, 1991), adaptive behaviour (Beer, 1990, 1997), coordination dynamics (Kelso, 1995), neuroscience (Skarda and Freeman, 1987), developmental psychology (Thelen and Smith, 1994) and philosophy of mind (Port and Gelder, 1995; Clark, 1997). Most of this change of paradigm has been labelled under the term of the *dynamical systems approach* to cognitive science (Chemero and Richardson, 2014).

In a nutshell, dynamical systems approaches propose two central contributions to cognitive science: (1) that cognitive mechanisms (neural or otherwise) could be better effectively modelled and understood in terms of dynamical systems (Haken, 1978; Kelso, 1995; Freeman, 2001) instead of symbolic representational algorithms (e.g. Fodor, 1983; Pinker, 1997; Carruthers, 2006) and (2) that cognitive behaviour could emerge out of recurrent sensorimotor

loops in a self-organized manner, without the need for explicit encoding and planning on the side of the agent. And yet the relationship between both contributions remains relatively under-explored: how does the self-organization of behaviour change the dynamical properties of brains? What is lost when we study brain dynamics in isolation from the sensorimotor loops they are naturally embedded in?

#### 2.1.4 NOT-SO-EMBODIED DYNAMICAL APPROACHES TO COGNITION AND NEUROSCIENCE

The apparent disconnection between these two issues is patent given that when some fields have just focused on one of these contributions neglecting the other. For example, some of the latest progress at both mechanistic (neurodynamic) and behavioural levels of dynamic modelling is related to *oscillatory dynamics*. Interactions between oscillatory components (neurons, brain regions, limbs or humans interacting with each other) are studied in terms of synchronization and phase-difference at various scales where macroscopic variables provide indexes of emergent collective behaviour (Strogatz, 2004; Buzsaki, 2006). Oscillations are ubiquitous in nature, from planetary motion to circadian rhythms (Pittendrigh, 1960), from predator-prey population dynamics (Lotka, 1920) to chemical systems (Kuramoto, 1984). Oscillatory activity is also present at different levels of the nervous system (Freeman, 2001). At the individual neural level, neurons undergo cyclic alterations on their membrane potential following different dynamical regimes depending on the cell properties (Izhikevich, 2006). At higher levels, global oscillations are observed as a collective phenomenon generated by groups of neural cells that fire synchronously (entrained by pacemaker cells or as a result of recurrent network activity with inhibitory and excitatory connections; Buzsaki, 2006). Different aspects of large-scale brain oscillatory activity (e.g. self-organization of emergent patterns, synchronization and oscillatory rhythms) have become a common explanatory resource in behavioural and cognitive neuroscience. Some of the phenomena that have been explored from this approach include the binding of the different perceived features of an object (Phillips and Singer, 1997), the representation of position information in navigation tasks (O'Keefe and Recce, 1993), attention (Deco and Thiele, 2009), memory (Jensen et al., 2007) and conscious experience (Crick and Koch, 1990; Engel et al., 1999; Varela et al., 2001b).

Despite the significant progress recently achieved by investigating oscillatory dynamics in cognitive neuroscience, existing theoretical frameworks and models are generally developed without taking into account sensorimotor dynamics and, and frequently are limited to the establishment of oscillatory correlations arising after a given stimulus onset. For example, the analysis of neural synchronization generally assumes dynamical mechanisms, but at the same time it typically assumes neural synchronization to play a key role in the generation of representations of sensory inputs. This idea is based on observations of how spatially separated cells responding to the same input display strong synchronization, and how this synchronization is reduced when cells respond to independent stimuli moving in different directions (Engel and Singer, 2001). However, evidence favouring this internalist view of neural synchronization has been obtained through equally internalist experimental designs. For example, many

experiments in perception and neural synchronization have been conducted analysing cortical and subcortical neurons in paralysed anaesthetized cats (Gray et al., 1989; Engel et al., 1991; Freiwald et al., 1995; Brosch et al., 1997; Brecht et al., 1999; Castelo-Branco et al., 2000) or cortical cells in anaesthetized monkeys (Livingstone, 1996). Moreover, behavioural experiments performed in awake monkeys analysing temporal encoding of information related to different tasks, has involved monkey training with simple input-output tasks in which the monkey has to fixate its attention on a sensory stimulus and provide a response after a cue was given (Abeles et al., 1993; Kreiter and Singer, 1996; Prut et al., 1998). As a result, the focus on oscillatory brain dynamics is often centred on those aspects of oscillatory activity within the brain that might carry information about a particular stimulus, without considering the coupled brain-body-environment dynamics. This is even true for non-representational approaches to cognition that acknowledge the theoretical relevance of situated cognition but conduct most of their studies in search of cognitive correlates in oscillatory brain activity leaving aside the potential effects of the sensorimotor coupling (e.g. Skarda and Freeman, 1987; Varela et al., 2001b).

On the modelling side, computational simulations are generally built without considering the body and the environment and often assuming a representational theory of brain function. That is, they assume that the main job of the brain is to create a representation or model of the environment, and focus on neuronal mechanisms capable of supporting the processing of such a model. For example, in a recent review paper, Andy Clark depicts a wave of work around Bayesian predictive neurodynamics (Clark, 2013). Here, Clark argues that brains behave essentially as prediction machines, where neural networks constantly attempt to match incoming sensory inputs with top-down expectations or predictions. The Bayesian predictive approach hypothesizes the existence of a hierarchical generative model that aims to minimize prediction error within a bidirectional cascade of cortical processing. Computationally, the appeal of this perspective lies in the fact that all that needs to be passed forward through the system is the error signal (which is what remains once predictions and driving signals have been matched). In these models it is therefore the backward (recurrent) connectivity that carries the main information processing load. Neural architecture is depicted as a kind of duplex architecture: one that at each level *combines* both traditional representation of inputs with processing of prediction errors. While ‘representation units’ generate the predictions necessary to instantiate a potent context-dependent dynamical system, ‘error units’ correct the error signals to continuously adapting brain dynamics to different situations. Clark’s view brings a number of interesting elements from a dynamical systems perspective. He proposes that it is not only sensory input that is predicted, but he rescues Karl Friston’s perspective to the framework, generalizing hierarchical predictive processing to include action, or ‘action-oriented predictive processing’ (Friston et al., 2010). Motor action would be predicted in the same way as perceptual inputs but with the difference that motor error signals would be corrected by movements instead of neural activity in an intimate coordination between action and sensation. Moreover, Friston’s free energy formulation postulates that prediction errors

are quantified by means of entropy, allowing the system to minimize prediction error by self-organizing to minimize the entropy levels in the system.

What is important here is that there is still a pervasive kind of ontological divide between agent and environment: what lies within the agent is a representational device, an inference engine, and outside lies the world to be represented. The interaction between the two is limited to prediction and error correction. The fine grained temporal structure of interaction including the sensorimotor entrainment of brain dynamics and the embodied shaping of experience is left out of the picture. Even the most dynamicist versions of Bayesian predictive approaches still assume that the brain is the indisputable centre of cognitive life. Ultimately, everything consists in generating predictive models *inside the brain* – i.e. representations in terms of probability distributions – about the outside world. These representations are stored in parameters distributed across a hierarchical neural architecture which are tuned to minimize prediction errors about the body and the environment. In that sense the framework proposed by Clark, as pointed out by Anderson and Chemero (2013), does not understand prediction in the broader sense of correlations (which may fit well with a dynamicist embodied perspective) but in the sense of abductive inference. Thus, cognition would consist in inferring the causes of our observations and predicting future ones. This inevitably leads to exclusively internalist processes, in which the world we experience is a projection of the predictive models generated by our minds. This assumption neglects the possibility that some aspects of the environment can be accessed directly from the constraints of the agent-environment interaction at an ecological scale (Calvo et al., 2012, see next section).

In general, the notion of embodied cognition is becoming increasingly popular, and it is widely accepted that cognition can be modulated or biased by the senses (Llinás et al., 1998), the body (Eerland et al., 2011) and the environment (Adam and Galinsky, 2012), or that cognitive processes are grounded in the biological substrate and our computational capacities are not abstract but shaped by the peculiarities of our bodies (Lakoff and Johnson, 1999). Still, these approaches do not break completely with the dominating internalist paradigm, and still characterize cognition as a fundamentally representationalist, computationalist and functionalist process, even when cognitive functions can be composed by parts of the body or the environment. Thus, under these assumptions embodied cognition does not follow through on the necessary consequences of allowing cognition to involve more than the brain (e.g. Dietrich and Markman, 2003).

### 2.1.5 RADICAL EMBODIMENT AND SENSORIMOTOR THEORIES OF COGNITION

As we have shown, throughout the history of cognitive science, problems posed by previous dualist philosophical frameworks were solved by relaxing some assumptions about how cognition may be bound to particular modules located in the brain. In contrast, other approaches proposed that cognition should be a more distributed process involving a network of processing units (connectionism), computation through the body and the environment (embodiment) and dissolution of representations in favour of emergent dynamical processes

(dynamical systems approaches). Nevertheless, these approaches have frequently been unable to escape a basically internalist perspective on cognition and giving the brain a privileged position as the domain within which cognition is confined (even if it can take advantage of some processes going in the body or the environment). In this context, some authors have proposed the need to completely abandon of the representational nature of cognitive architectures. This has given rise to what has been labelled as the *radical embodiment* approach to cognitive science, based on both dynamical systems theory approaches to cognition and ecological psychology (Chemero, 2009).

#### SENSORIMOTOR CONTINGENCIES

Sensorimotor coordination implies more than the statement that sensory input, through its influence on brain dynamics, creates an action that, in turn, produces a change that leads to a new perceptual state. The central claim of situated approaches to cognitive behaviour is that the agent-environment coupling shapes brain dynamics in a manner that is essential to behavioural or cognitive functionality (Steels, 1990; Beer, 1995a; Clark, 1997). In other words, macroscopic functional behaviour (e.g. intentional grasping or perception) emerges from microscopic sensorimotor dynamics (e.g. proprioceptive and visual feedback in grasping or saccadic movements in visual perception). More recently, the Sensorimotor Contingency Theory (O'Regan and Noë, 2001) has defended the notion that what is constitutive of perceptual awareness (and, it could be argued, of other cognitive states) is not a specific internal state of an agent, but the structure of sensorimotor contingencies. To perceive is to act in a specific manner that brings forth the structure of sensory changes in relation to the activity of the agent. To see or to perceive is something that is done and relies on the very sensorimotor coupled dynamics of an agent. In the words of Kevin O'Regan and Alva Noë, 'to have a sensation is to exercise one's mastery of the relevant sensorimotor contingencies and in this sense to be 'attuned' to the ways in which one's movements will affect the character of input' (O'Regan and Noë, 2001, p.84). They use the example of Braitenberg *vehicles* (Braitenberg, 1986), in which a wheeled vehicle is equipped with two front mounted light sensors, the left sensor being linked to the right wheel and the right sensor to the left wheel. As result of its wiring, the vehicle drives itself toward light sources exhibiting a phototactic-like behaviour:

the ability of this extremely simple mechanism to 'sense light' does not consist just in the fact of activity in the sensor. Light causally affects the sensors, to be sure. But the vehicle's ability to sense the light consists in the ways in which it makes use of the raw stimulation of the light on the sensor [...] The sensation does not occur in the brain any more than it occurs in the sensors of the vehicle. The occurrence of sensation is a system or creature-wide phenomenon. Mastery of sensorimotor contingencies may be neurally encoded, but this mastery does not itself reside in the brain. To the extent that it makes sense to speak of mastery

residing anywhere, then it resides in the creature as a whole, in the whole neurally enlivened body (O'Regan and Noë, 2001, p.84-85).

An illustration of this may be the sensation of softness one experiences when manipulating a sponge (Myin, 2003). We perceive the softness of the sponge because when we press it, it squishes. 'Having the sensation of softness consists in being aware that one can exercise certain practical skills with respect to the sponge: one can for example press it, and it will yield under the pressure. The experience of softness of the sponge is characterized by a variety of such possible patterns of interaction with the sponge, and the laws that describe these sensorimotor interactions' (O'Regan et al., 2005).

#### ENSLAVING OF THE BRAIN BY THE AGENT-ENVIRONMENT SYSTEM

How can perceptual awareness or a cognitive state emerge from a structure of sensorimotor interactions? In Chapter 1 we presented the notions of collective variable, order parameter and the slaving principle, and how nonlinear dynamics can provide the analytical tools to understand the circular relation between top-down and bottom-up influences between microscopic and macroscopic variables *i.e.* However, this approach has been frequently applied to brain activity alone. For example, Haken states that while the chemical and electrical activities of neurons constitute the microscopic variables, 'order-parameters are ultimately the thoughts' (Haken, 1978, p.15). That is, order parameters are strictly defined in terms of neural activation. However, a careful analysis of Haken's work shows that his approach assumes an internalist perspective of cognitive self-organization. For example, he describes the 'emergence of meaning' in pattern recognition in purely internal terms (Haken, 1995). According to Haken, semantic information of different patterns arises because they are able to produce an effect in the dynamical landscape of neural networks, 'opening or closing of perception attractors can be achieved by changing attention parameters [...]. Therefore, we may say that meaning is attached in each case to a specific order parameter' (Haken, 1995, p.42).

A different approach has been proposed from ecological psychology (Reed, 1996; Warren, 2006; Chemero, 2009). Ecological psychologists do not assume that order parameters governing cognitive behaviour are in the brain but at the level of the agent-environment system. Therefore, organized behaviour does not need to be attributed to a centralized controller, or internal models or representations, but instead they present models in which control is distributed over the whole agent-environment behavioural dynamics. 'Interactions between the agent and environment under physical, informational, and task constraints give rise to the behavioural dynamics, an evolving vector field that reciprocally captures the agent's behavior' (Warren, 2006). That is, it is the existence of supra-brain order parameters (in the form of behavioural dynamical patterns) which control behaviour. This approach has been proven to be quite fruitful at least at a behavioural level (Warren, 2006; Dotov et al., 2010; Gibson, 1986; Lee, 1976). However, ecological psychology perspectives typically rule out the neurodynamics of the system and just focus on finding the higher behavioural order parameters that govern the interaction.

The assumption underlying these perspectives, is that neural substrate is enslaved by higher-order dynamical patterns of motor behaviour at the level of behaviour in an environment (Dotov, 2014). This idea was proposed by Van Orden et al. (2012) in his *blue collar brain* hypothesis, suggesting that since the timescales of body and environment are slower than the timescales of the brain, the latter must be constrained and controlled by slower flow of context and behaviour, limiting the degrees of freedom about what can happen next<sup>1</sup>. In a similar vein, Fuchs (2011) conceives the brain as a large set of latent open loops or attractors coupled to the environment, and the sole role of the brain is to mediate and modulate the cycles of embodied interaction (i.e. not to produce its content). This idea is supported by Dotov's provocative claim stating that we may already 'know most of what we need to know about the brain, there is no hidden neural code to be decoded' (Dotov, 2014, p.7). Reviewing the successes of the Blue Brain Project (Markram, 2006), which used the computational capacities of IBM's Blue Gene supercomputer for building biologically accurate models of the brain, Dotov claims that detailed dynamical modelling of neuronal cortical columns displays patterns with the form of emergent collective phenomena at the level of the whole column, such as propagating waves of synchronous activation. Moreover, these patterns are similar to the ones generated with relatively simple neurodynamical models running in a regular laptop (Izhikevich, 2006). Dotov's proposal is that neural tissue by itself does not exhibit much more than spontaneous synchronization and other phenomena typical of nonlinear dynamical systems. Therefore, complex behaviour does not arise from a complex 'self complicating' neural structure. Instead, it arises when neural tissue is coupled into a field of environmental and bodily activity. That is, the brain has not evolved to 'generate' complicated behaviour but to coordinate with the body and environment in complex ways.

### 2.1.6 FILLING THE GAP

We have described how the dominant views in cognitive systems have been mainly focused on an internalist perspective, where the processes constitutive of cognition are bound to the limits of the brain. In contrast, different sensorimotor and ecological approaches have proposed a radically different perspective in which the emergence of agent-environment order parameters is the constitutive element of cognition. Filling the gap between the study of neural and situated sensorimotor dynamics is essential if we want to understand the nature of neurocognitive autonomy. However, despite the repeated emphasis on the importance of sensorimotor coupling for neurodynamic approaches<sup>2</sup>, current understanding of brain oscillatory dynamics mostly focuses on 'passive' conditions. The dynamical properties of oscillatory networks

---

<sup>1</sup> This assumption about the timescales of body and brain has not been proven. However, it has been proposed that the body, as the brain, is self-organized in a (multi)fractal tensegrity structure comprising from individual cells, to muscular-skeletal structures which may embed the brain (Turvey and Fonseca, 2014)

<sup>2</sup> By neurodynamic approaches we refer to dynamical system approaches to the understanding of neural activity (Gelder, 1998; Freeman, 2001; Buzsaki, 2006, etc.). There are very few examples of these types of models that exploit sensorimotor coupling and almost none for oscillatory models (some recent exceptions include Santos et al., 2011; Moiola et al., 2010).

(even when studied within the context of behavioural or cognitive neuroscience, see Strogatz, 2004) are deduced from mathematical and computational models that have constant or no input at all, and the effect of sensorimotor or situated dynamics on the oscillatory properties of such networks is rarely considered.

The goal of this chapter is to make a theoretical contribution in the direction of explicitly quantifying the difference between the dynamics that results from isolated versus situated oscillatory controllers, and those that result from actively versus passively coupled systems.

Inspired by Held and Hein's experiment above in which two kittens were compared, one of them moving freely and the other being only allowed to passively receive a visual pattern product of the movement of the other kitten (Held and Hein, 1963), we have designed an experiment to study a similar situation in a minimal dynamical system<sup>3</sup>. Our objective is to determine the simplest system that can reproduce the difference between these two modes of perception (coupled to the environment and 'passively coupled'). For doing so, we have chosen the Haken-Kelso-Bunz (HKB) model presented in section 1.2.2 as a paradigmatic example of oscillatory dynamics and behaviour to address these questions. There are a number of good reasons to choose the HKB model. First, the HKB model is simple enough to be treated analytically. Second, it has been used both to model behavioural phenomena and to model brain dynamics. Finally, to our knowledge, no variation of the HKB model exists that uses it as a controller of a sensorimotor system, and no analytical study exists of a comparison between the dynamics of the HKB studied in isolation (with a parametric analysis) and its dynamics under sensorimotor loop conditions (with few exceptions as, for instance, Kelso et al., 2009).

By filling in this gap we can address the following questions: How does the HKB model change its properties when it is situated (i.e. under closed-loop sensorimotor coupling in a spatial environment)? Is there any qualitative change that comes out of this coupling? Can the behavioural properties of an oscillatory 'brain', or controller, be inferred from the study of the brain in isolation or under constant input, or even from variation of the input corresponding to those found in the coupled system? In this chapter, we explore the answer to these questions by modelling a 'situated HKB' and analytically solving the coupled agent-environment system and comparing it with isolated and passively coupled conditions. The structure of the rest of the chapter is as follows: (1) First, we introduce the well known HKB model and the coordination dynamics framework; (2) next, we characterize the notion of dynamically coupled and spatially situated system and present a novel extension of the HKB model with sensorimotor embodiment that we call the *situated HKB* model; (3) then, we analytically solve a particular case of the situated HKB model performing a gradient climbing task in a 2D environment. Subsequently, (4) we compare the obtained dynamics of the coupled system with the dynamics of a *decoupled HKB* and with a *passively-coupled HKB* model for an equivalent parametric analysis. Qualitative changes between the eigenvalues describing

---

<sup>3</sup>Although the developmental aspect of Held and Hein's experiment was a fundamental part of it, for simplicity, our design only considers the difference between a system coupled to the environment and one that is only 'passively coupled' to it.

the three HKB-system dynamics are identified, as well as experimental measures characterizing the transformation of the complete phase space of the agent produced by sensorimotor coupling. Finally, (5) we discuss some implications for the study of oscillatory brain dynamics.

## 2.2 THE HKB MODEL

We have chosen the HKB model because (1) it is one of the simplest models describing coordination phenomena, and (2) despite its simplicity it has been widely used as representative model of dynamicist approaches and it describes a rich spectrum of possible behaviours. Thereby, the HKB model has been a driving example of coordination dynamics, describing the behaviour of two non-linearly coupled oscillators. The model was originally formulated in 1985 to explain experimental observations in the relative phase dynamics of bimanual coordination (Haken et al., 1985) but it has also been shown to capture the coordination dynamics of different behavioural (Kelso, 1995), neural (Jirsa et al., 1994) and social (Oullier and Kelso, 2009a) phenomena. Using the language of synergetics (order parameters, control parameters, instability, etc., see Haken, 1978), the HKB describes a simple nonlinearly coupled dynamical system that captures the self-organized behaviour of two generic coordinated nodes or units (Fuchs et al., 1995). More specifically, the HKB model was conceived to provide insights into: (1) the formation of ordered states of coordination; (2) the multistability of those states; and (3) the conditions that give rise to switching among coordinative states (Kelso, 1995). Moreover, the HKB model has been proven to describe fundamental features of self-organization such as multistability, phase transitions and hysteresis (Kelso, 1995).

Here we use the ‘extended HKB’ equation (Kelso et al., 1990)<sup>4</sup>, in which a system composed of two coupled oscillators is reduced to a single equation where the main variable is the relative phase between the two oscillators, and whose dynamics are shaped by the difference between the natural frequency of the oscillators and their coupling strength:

$$\dot{\varphi} = \Delta\omega - a \cdot \sin(\varphi) - 2b \cdot \sin(2\varphi) \quad (2.1)$$

The relative phase or phase difference,  $\varphi$ , represents the order parameter or collective variable that emerges from lower-level interactions of the two coupled oscillators,  $a$  and  $b$  are the coupling coefficients between the two oscillators, and  $\Delta\omega$  is the difference between their intrinsic frequencies. Despite its simplicity, this equation captures a wide range of self-organized phenomena. Different combinations of the control parameters  $a$ ,  $b$  (or rather  $b/a$ ) and  $\Delta\omega$  give rise to different collective behaviours. For example, when shifting the value of  $\Delta\omega$  while the values of  $a$  and  $b$  are held fixed, the system experiences phase transitions between three different modes of behaviour: monostable, bistable and metastable (Figure 1.6).

---

<sup>4</sup> We shall hereafter use the terms ‘HKB model’ or ‘HKB system’ to refer to the ‘extended HKB model’ (Kelso et al., 1990) rather than the original (and simpler) HKB model (Haken et al., 1985). The reason for this is that we are going to distinguish additional situated, decoupled and passively-coupled versions of the extended HKB model and names such as ‘passively-coupled extended HKB’ would interrupt the flow of the text.

The HKB has been used to model different kinds of coordination phenomena but rarely used as a controller of an embodied agent. To be fair to the HKB model, this was never intended by the original authors. The HKB model (and its extended version) was rather conceived to describe the behavioural dynamics at the macroscopic level (i.e.  $\varphi$  representing the collective variable of phase difference between two ‘behaving’ oscillatory components, like fingers, oscillating armchairs in social coordination, etc.). It can be said that the HKB model was designed to capture the global agent-environment dynamics, not any explicit behaviour generating mechanism that is coupled through sensors and motors to an environment. In addition, the HKB was also used to model inter-areal coordination in the cortex (Tognoli and Kelso, 2009), ignoring the potential influence of the coupling between brain and environment. In short, previous uses of the HKB model involved either full behavioural phenomena or ‘isolated’ brain dynamics. Thus, there is a theoretical modelling gap that remains under-explored: namely the HKB as a controller of an agent that can modulate the control parameter (influenced by sensory input) through the behaviour it generates when embodied in a robot.

### 2.3 THE SITUATED HKB MODEL

#### 2.3.1 SENSORIMOTOR EMBODIMENT OF THE HKB EQUATION

In this section we describe what we have called the situated HKB system: a robotic model where the HKB equation describes the ‘neural system’ of the agent which is embodied with sensors and motors and, in turn, situated in an environment. The agent has a circular body of radius  $R$  with two diametrically opposed motors (see Figure 2.1.a), that can move forward or backwards with different velocities in a 2D arena, and a single sensor that provides an input to the HKB neural controller.

Thereby, the HKB equation provides the macroscopic description of the dynamics of the two coupled oscillators. It allows us to describe the behaviour of the situated HKB system in the following manner: (1) the agent has a ‘brain’ where two regions (e.g. sensory and motor cortex) oscillate with their corresponding natural or intrinsic frequency (the difference between these frequencies is expressed by  $\Delta\omega_0$ ); (2) when a sensory input  $l$  modifies the natural frequency of one of the oscillators (e.g. sensory cortex), the frequency difference term changes<sup>5</sup>; (3) since the frequency difference term is the control parameter of the phase difference,  $\varphi$ , we can consider that the situated HKB agent modulates its control parameter through sensorimotor contingencies: i.e. through the sensory changes that result from motor actions and the displacements they generate.

The dynamics of our agent’s controller is driven by the following equation:

$$\dot{\varphi} = (\Delta\omega_0 + l) - a \cdot \sin(\varphi) - 2b \cdot \sin(2\varphi) \quad (2.2)$$

---

<sup>5</sup>The term  $\Delta\omega$  of the original extended HKB equation has been substituted by the term  $\Delta\omega_0 + l$ , where  $l$  is the sensory input term whose effect is mathematically equivalent to changing the natural frequency of one of the oscillators (see Figure 2.1.a)

It is assumed that the agent is situated in a two-dimensional environment where a radial gradient of a stimulus  $\eta$  is present with its peak on the origin of coordinates (one can interpret this environment in different manners, e.g. as a light source or a chemical gradient diffused in all directions from a central point with radial symmetry).

With regard to the ‘sensory system’ of the model, since the agent lives in a world of gradients, we have designed its sensor not to perceive the absolute amount of stimulus present in the environment but its change. Thus, the agent is sensitive to changes in  $\eta$  mediated by a sensitivity factor  $s$  that characterizes the sensor gain:

$$l = \dot{\eta} \cdot s \quad (2.3)$$

where the perceived intensity signal decays linearly with the distance  $d$  to the source of the gradient. With respect to the ‘motor system’ of the model, we define the activations of the motors as functions of the state of the controller,

$$\begin{aligned} M_r(\varphi) &= m \cdot \cos(\varphi) \\ M_l(\varphi) &= m \cdot \cos(\varphi + c) \end{aligned} \quad (2.4)$$

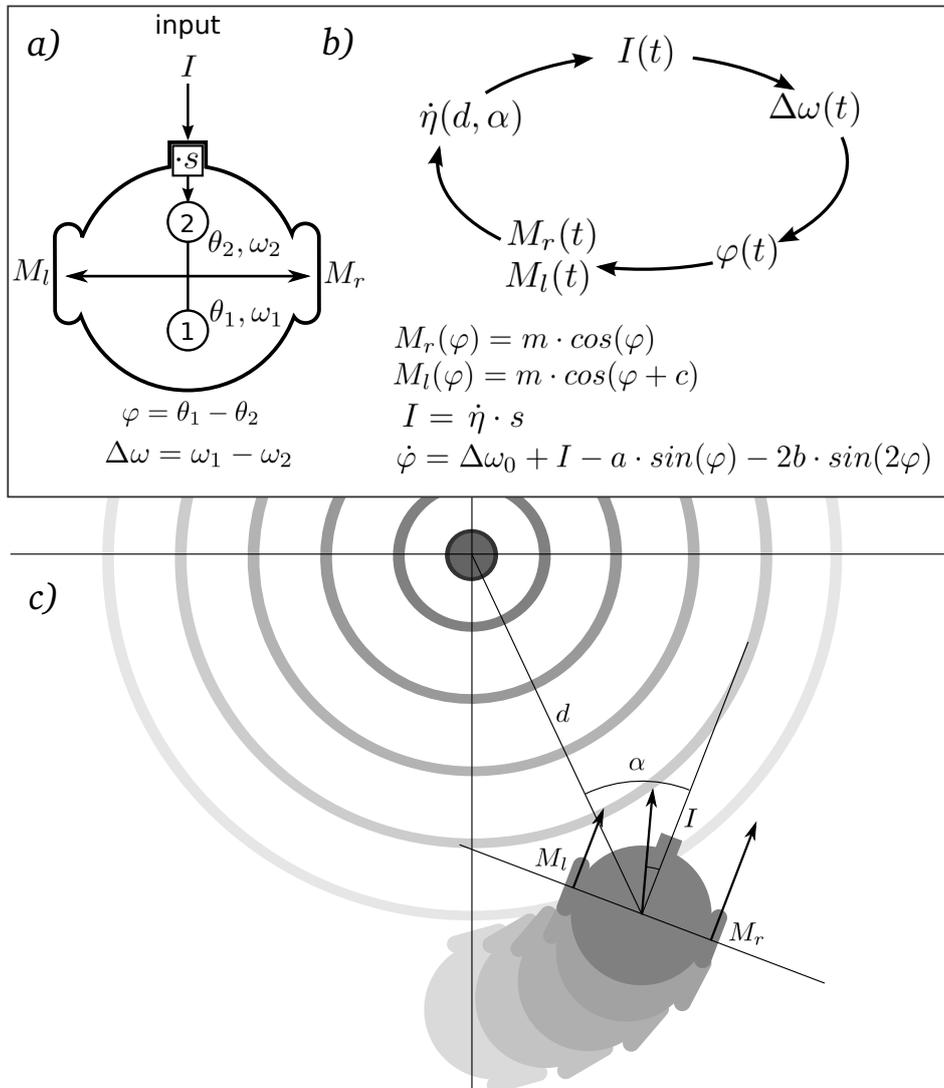
where  $M_r$  and  $M_l$  represent the right and left motors respectively,  $m$  is a speed parameter, and  $c$  is a bias parameter that breaks the symmetry between the right and left motors.

Thus, brain-body-environment coupling can be understood as a process that repeats a cycled-course of four stages involving successive transformations of  $l(t)$ ,  $\Delta\omega(t)$ ,  $\varphi(t)$ ,  $M(\varphi(t))$  and back to  $l(t)$  (see Figure 2.1.b).

#### 2.4 THREE-DIMENSIONAL DESCRIPTION OF THE SITUATED-HKB MODEL

To fully describe the movement of the agent we include an additional variable describing the angle  $\alpha$  of the agent’s orientation relative to the peak of the gradient (see Figure 2.1.c). The behaviour of the *situated* HKB agent is completely addressed by the variable  $\varphi(t)$  and the position of the motors of the agent  $M(\varphi(t))$  at each time  $t$ . The process can be characterized as follows: (1) with the movement of the agent in the environment (variation of  $d$ ) the sensor receives a new input  $l$ , (2) the input influences the firing rate of the oscillators, changing the frequency difference between the HKB nodes (variation of  $\Delta\omega$ ); (3) these frequency differences translates into a change in the phase difference between the oscillators (variation of  $\varphi$ ) and, finally, (4) the new value of the phase difference changes the state of the motors (variation of  $M(\varphi)$ ) moving the robot (variation of  $\alpha$ ) and starting the cycle again.

As a direct result of assuming radial symmetry in the problem: (1) we can use a polar coordinate system (distance  $d$  of the agent to the centre of the gradient, and angle  $\alpha$  of orientation of the agent relative to the peak of the gradient) as the reference frame, and (2) it is considered that the variation of the gradient in terms of the polar coordinates does not depend on the angle,  $\dot{\eta}(d, \alpha) = \dot{\eta}(d)$ .



**Figure 2.1: Situated HKB agent.** a) Structure of the agent, consisting in a sensor, two oscillatory controllers and two motors. b) Sensorimotor loop of the agent. c) Representation of the agent interacting with its environment. The position and orientation of the agent with respect to the centre of the gradient are represented through the variables  $d$  and  $\alpha$ .

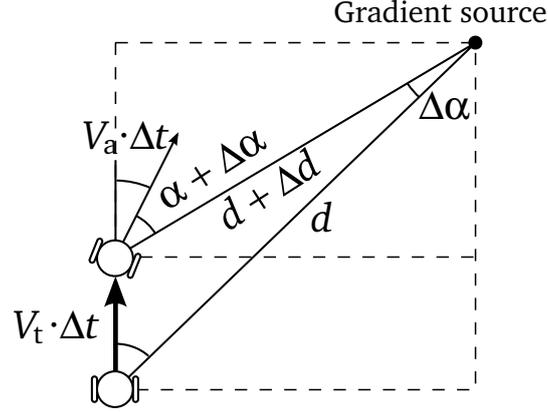


Figure 2.2: Schema of the situated HKB system equations. The equations of the movement of the agent can be deduced from this schema

- ASSUMPTION 1: The choice of polar coordinates ( $d$  representing the distance to the source and  $\alpha$  representing the relative angle between the agent's orientation and the source), instead of cartesian coordinates allows us to reduce the dimensionality of the model.

Assuming that the agent's mass is small enough to be discounted (in order to avoid inertial resistance), we can describe the speed of the agent as follows: the translational speed of the robot  $V_t$  is calculated as the vectorial average of the motor velocities, and the angular speed  $V_a$  as the difference of the motor velocities divided by the body diameter.

$$\begin{aligned} V_t(\varphi) &= (M_r(\varphi) + M_l(\varphi))/2 \\ V_a(\varphi) &= (M_r(\varphi) - M_l(\varphi))/2 \cdot R \end{aligned} \quad (2.5)$$

Assuming the tangential and angular velocities of the robot  $V_t$  and  $V_a$  are known, we can deduce the movement of the agent in terms of polar coordinates (see Figure 2.2):

Considering  $\Delta t, \Delta d, \Delta \alpha \rightarrow 0$ , we get

$$V_t \cdot \Delta t \cdot \cos(\alpha) + (d + \Delta d) \cos(\Delta \alpha) = d \quad (2.6)$$

from which, the next equation is easily obtained

$$\dot{d} = -\cos(\alpha) \cdot V_t \quad (2.7)$$

Similarly, if we take

$$V_t \cdot \Delta t + (d + \Delta d) \cdot \cos(\alpha + \Delta \alpha + V_t \cdot \Delta t) = d \cdot \cos(\alpha) \quad (2.8)$$

we find that

$$V_t + d \cdot \sin(\alpha) \cdot (\dot{\alpha} + V_a) + \dot{d} \cdot \cos(\alpha) = 0 \quad (2.9)$$

Putting together (2.7) and (2.9), we obtain the movement equations:

$$\begin{aligned} \dot{d} &= -\cos(\alpha) \cdot V_t \\ \dot{\alpha} &= \sin(\alpha)/d \cdot V_t + V_a \end{aligned} \quad (2.10)$$

where we know that  $V_t$  and  $V_a$  depend on the speed of the motors:

- **ASSUMPTION 2:** For simplicity, it is considered that  $\eta = d_0 - d$ , where  $d$  is the distance to the centre of the gradient and  $d_0$  is the intensity of the stimulus at the centre of the gradient, which decreases linearly with  $d$ .

Since  $\eta$  and  $d$  are inversely proportional and radial symmetry allows us to dismiss the angle of the position of the agent respect to the gradient when we describe it in polar coordinates. The change of the perceived gradient only depends on the position of the agent and the gradient itself,

$$\dot{\eta} = \dot{F}(d, \alpha) = \partial F(d, \alpha)/\partial d \cdot \dot{d} + \partial F(d, \alpha)/\partial \alpha \cdot \dot{\alpha} \quad (2.11)$$

and the set of equations describing the system-environment coupling can be reduce to only three:

$$\begin{aligned} \dot{\varphi} &= \Delta\omega_0 + \dot{\eta} \cdot s - a \sin(\varphi) - 2b \cdot \sin(2\varphi) \\ \dot{\eta} &= \cos(\alpha) \cdot (M_r(\varphi) + M_l(\varphi))/2 \\ \dot{\alpha} &= \sin(\alpha)/(d_0 - \eta) \cdot (M_r(\varphi) + M_l(\varphi))/2 + (M_r(\varphi) - M_l(\varphi))/(2R) \end{aligned} \quad (2.12)$$

- **ASSUMPTION 3:** Without loss of generality, we choose  $m = 2$ ,  $\Delta\omega_0 = 1$ ,  $d_0 = 0$  and radius of the body  $R = 1$ , obtaining:

$$\begin{aligned} \dot{\varphi} &= 1 + \dot{\eta} \cdot s - a \cdot \sin(\varphi) - 2b \cdot \sin(2\varphi) \\ \dot{\eta} &= \cos(\alpha) \cdot (\cos(\varphi) + \cos(\varphi + c)) \\ \dot{\alpha} &= -\sin(\alpha)/\eta \cdot (\cos(\varphi) + \cos(\varphi + c)) + (\cos(\varphi) - \cos(\varphi + c)) \end{aligned} \quad (2.13)$$

where  $a$ ,  $b$ ,  $c$  and  $s$  are the parameters of the system.

## BEHAVIOURAL ANALYSIS

We have chosen a basic gradient climbing<sup>6</sup> task for our agent to solve. That is, we ask the agent to climb up a linear gradient and maintain itself as close as possible to the maximum peak. A simple trial-and-error hand-tuning of the parameters gives us combinations that perform the desired behaviour. We chose to adjust the parameters to  $a = 5$ ,  $b = 1$ , and  $c = 5$ , leaving the sensitivity parameter  $s$  unspecified in order to have one free parameter that allow us to explore different kinds of behaviour. This selection is arbitrary (except for the relation of  $a/b$ , which was chosen to ensure that the HKB is always in a monostable mode of functioning) but other combinations of parameters which result in gradient climbing behaviour lead us to similar results in the analysis. For the experiments, the value of  $s$  is defined in the interval  $[0, 15]$ .

For these parameters, we see that the agent displays different behavioural strategies depending on the value of its sensitivity parameter  $s$  (Figure 2.3): (1) for values of  $s \in [0, 2.4]$  displaying cycloidal strategies where the agent turns over itself with a corkscrew-like movement, to (2) spiral paths where the agent slowly climbs the gradient, when  $s \in [2.6, 15]$ . At the frontier between these two behavioural strategies, we find (3) a critical region ( $s \approx 2.5$ ) where the agent displays the most efficient gradient navigation (in terms of time and trajectory efficiency), taking a curved approximation path ending in a spiral-circular pattern around the peak of the gradient. These different behaviours are shown in Figure 2.3, where the efficiency of each gradient climbing strategy is computed with a parameter  $F_d = 1 - \frac{d(t=t_1)}{d(t=0)}$ , that represents how close the agent gets to the centre of the gradient in a given time (we use  $t_1 = 40s$ ).

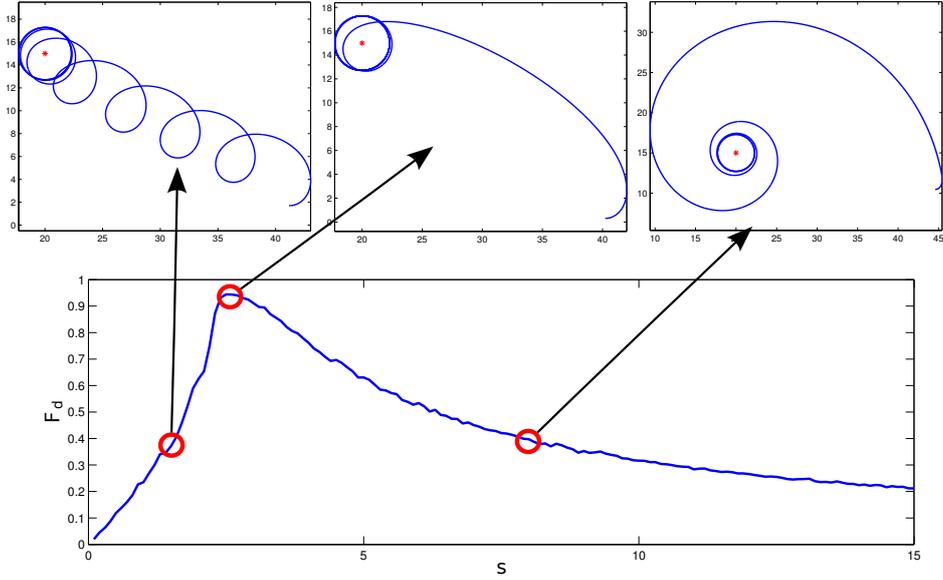
How do these behavioural strategies work? In the critical region ( $s \approx 2.5$ ) the agent maps the highest gradient ‘sensation’ with high activation of both motors, moving effectively towards the peak of the gradient. On the one hand, as we increase the value of  $s$ , sensory stimulation is more intense, so the agent needs an strategy to make the approach to the gradient peak slower in order to maintain a compensated activation between the motors. At the same time, as the value of  $s$  is decreased, the agent experiences difficulties maintaining an equilibrated velocity for both motors, having to turn around periodically to regain a trajectory where a high sensor input is perceived.

### 2.4.1 ANALYTICAL SOLUTION FOR THE SITUATED-HKB SYSTEM

We shall now analytically solve the coupled brain-body-environment system in order to understand the emergence of the qualitatively different kinds of behaviour that appear as we

---

<sup>6</sup>Gradient climbing is a minimal (yet not totally trivial) task, which is widespread in nature. Most of small scale adaptive behaviour occurs along chemical gradients. The microscopic world is a world of gradients (like thermal gradients or light gradients but mostly chemical gradients). The adaptive behaviour of small animals (e.g. *C. elegans*) and individual motile cells (e.g. bacteria but also animal cells migrating during development) is mostly a gradient-related adaptive behaviour. Navigating smell and heat gradients are also stereotypical adaptive tasks for higher animals. Moreover many instances of higher-level behaviour can also be interpreted as abstract gradient climbing (e.g. a human can move up a gradient of social popularity or economic wealth involving complex strategic decisions combined with an emotional gradient climbing of the perceived result of such strategies).



**Figure 2.3: Behaviour of the situated HKB system.** Different behaviours performed by the situated HKB system. We observe how different gradient climbing strategies arise depending on the value of  $s$ . For  $s = 1.5$ , the agent follows a cycloidal trajectory continuously turning over itself, for  $s = 2.5$ , the agent finds a direct path towards the peak of the gradient, and for  $s = 8$  the agent slowly approaches the gradient peak following a spiral path.

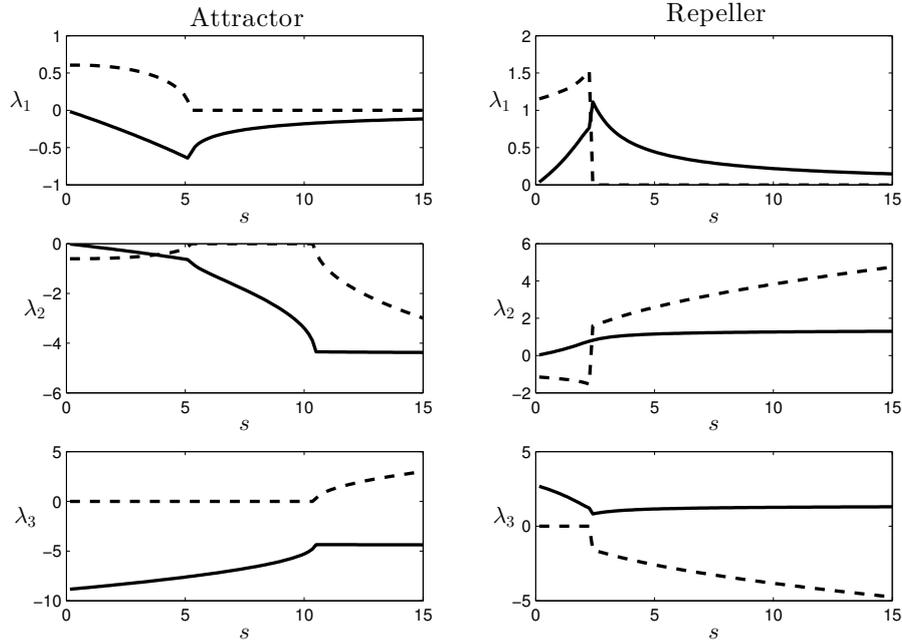
increase the sensitivity  $s$  of the agent. As usual, if we want to understand the behaviour of an artefact modelled by a dynamical system, we need to calculate the linearization of the system around its fixed points (Strogatz, 2001).

Thus, if we take the situated-HKB system of equations to be solved,

$$\begin{aligned}
 \dot{\varphi} &= (1 + \dot{\eta}(d, \alpha) \cdot s) - a \sin(\varphi) - 2b \cdot \sin(2\varphi) \\
 \dot{\eta} &= \cos(\alpha) \cdot (\cos(\varphi) + \cos(\varphi + c)) \\
 \dot{\alpha} &= -\sin(\alpha) / \eta \cdot (\cos(\varphi) + \cos(\varphi + c)) + (\cos(\varphi) - \cos(\varphi + c))
 \end{aligned} \tag{2.14}$$

where the parameter  $s$  (sensitivity) is used as a control parameter to analyse the solutions in our range of interest ( $s \in [0, 15]$ ), the two fixed points can be obtained as follows: (1) the first one is an attractor with values of  $\varphi, \eta, \alpha$  at  $(0.11, 2.28, -\pi/2)$  and (2) the second one is a repeller at  $(2.53, 0.43, \pi/2)$ .

Computing the Jacobian matrix of the system at the fixed points, and making an analysis of the eigenvectors and eigenvalues of the system, we get the behaviour of our dynamical system around the regions of its state space that bear qualitative significance. Figure 2.4 illustrates the range of different values of the eigenvalues (denoted by  $\lambda_1, \lambda_2$  and  $\lambda_3$ ) at each of the fixed points, depending on the parameter  $s$  (that corresponds to different observed behavioural patterns for gradient climbing, see Figure 2.3). We find regions that present sim-



**Figure 2.4: Solution of the situated HKB system.** Eigenvalues ( $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$ ) of the attractor (left) and repeller (right) fixed points of the situated-HKB: force of the attraction/repulsion vs. variation of the control parameter  $s$ . Real part (solid), Imaginary part (dashed).

ple attractor/repulsion dynamics (when  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$  are real numbers) whereas other regions present a spiral attraction or repulsion (when  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$  have complex values).

Now, we show a detailed description of the relations between eigenvalues and behaviour, analysing the transitions of the eigenvalues for both the attractor and the repeller (see Figure 2.4) and focusing on the correspondences between those transitions and the respective transitions in the behavioural modes (Figure 2.5). Specifically, as depicted in the Figure we analyse the transitions from real to complex eigenvalues (from regular attraction to spiral attraction), and the behavioural transitions from underdamped to overdamped behaviour (the system finds equilibrium with or without oscillating) on the one hand and from spiral to cycloidal movement of the agent on the other.

*ATTRACTOR:*

- **TRANSITION AT  $s = 5.1$ :** At this point, the attractor experiences a change in its dynamics. A spiral attraction in the plane  $\lambda_1\lambda_2$  disappears and the attraction of the system no longer has a spiral shape. At the same point, the approaching strategy of the agent experiences a change. For  $s < 5.1$  the agent enters the final stable circular trajectory from

within, whereas for  $s > 5.1$  the agent enters from outside the circle. These approaching strategies correspond respectively to an under-damped and an over-damped behaviour of the situated HKB system.

- **TRANSITION AT  $s = 10.4$ :** The attractor changes again into a spiral shape, now in the plane  $\lambda_2\lambda_3$ . The behavioural change here is more subtle. It appears at the initial turning behaviour of the agent until it finds a stable trajectory and enters into the spiral trajectory of the robot. When  $s < 10.4$ , the robot enters into the trajectory by softly adjusting the value of  $\alpha$  with an over-damped behaviour, while when  $s > 10.4$  the value of  $\alpha$  oscillates around the trajectory, thus adjusting to the optimal value with an under-damped behaviour. This damping behaviour also affects  $\varphi$  and  $\eta$ . Oscillations are too small to be clearly appreciated in the trajectory of the robot. That is why in Figure 2.5, in the enlarged boxes, we just represent the orientation of the agent  $\alpha$ , which shows how the robot adjusts its behaviour to the final trajectory.

#### *REPELLER:*

- **TRANSITION AT  $s = 2.4$ :** At this point, a different kind of transition takes place. While transitions in the attractor were gradual, the change in the repeller at this value is an abrupt bifurcation. The system suddenly changes from a spiral in the plane  $\lambda_1\lambda_2$  to a spiral in the plane  $\lambda_2\lambda_3$ . Also, a redistribution of the values of the real part of  $\lambda_1$  and  $\lambda_3$  (i.e. the ‘strength’ of the repulsion) takes place in the transition. Consequently, in this case the change in the agent’s behaviour is more dramatic. As we saw in Figure 2.3, the agent changes from a cycloidal trajectory to a more stable strategy where the agent continuously approaches to the gradient source (either directly or following more pronounced spirals as  $s$  increases).

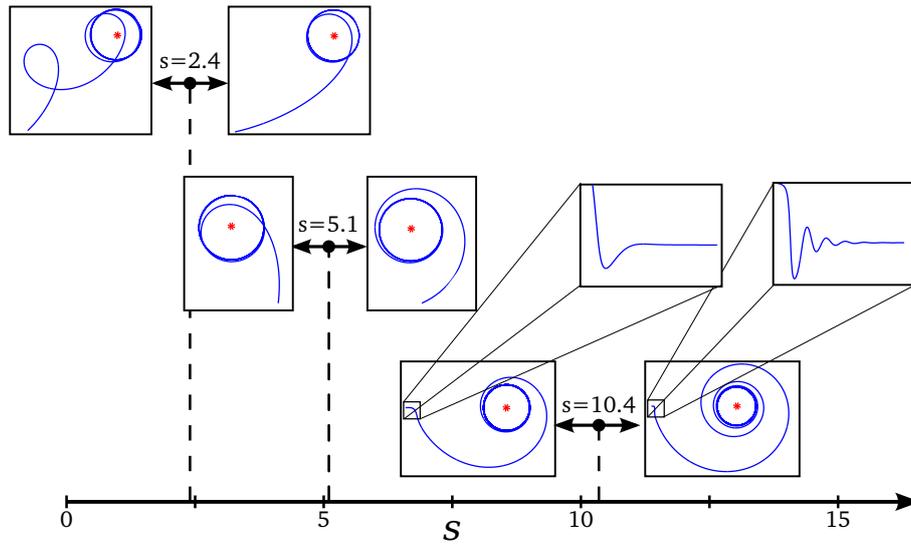
of its state space that bear qualitative significance. In Figure 2.4, it is illustrated

## 2.5 COMPARISON BETWEEN SITUATED, PASSIVELY-COUPLED AND DECOUPLED HKB SYSTEMS

In the previous section we provided a full dynamical analysis and explained how, in the situated HKB, the coupled brain-body-environment system gives rise to a gradient climbing behaviour. We also analysed the transitions that took place as we increase the sensitivity parameter.

Now, we want to explore the effect of the sensorimotor situatedness of the system (i.e. the role of the closed sensorimotor loop) upon the dynamics of the HKB equation by comparing the situated HKB with two homologous systems (see Figure 2.6):

1. A decoupled HKB system, such as the one originally used by Kelso, in which the effect of situated interaction with an external environment is not taken into account.



**Figure 2.5: Transitions in the phase space and changes in the agent's behaviour.** We can observe how changes in behaviour coincide with transition points in the dynamical description of the system: ( $s = 2.4$ ) an abrupt transition of the repeller's plane of spiral repulsion explains the switch from cycloidal to stable movement in the agent, ( $s = 5.1$ ) as the effect of spiral attraction vanishes, the agent changes from entering from inside to the final circular trajectory to entering from outside, and ( $s = 10.4$ ) when the spiral attraction appears in a different plane, the agent displays an oscillatory movement for adjusting the followed trajectory (the effect of the oscillations is shown over the value of  $\alpha$  in the enlarged box, its amplitude is damped too fast to be appreciated in the plot of the agent's trajectory)

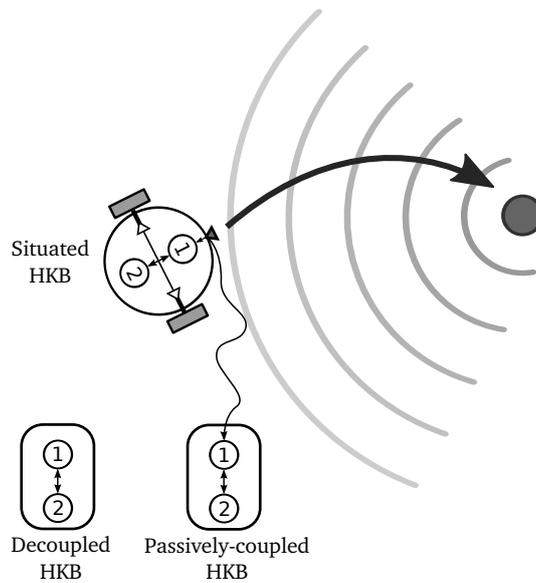
2. A passively-coupled HKB system, where the HKB equation receives a structured input resulting from a real interaction between a situated HKB system and its environment, but where this input does not directly correspond to the activity of the system but is received from a recording or virtual input of a truly behaving agent.

The second condition is equivalent to Held and Hein's classic experiment (Held and Hein, 1963) where two kittens were reared by holding one immobile and attached to the other, so that both received the same sensory stimulation, yet only one had freedom to control movement. In the following subsections, we analyse the dynamics of these HKB models, under equivalent parametric conditions, searching for the qualitative difference that highlight the functional and neurodynamic significance of the closed sensorimotor loop.

### 2.5.1 CASE I: NO SENSORY INPUT. THE DECOUPLED HKB SYSTEM

The decoupled HKB system simply consists of the classical HKB equation (2.1) whose dynamics have been widely analysed. Given the classical HKB equation with the parameters used above for the situated HKB ( $\Delta\omega_0 = 1$ ,  $a = 5$  and  $b = 1$ ) and after removing the sensory input, we get:

$$\dot{\varphi} = \Delta\omega_0 - a \cdot \sin(\varphi) - 2b \cdot \sin(2\varphi) \quad (2.15)$$

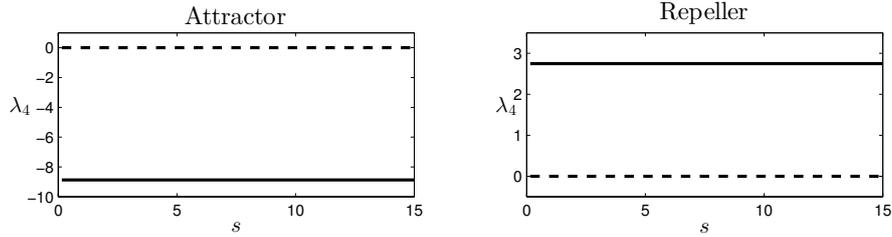


**Figure 2.6: Representation of a situated, decoupled and passively-coupled HKB systems.** The situated HKB system receives an input from its interaction with a gradient source. The passively-coupled HKB receives the copy of the input generated by a situated HKB. The decoupled HKB receives no input at all.

It is easy to see that, for this equation, two fixed points (or equilibrium points) are obtained by finding which values of  $\varphi$  make  $\dot{\varphi} = 0$ . The fixed or equilibrium points are found at  $\varphi = 0.11$  and  $\varphi = 2.53$ . Computing the Jacobian matrix of the equation for these values of  $\varphi$ ,  $J(0.11) = -8.87$  (attractor) and  $J(2.53) = 2.75$  (repeller), provides us with the values for the eigenvalue of the decoupled HKB at each point (denoted as  $\lambda_4$  and represented in Figure 2.7).

As we see in Figure 2.7, the eigenvalue is a real number at each fixed point, generating a simple pattern of attraction/repulsion in the dynamics of the system. Thus, the decoupled HKB alone cannot explain the behavioural changes shown in Figure 2.4, where simple patterns of attraction/repulsion are transformed into spiral cycles, or abrupt changes arise changing the plane of the resulting patterns. The decoupled HKB system displays simple attraction and repulsion forces around every fixed point and, therefore, the dynamics of the system can be described by constant attraction and repulsion forces regardless of the value of the parameter  $s$ .

In the situated HKB however, we observed how the system displays ‘qualitatively different’ behaviours, that is, behaviours that are not just due to gradual variations of a single dynamical regime but are a consequence of a phase transition in the system dynamics. This is a phenomenon that can also appear for the original HKB equation (under certain conditions the system is able to switch from one attractor to another; see Kelso, 1995). However, for the situated-HKB with the parametric configuration used here ( $a = 5$  and  $b = 1$ )  $\varphi$  remains within the monostable region (except for brief instants of time when the system visits the



**Figure 2.7: Solution of the decoupled HKB system.** Eigenvalues of the decoupled HKB ( $\lambda_4$ ): Force of the attraction/repulsion vs. variation of the control parameter  $s$ . Real part (solid), Imaginary part (dashed).

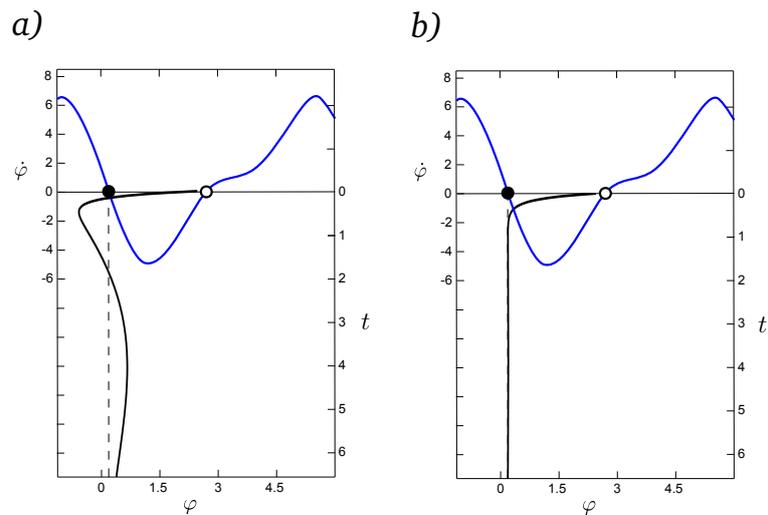
metastable region before stabilizing in the monostable region). Thus, it cannot display the phase transitions observed in the bistable configuration of the HKB. Taking the decoupled HKB as a reference, the situated HKB should not present, in principle, qualitative changes that are not due to external factors.

Thus, the observed phase transition in the situated HKB system cannot be explained by the dynamics of the HKB model alone. Instead, the reason for this transition lies in the joint dynamics of the agent-environment system, as we illustrated when we solved the eigenvalues of the system. However, it is true that, in a certain sense, the difference of dimensionality of the two models is enough to substantially modify the dynamics of the system, independently of the fact that these extra dimensions correspond to the agent or the environment. The very fact that the situated HKB has three dimensions instead of one makes both systems somewhat incommensurable.

This incommensurability further supports the situatedness thesis. It is clearly proven by the fact that the eigenvalues that determine the qualitative evolution of the system cannot be translated or mapped from the situated to the decoupled conditions: the whole brain-body-environment system defines a new coordinate system where the ‘brain’ contribution cannot be isolated.

The decoupled HKB is affected by a constant force of attraction/repulsion (see Figure 2.8.b) while the situated HKB is subject to forces in three different dimensions that continuously modulate each other (Figure 2.8.a). Note that even if we were feeding a constant input into the system (anywhere in the input range displayed by the situated system) the result would be equivalent.

The next logical step is to question whether the specific structure of the input is the crucial factor when comparing the decoupled and situated HKB systems. In order to address this question we introduce the passively-coupled HKB model where the HKB equation receives the exact same input as the freely behaving situated HKB.



**Figure 2.8: Attractor's behaviour.** Comparison of the evolution of the system around the attractor for (a) the situated HKB with  $s = 2.5$  and (b) the decoupled HKB (with  $\Delta\omega = 1$ , both with  $a = 5$  and  $b = 1$ ). The black line in the vertical axis represents the evolution over time (right vertical axis) of  $\varphi$ , which has been simulated during 6.5 seconds with an Euler step of 0.1 with arbitrary initial values of  $\varphi = 0.65$ ,  $\eta = -2.78$ , and  $\alpha = -2.07$ . The blue line represents the phase space of the HKB, representing the attractors as filled dots and the repellers as empty dots. We observe how the decoupled HKB is only affected by a simple attraction force with constant strength while a much richer dynamics is shown in the situated HKB, where different forces of attraction interact to modulate the systems evolution

### 2.5.2 CASE 2: EXTERNALLY STRUCTURED SENSORY INPUT. THE PASSIVELY-COUPLED HKB SYSTEM

We can model the passively-coupled HKB system by just adding a new variable  $\varphi^*$  that receives the same input as  $\varphi$  (i.e., a new equation  $\dot{\varphi}^* = 1 + \dot{\eta} \cdot s - \alpha \cdot \sin(\varphi^*) - 2b \cdot \sin(2\varphi^*)$  is added to the previous system (equation 2.14), see Figure 2.6). As a result, we have to solve a four-dimensional system under the same conditions as before (parameters  $\alpha = 5$ ,  $b = 1$ , and  $c = 5$ , and the value of  $s$  defined in [0, 15]).

Analogously to the previous system we get two fixed points. There is an attractor at position  $(0.11, 2.28, -\pi/2, 0.11)$  and a repeller at  $(2.53, 0.43, \pi/2, 2.53)$  (as shown in Figure 2.7), and through the diagonalization of the respective Jacobian matrices, four eigenvalues  $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$  are obtained.

As we will see below, although the term  $\lambda_4$  is decoupled from the activity of the situated HKB (and therefore independent of the type of coupling, i.e. independent of the value of  $s$ ), the behaviour of the new variable  $\varphi^*$  is necessarily described by a combination of the eigenvalues of the situated HKB system  $(\lambda_1, \lambda_2, \lambda_3)$  and the decoupled HKB equation  $(\lambda_4)$ . This provides us with a qualitative difference between the activity in  $\varphi$  and  $\varphi^*$ .

In order to show this, we need to remind ourselves how the general solution of a dynamical system is built. It is well-known that the eigenvalues of a *linearized* dynamical system define the behaviour of the system around the fixed points as a series of exponential functions which converge to an attractor or diverge from a repeller (Strogatz, 2001). For both the attractor and the repeller, the solution of any variable of the passively-coupled HKB system has the form:

$$x(t) = x_0 + A_{\lambda_1} \cdot v_{1x} e^{-\lambda_1 t} + A_{\lambda_2} \cdot v_{2x} e^{-\lambda_2 t} + A_{\lambda_3} \cdot v_{3x} e^{-\lambda_3 t} + A_{\lambda_4} \cdot v_{4x} e^{-\lambda_4 t} \quad (2.16)$$

where  $x$  can stand for  $(\varphi, \eta, \alpha$  or  $\varphi^*)$ ,  $\lambda_i$  and  $v_{ix}$  in general have complex values (where  $v_{ix}$  are the eigenvectors of the solution and  $\lambda_i$  are the eigenvalues,  $i = 1, 2, 3, 4$ ),  $x_0$  represents the position of the fixed point and  $A_{\lambda_i}$  are the coefficients that fix the initial state of the system (with  $i = 1, 2, 3, 4$ ).

After making some simple calculations, we find that: (1) some eigenvectors are null (i.e.  $v_{4\varphi}, v_{4\eta}, v_{4\alpha} = 0$ ) so  $\lambda_4$  are only be part of the solution of  $\varphi^*$ ; (2) moreover, others share the same values ( $v_{1\varphi} = v_{1\varphi^*}, v_{2\varphi} = v_{2\varphi^*}$  and  $v_{3\varphi} = v_{3\varphi^*}$ ), so  $\lambda_1, \lambda_2$  and  $\lambda_3$  are equally present in the dynamics of  $\varphi$  and  $\varphi^*$ , (3) finally, we can simplify the system a little more because ( $v_{4\varphi^*} = 1$ ) and ( $\varphi_0 = \varphi_0^*$ ) for any value of  $s$ .

Given the previous simplifications, the solutions of the system around  $x_0$  become:

$$\begin{aligned} \varphi(t) &= \varphi_0 + A_{\lambda_1} \cdot v_{1\varphi} e^{-\lambda_1 t} + A_{\lambda_2} \cdot v_{2\varphi} e^{-\lambda_2 t} + A_{\lambda_3} \cdot v_{3\varphi} e^{-\lambda_3 t} \\ \eta(t) &= \eta_0 + A_{\lambda_1} \cdot v_{1\eta} e^{-\lambda_1 t} + A_{\lambda_2} \cdot v_{2\eta} e^{-\lambda_2 t} + A_{\lambda_3} \cdot v_{3\eta} e^{-\lambda_3 t} \\ \alpha(t) &= \alpha_0 + A_{\lambda_1} \cdot v_{1\alpha} e^{-\lambda_1 t} + A_{\lambda_2} \cdot v_{2\alpha} e^{-\lambda_2 t} + A_{\lambda_3} \cdot v_{3\alpha} e^{-\lambda_3 t} \\ \varphi^*(t) &= \varphi_0 + A_{\lambda_1} \cdot v_{1\varphi} e^{-\lambda_1 t} + A_{\lambda_2} \cdot v_{2\varphi} e^{-\lambda_2 t} + A_{\lambda_3} \cdot v_{3\varphi} e^{-\lambda_3 t} + A_{\lambda_4} \cdot e^{-\lambda_4 t} \end{aligned} \quad (2.17)$$

Here, we can see that the dynamics of  $\varphi^*$  corresponds to the dynamics of  $\varphi$  plus an extra

term which determines the difference between the situated and the decoupled model:

$$\varphi^*(t) = \varphi(t) + A_{\lambda_4} \cdot e^{-\lambda_4 t} \quad (2.18)$$

From now on, this term will be denoted as  $\Delta\varphi^*(t) = A_{\lambda_4} \cdot e^{-\lambda_4 t}$ , representing the difference between the situated and passively-coupled HKB. We now quantify the influence of this extra term. The solution in  $t = 0$  around the fixed points is:

$$A_{\lambda_4} = \varphi^*(0) - \varphi(0) \quad (2.19)$$

That is, the influence of the decoupled term around the fixed points depends on the linear difference between the initial conditions of  $\varphi^*$  and  $\varphi$ .

Therefore, we can interpret the dynamics of the partially-coupled HKB,  $\varphi^*(t)$ , as composed of two ‘partially-decoupled’ terms  $\varphi(t)$  and  $\Delta\varphi^*(t)$  (partially decoupled because  $\varphi(t)$  influences  $\Delta\varphi^*(t)$  but not the other way around). What is the difference between these two terms?

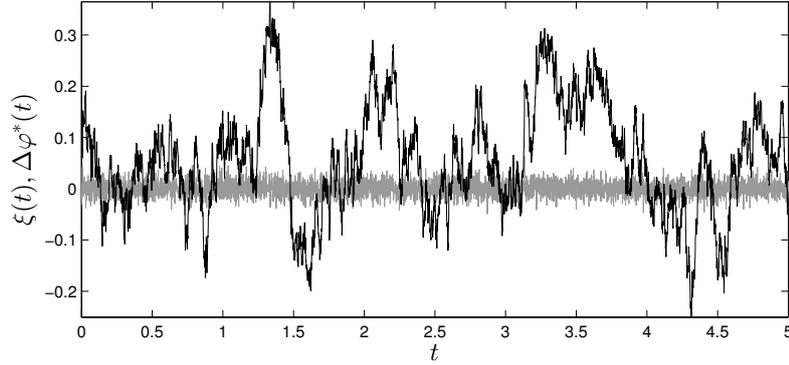
On the one hand,  $\varphi(t)$  follows a complex dynamic unfolding, intertwined with and modulated by the dynamics of  $\eta(t)$  and  $\alpha(t)$  as a combination of the eigenvalues  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$  (Figure 2.8.a). On the other hand, the dynamics of  $\Delta\varphi^*(t)$  are much simpler as they are defined by a unique eigenvalue  $\lambda_4$  (Figure 2.8.b). However, we are analysing a highly idealized situation, where the system easily converges into its attractor without having to deal with any kind of perturbation.

In the subsection below we quantify the contribution of the  $\Delta\varphi^*(t)$  term to the dynamics of the system in a more realistic situation. We analytically derive a theoretical expression to calculate  $\Delta\varphi^*(t)$  in the presence of persistent perturbations and we validate the results with numerical experimentation.

#### COMPARING SITUATED AND PARTIALLY-COUPLED HKB SYSTEMS UNDER PERTURBATIONS

Typically, in a real system, variables are not only subject to differences in initial conditions as the term  $\varphi^*(0) - \varphi(0)$  represents, but also suffer continuous fluctuations under different forms. Each fluctuation in the difference between  $\varphi(t)$  and  $\varphi^*(t)$  (that is, the difference between the fluctuations of  $\varphi(t)$  and the fluctuations of  $\varphi^*(t)$ ) will provoke an effect as a function with the form  $A_{\lambda_4} \cdot e^{-\lambda_4 t}$ , with  $A_{\lambda_4}$  being equal to the amplitude in the fluctuation at time  $t$ , and  $\lambda_4$  the eigenvalue of the decoupled HKB in the attractor (thus assuming that the value of  $\Delta\varphi^*(t)$  is small). If fluctuations are present at different instants, the result is a linear combination of all the exponential functions multiplied by the respective values of  $A_{\lambda_4}$  for each instant of time. If the fluctuations between  $\varphi(t)$  and  $\varphi^*(t)$  (i.e. the difference of the fluctuations affecting the two variables) are given by the function  $\xi(t)$ , we can compute the final expression of the passively-coupled HKB around the fixed points as a Green’s function (Alimov, 2001):

$$\varphi^*(t) = \varphi(t) + \Delta\varphi^*(t) = \varphi(t) + \int_0^t \xi(\tau) \cdot e^{-\lambda_4(\tau-t)} d\tau \quad (2.20)$$



**Figure 2.9: Effect of fluctuations in the passively coupled system.**  $\xi(t)$  (grey line): difference of the introduced fluctuations in  $\varphi(t)$  and  $\varphi^*(t)$ , and  $\Delta\varphi^*(t)$  (black line): fluctuations in the difference between  $\varphi(t)$  and  $\varphi^*(t)$  computed through equation (2.20).

Without internal fluctuation the value of  $\varphi^*(t)$  would converge to the value of  $\varphi(t)$  after an initial phase of adjustment. But if fluctuations are present, we can use the expression above to compare the fluctuations in the situated model with the fluctuations in the passively-coupled model.

As an example, we simulate the situated HKB system with a passively-coupled HKB connected to it with an Euler step of 1ms during a period of 5s. We introduce an additive white noise to the variables  $\varphi$  and  $\varphi^*$  with a variance of  $10^{-4}$ . Then,  $\xi(t)$  is equal to the difference between these two sources of noise, which conserve their white noise structure with twice its variance ( $2 \cdot 10^{-4}$ ) (Figure 2.9, grey line). By applying equation (2.20) we can compute the resulting fluctuation  $\Delta\varphi^*(t)$  that determines the differences between the values of  $\varphi(t)$  and  $\varphi^*(t)$  (Figure 2.9, black line). We can observe how the fluctuations in  $\Delta\varphi^*(t)$  lose the uncorrelated white noise structure of the initial fluctuation, and get a radically different structure with different temporal correlations induced by the  $e^{-\lambda_4 t}$  term. We can validate this result by comparing  $\Delta\varphi^*(t)$  computed with equation (2.20) with the difference between  $\varphi(t)$  and  $\varphi^*(t)$  measured experimentally without the effects that reduce the system to a linear system around the attractor. That is, we can measure the error in the estimation of the fluctuation as follows:

$$e(t) = (\varphi(t) - \varphi^*(t)) - \Delta\varphi^*(t) \quad (2.21)$$

By computing  $e(t)$  we find that the amplitude of the error is significantly smaller than the amplitude of the theoretical measures of fluctuation  $\Delta\varphi^*(t)$ . Calculating how well the theoretical results adjust to experimental data, we obtain that  $(\varphi(t) - \varphi^*(t))$  fits  $\Delta\varphi^*(t)$  with an  $R^2$  coefficient of 0.95, which indicates a good fit.

If we compute the variance of  $\Delta\varphi^*(t)$  we find that it is equal to  $12.3 \cdot 10^{-4}$ , that is, about 6 times larger than the fluctuation introduced to the system. Thus, when measuring a passively-coupled version of a phenomenon we are interested in, we would expect a bigger amount of

fluctuation than in the real situated case. Furthermore, in the passively-coupled model fast fluctuations are modulated by slow fluctuations, since we can see that low frequencies dominate the signal. This may represent an important advantage in different tasks; whereas the original white noise introduced to the system may be averaged out and removed, the fluctuations added to the passively-coupled variable  $\varphi^*(t)$  present correlations at different scales that cannot be filtered easily.

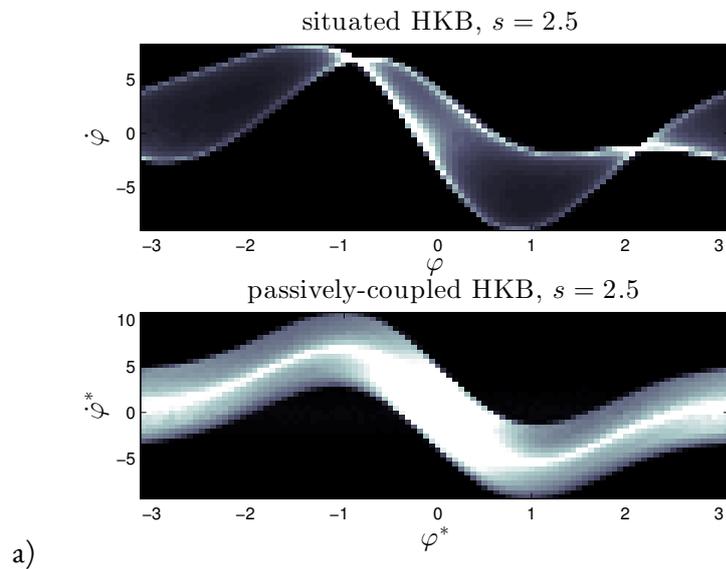
## DYNAMICAL SIGNATURE

Above we have presented proof of the dynamical differences between a situated HKB system and a passively-coupled HKB system in the transients around the attractor and repeller of the system. Also, we have compared the results with experimental measures during the whole trajectory and without the effects of the linearization of the dynamical system, and demonstrated similar results. We have concluded that these changes are the product of the different modes of interaction of the situated and passively-coupled agent, which modulate the dynamical landscape of the brain-body-environment system. What do we mean when we say that the different types of coupling transform the dynamical landscape of the system? To clarify this, we analyse the *dynamical signature* of the HKB equation for the situated HKB and the passively-coupled HKB.

With the idea of a dynamical signature, we intend to obtain a simplified representation of the phase space of the system which allows us to directly observe some of the characteristics of the system. In our case, we extract the dynamic signature of the system by collapsing the phase space of the agent to a statistical description of the behaviour of one variable and its representative. We can interpret this *dynamic signature* as the functional ‘brain correlate’ of gradient climbing behaviour. No single brain ‘state’ (i.e. value of  $\varphi$ ) is functionally significant in terms of behaviour. What matters is the shape of the temporal pattern of phase relation among oscillators. In this sense, the ‘gradient-climbing behaviour’, as a unit of explanation, is not the result of a set of brain states encoding a decision or a motor-program output, but results from a specific coordination pattern between sensor and motor surfaces, mediated by a specific coordination pattern between ‘brain oscillations’. The specific pattern of internal coordination that corresponds to gradient climbing behaviour is here called ‘dynamic signature’ and refers to the temporal structure of internal changes that is both cause and effect of different instances of a particular behaviour.

To obtain and compare the dynamic signature of our agents, we have simulated the situated HKB system (with the sensitivity parameter  $s = 2.5$ ) and a passively-coupled HKB connected to it, with an Euler step of 1ms and a duration 10000s, and we periodically reset the variables of the system ( $\varphi$ ,  $\eta$  and  $\alpha$ ) to new randomized values with intervals of 20ms. The goal of this randomization is to sample a wide range of initial conditions of the system, that is, to capture a sufficient range of different situations to constitute the abstract category of ‘gradient climbing behaviour’. This way we can identify what a ‘neural signature’ or dynamic pattern trace corresponds to all the instances of this form of behaviour. We formally

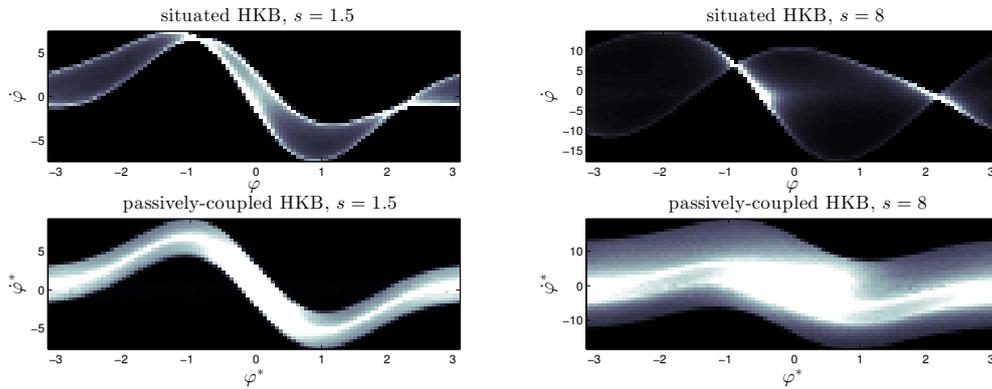
define the dynamic signature of the HKB system as the density distribution of the derivative of the relative phase  $\dot{\varphi}$  with respect to  $\varphi$  (or in terms of  $\varphi^*$  in the passively-coupled case).



**Figure 2.10: Dynamical Signature.** Signature of the situated HKB with  $s = 2.5$  and the corresponding passively-coupled HKB. It represents the density distribution of the effective phase space of the HKB equation when it is coupled with an environment, showing the difference between situated and passive coupling.

What we see (Figure 2.10) is that the dynamical signature of the system changes severely when the system is situated in an environment. Whereas the passively-coupled HKB displays the shape of the original HKB phase space with a ‘blurring’ effect created by the addition of a structured input (we can see it in Figure 2.10 as a ‘thick’ line shaped with the form of the HKB original phase space), in the situated system the structure of the dynamical signature no longer resembles the original HKB phase space. The situated system has modulated or re-shaped its state space into a specific pattern through sensorimotor coordination.

If we look for the signature for different configurations of the Situated HKB model, we observe that each value of  $s$  produces a particular signature pattern (Figure 2.11), representing the different dynamical configurations of the ‘brain correlate’ of a particular sensorimotor pattern of the agent. As we see, the dynamic shape of the neural controller changes when it is engaged in different sensorimotor loops (even although the HKB equation has exactly the same attractor in all cases). This shows that, even in relatively simple models like this, analysing correlations of brain activity as if they were in isolation from their surrounding world, ignores the richer dynamics that one would expect from agents strongly coupled to their environments.



**Figure 2.11: Dynamical Signature.** Signature of the situated HKB with  $s = 1.5$  (top left) and  $s = 8$  (top right), and the corresponding passively-coupled HKB (bottom left and right). It represents the density distribution of the effective phase space of the HKB equation when it is coupled with an environment, showing the difference between situated and passive coupling.

## 2.6 DISCUSSION

We have just shown how the HKB system displays qualitatively different dynamics under situated conditions, as compared to decoupled and passively-coupled conditions. We can restate the main results as follows:

1. Transitions between qualitatively different types of sensorimotor behaviour, that are generated by the situated HKB with increasing sensitivity, cannot be deduced from the behaviour of the decoupled HKB nor from the analysis of the passively coupled HKB alone. The nature of these transitions can only be revealed through the analysis of the whole brain-body-environment system.
2. Even for a single fixed value of the sensitivity parameter, the transient trajectory of the situated HKB system towards its attractor is far from trivial and unfolds at different temporal scales. The transient trajectory of the decoupled HKB system, instead, is relatively simple and monotonic.
3. A passively-coupled agent receiving an input generated by a situated agent shows inter-correlated and amplified fluctuations that are not present in the situated agent.
4. Finally, the dynamical signature of the agents shows how the type of coupling (passive or situated) severely transforms the phase space of the HKB system. The specificity of functional neural signatures is lost when studying the brain out of the closed sensorimotor loop, even if it is subjected to exactly the same input.

The differences between situated and decoupled oscillatory controllers illustrate how much we would miss if we analysed the ‘brain’ of an agent isolated from its embodied situatedness

in an environment. On the one hand, the brain-body-environment system constitutes a dynamical holistic continuum where phase transitions can take place, without necessarily corresponding to phase transitions that would occur in an isolated brain. On the contrary, this joint dynamical structure would generally not be deduced from the isolated controller of an agent. On the other hand, the unfolding of behaviour shown by a system is modulated by the continuous interplay with the environment at different time-scales that generally are not present in the dynamics of the controller system alone, giving rise to much richer behavioural dynamics.

Our situated HKB is a case of double coordination: sensorimotor coordination of the coordinated dynamics of the two oscillatory components (modelled by the HKB equation). What is crucial is the fact that *under situated conditions the dynamics of the HKB can be modulated by the precise and interactively structured coordination between its internal dynamics and the sensorimotor environment*. The mode in which sensory input (the control parameter) changes as a function of the motor output (which is in turn generated by internal dynamics) through the environment, enables this higher order coordination. It can be said that the agent modulates its internal dynamics through sensorimotor coordination, in a manner that is not available to the decoupled or passively-coupled system, resulting in functionally specific internal patterns that we have demonstrated with the dynamic signature.

It is important to note that we have focused on the most simple configuration of the HKB model; with an extremely simplified body and environment. Regarding the internal configuration of the HKB, we have studied it under the simplest parametric configuration that produces a single attractor instead of two or none, which is due to the  $a/b$  coefficient value that we kept fixed. Even for this simple configuration we have found qualitative differences between the transient dynamics of the system before falling into the attractor — see  $\varphi = 0.11$  and transient dynamics around  $(-1, 2.50)$  in Figure 2.8. The behaviour of the system shows strong differences under the situated and decoupled conditions. But the HKB can display much richer behaviour on its bistable configuration (with two attractors and repellers) or along the metastable region (see light grey area in Figure 1.6) where there are no fixed points but just attractor and repeller shadows, a phenomenon that appears for high values of difference between the natural frequencies  $\Delta\omega$  (Kelso, 1995, this is something that did not occur on our model given the low value of  $\Delta\omega_0$ ).

The differences we found between decoupled, passively-coupled and situated dynamics can be expected to be amplified for richer parametric configurations of the HKB model. For example, the dynamics of the same situated HKB model have been analysed performing a gradient climbing task under metastable regimes of  $\varphi$  (Santos et al., 2012). In this case the effect of sensorimotor modulation of the HKB dynamics was much stronger, shaping specific metastable regimes and transitions between them, whereas passively-coupled agents showed different regimes and transitions. We want to stress that the appearance of qualitative differences between situated and decoupled systems is not a contingent result of the parameter values chosen here (nor those of Santos et al., 2012). We observed the same changes for dif-

ferent parameter values and also with different oscillatory controllers (such as homeostatic Kuramoto oscillators, as we see in the next chapter).

As far as the experimental setup is concerned, we demonstrated that a simple linear and radially symmetric 2D gradient environment, a single sensor and two motors were sufficient for the HKB to exploit the sensorimotor coupling to modulate its internal dynamics in qualitatively different manners under different coupling conditions. Richer environments, more complex tasks and multimodal and higher dimensional sensory and motor surfaces could increase the divergences we showed here. It was also shown that in cross-modal perception, perturbations to one sensory modality can be compensated by other sensory modalities (Ernst and Banks, 2002), it is therefore to be expected that multi-modal sensorimotor engagement could have an even greater effect on brain dynamics than a single sensorimotor modality.

Could these results be generalized to other neurocognitive agents? Not directly. We have just provided a proof of concept of how severely oscillatory brain dynamics can be altered by sensorimotor coordination. However, even if true for our extremely simplified model this conclusion has still to be proven for neuro-biological systems. We are not aware of any neuroscientific study comparing situated and passively coupled recordings for perceptuomotor tasks, but new recording techniques (Linderman et al., 2006; Santhanam et al., 2007; Fan et al., 2011) might help replicate the experiments we have developed in this chapter. Although it is much more complicated to measure neural activity in a moving animal due to noise from motor activity, an interesting technique has been developed recently to overcome some of these limitations. The technique consists of a spherical treadmill where a head-restrained mouse interacts with a virtual-reality system. The mouse is surrounded by a toroidal screen that covers its field of vision, and images are updated on the basis of the movements of the animal and they are measured as rotations of the spherical treadmill using an optical computer mouse (Dombeck et al., 2007; Harvey et al., 2009; Dombeck et al., 2010; Harvey et al., 2012). Some of the results using this technique highlighted the importance of analysing situated activity to understand neural mechanisms. In (Harvey et al., 2012) it was found that activity in the prefrontal cortex in mice in a navigation decision task showed much more heterogeneity than in previous recordings and models of prefrontal activity in mice in other experimental settings, in which the activity is much more stereotyped and low-dimensional. A key aspect of such result is that the behavioural task in the experiment did not isolate the decision-making process, so the trajectories of neural activity may have been modulated by sensory information (Harvey et al., 2012, p.67).

What would the results of applying passively coupled conditions to humans or other animals be? The cognitive or psychological effects of different degrees of disruption of the sensorimotor loop could range from a complete lack of perceptual capacity (e.g. when inducing sensory streams resulting from input recorded from saccadic exploration) to a loose sense of reality when sensorimotor coupling conditions in virtual reality are not optimal. In humans, new experimental paradigms in substitutional reality (playing back recorded visual experience to re-create realistic scenes, see Suzuki et al., 2012) showed that ‘a major factor influencing suc-

successful substitution in the SR [Substitutional Reality] system is consistent visuo-motor coupling throughout the experience' (Suzuki et al., 2012, p.6).

At a more abstract level, the present model makes a theoretical contribution to the ongoing debate around the *causal* vs. *constitutive* role of action in perception. If we consider a perceptual state  $X$  as a collection of past and present processes and relations, we may consider different situations according to  $X$ 's context and the causal role of this context. We may summarize the different explanatory possibilities for the elements influencing  $X$  (De Jaegher et al., 2010, p.443) as (1)  $F$  can be considered a *contextual factor* if variations in  $F$  produce variations in  $X$ , (2)  $C$  is an *enabling condition* if the absence of  $C$  prevents  $X$  from occurring, and (3)  $P$  is a *constitutive element* if  $P$  is part of the processes that produce  $X$ . Roughly speaking, causal theories (Prinz, 2006; Adams and Aizawa, 2008) claim that movement can certainly be a cause of the right sensory input that in turn causes perceptual states but it is not strictly necessary. Constitutive theories on the other hand (O'Regan and Noë, 2001; Noë, 2004) claim that movement itself is part of the perceptual process. Causal theories are generally internalist by asserting that what matters is the brain state (caused by the sensory input), whereas constitution theories tend to align with externalism (perception is a process that involves a distributed set of brain, body and environmental components, all of them constituting the same percept). Our model can be used to show how, even if favouring internalism, the neural signature that corresponds to a given cognitive episode can be qualitatively different from the neural signature obtained when the very same input (cause) is induced into a passively-coupled system and fine grained sensorimotor contingencies become strictly necessary or constitutive of functionally distinct neural signatures. Whether this holds also true for natural systems is a matter for further experimentation but the conceptual discussion, which is often obscured by a lack of clear models, can benefit from the findings presented here.

### 2.6.1 CONCLUSION

Contemporary neuroscience often assumes that it is possible to deduce the behavioural properties of the brain by simply studying its dynamics under 'passive' input conditions (e.g. neural recordings in anaesthetized animals) or building models that ignore sensorimotor dynamics (like large scale networks with noise-inputs or otherwise non behaviourally controllable input). The brain-body-environment coupled dynamics is rarely considered as a unified dynamical system and there is still a limited understanding of the interplay between sensorimotor and neural dynamics.

In this chapter we have illustrated what neuroscience misses when ignoring the role of sensorimotor coordination, particularly when drawing models of brain dynamics out of neural recordings in the absence of closed sensorimotor loops. We have presented a minimal model that shows the qualitative differences that can arise under situated and decoupled sensorimotor conditions. We have shown how some features – such as (1) the phase transition that takes place modifying a sensitivity parameter, (2) the attraction patterns, (3) the neurodynamic signatures and (4) the modulatory capacity of the situated system – need to be explained by a

framework that takes into account the coupled dynamics of the brain-body-environment system.

The model presented in this chapter allows us to understand the importance of a dynamic sensorimotor landscape for the emergence of different modes of engagement with the environment. By varying a parameter  $s$ , we can observe how different modes of coordinated sensorimotor activity emerge (Figure 2.3). Each one of these modes is characterized by a dynamical signature reflecting a particular pattern of sensorimotor interaction (Figures 2.11 and 2.11). With this model we have shown that sensorimotor dynamics can be constitutive of the specific patterns of oscillation of a controller. This challenges some notions of autonomy: those that emphasize closure and openness as reactions to simple perturbations from the environment. We have also shown how strong sensorimotor situatedness might involve a fine grained sensorimotor coordination that is necessary to produce the functionally relevant brain dynamics.

In order to illustrate the simplest case of strong sensorimotor situatedness we have made use of the HKB model, whose 'internal' dynamics make it relatively simple to address some key aspects of autonomy (such as homeostasis or maintenance of organizational invariants). Somehow, the neural controller of the situated HKB is 'enslaved' into a specific brain-body-environment dynamical system, and we can only modulate the dynamical landscape of the agent by externally tuning its sensitivity parameter. However, an agent permanently trapped into the coupling of an environmental dynamical field cannot be considered as autonomous. How does autonomous agency arise from strongly coupled sensorimotor agents? What kind of modulation do they exert over the control parameters that regulate behaviour? This is what we will explore in the next chapter. Beyond the constitution of the agent in a sensorimotor domain, we will analyse another fundamental aspect of cognitive autonomy: how an agent may be able to modulate the flow of sensorimotor interactions, breaking the symmetry between agent and environment. We will present a more complex model involving plasticity able to capture the regulation of sensorimotor engagement.

# 3

## Situated Autonomy: organizational homeostasis in the sensorimotor loop

**ABSTRACT:** We explore a definition of neurocognitive autonomy through a simulated robotic agent. We find that the coordination of internal plastic mechanisms and sensorimotor adaptive regulation is able to create an agent-environment asymmetric circular loop which supports the constitution of the agent in front of its environment.

**CONTRIBUTIONS:**

- We propose a neurodynamic model based on relational homeostasis in an oscillator network.
- We test our model in a behavioural preference task, observing that our model presents some advantages in respect to previous models based in component-based homeostasis.
- Using information theory tools we are able to analyse the relations between neurodynamic plastic regulatory loops and sensorimotor coupling, depicting a circular relationship between levels of autonomous organization.
- Based on experimental results, we propose new perspectives to improve current definitions of neurocognitive autonomy.

**RELATED PUBLICATIONS:**

- Aguilera, M., Barandiaran, X. E., Bedia, M. G., and Seron, F. (2015). Self-Organized Criticality, Plasticity and Sensorimotor Coupling. Explorations with a Neurorobotic Model in a Behavioural Preference Task. *PLoS ONE*, 10(2):e0117465.

In previous chapters we reviewed different notions around the ideas of persistence, self-regulation, identity, and so forth, trying to characterize what kind of mechanisms may be suited for generating autonomous behaviour. We concluded that two crucial aspects for autonomous organization, first, the ability of the system to constitute its own identity, generating some form of cohesive unity through the recurrent interaction of its components, and secondly, the ability of this unity to regulate the interaction with its environment, creating a domain in

which the system emerges as a differentiated entity with respect to its environment. We have proposed that these aspects are captured by the notions of operational closure, interactive asymmetry and precariousness in the behaviour of an agent, which we considered as sufficient for characterizing it as autonomous.

As we described in Chapter 1, the organization of unicellular organisms is a clear example of autonomous organization: the recurrent metabolic processes within the cell build and sustain a membrane that encloses it in space as a unit and, at the same time, the membrane regulates material and energetic flows through the cell to constitute the cell as a differentiated entity in continuous and asymmetrical interaction with its environment. Nevertheless, for other levels like sensorimotor agency, the agent-environment interface is no longer a self-produced physical boundary which confines the autonomous self. Instead, the interface lies in an interactional domain of sensorimotor processes. This chapter focuses on how we can define such autonomous closure presenting a robotic model controlled by an adaptive neurodynamic controller.

### 3.1 DECISION-MAKING, HABITS AND AUTONOMY

How do autonomous agents constitute a dynamical coherent unit in front of a changing environment? What neural and bodily mechanisms are involved in this process and how can we model them and their interactions? In this chapter we explore these questions using a robotic agent in a behavioural preference task. Imagine a simplified world with two possible routes of action: action A and action B. An autonomous agent would start interacting with its world, learning ways to perform A or B as it encounters opportunities to do so. At some point, the agent may encounter *preferred* forms of performing one or other action, it will develop a habit. However, habits are not permanent, and will be influenced by new experiences encountered by the agent. Let's explore this idea with an everyday life example.

#### 3.1.1 A BIKE-SCIENTIST-ROUTE AUTONOMOUS COMPLEX

We can imagine a cognitive scientist who bikes to work every day. She has to choose between a more direct road to the university, which she has to share with regular traffic, or using some extra time to take a small detour through a more pleasant bike lane traversing a park. There are several factors influencing the decision, the amount of traffic on the road, whether or not she is in a hurry, or simply the fact that it is a nice day conducive to enjoying the ride. As an autonomous agent, every morning our scientist is faced with the decision of which route to take. In the previous chapter, we argued that cognitive autonomy is not a process that can be restricted to events happening inside the brain. Indeed, the decision of taking one route or another will be influenced by events happening in a sensorimotor space including her body and the environment. For example, the decision to avoid the road may be influenced by the perception of danger as the cars pass by in a heavy traffic day, or a sensation of well-being when feeling the sun and breeze on her face while cycling. Similarly, the decision of taking the road after a bad night's sleep may be influenced by a sensation of weariness when starting cycling,

making her feel that it is not worth travelling distance of the bike lane. Alternatively, on a blustery day, the feeling like pedalling with the brake blocks rubbing on the rim when facing headwind may influence her to cycle through the road where nearby buildings act as parapets for the wind.

Moreover, these everyday decisions do not only depend on the instant correlation of factors favouring one or other option. After several days taking the bike lane, the scientist will develop a habit displaying a preference for that behaviour. The more this route is taken, the more she may become familiar with it reinforcing the preference towards using the bike lane. Of course, this preference will not be permanent, and a series of windy days may make her change her preference and develop a new habit in which she takes the road every morning. Thus, the constitution of our biking scientist as an autonomous entity is not restricted to a single identifiable material domain. Her brain obviously has a fundamental role, but also the state of her body, the state of the traffic, together with an intricate history of previous decisions which shape the neurodynamic structures of the scientist making her more or less sensitive to different aspects of her body and the environment. Even the bike itself and how it feels when she rides it is a fundamental part in this autonomous decision-making. A lighter bike would make pedalling easier when facing headwind. A road bike may feel safer when riding into traffic since it is faster and allows the rider more control, whereas a touring bike might be more comfortable when enjoying the park's bike lane.

In our example, autonomy is not a property or a function that is created by any independent part of the system, but is a process that cuts across many scales of the system: the bike, the scientist, the sensorimotor perception of the conditions for biking in a precise instant, the constraints of the environment (traffic, weather, an appointment at the university), the history of previous decisions is solidified in particular behavioural habits and so forth. If one of these elements dominates the others, we cannot say that the decision is being taken autonomously. For example, if the traffic is so heavy every day that the road is no longer a viable option, we can no longer consider the decision autonomous, since there is only one real option to be chosen (under some limits ensuring the survival of the scientist). Returning to the analogy with autonomy at the level of the cell: what is the unity that is constructed here and what is the environment that interacts with it? Is the bike part of the environment or of the autonomous entity? What about the wind or the traffic? We can clearly identify the scientist as a unity because she is a biological entity based on metabolic and organismic autonomy. But what if we were to replace her by a robotic bike controlled by an artificial neural network which is not alive in a biological sense? What if the bike was to be remotely controlled from a software computer using meteorological and traffic data as inputs? Is there a systematic operational manner to distinguish an autonomous identity out of a complex network of interacting processes? In this chapter we present a model of neurocognitive autonomy that proposes answers for such questions, based on recent developments in neurodynamics, radical embodiment, and autonomous robotics.

### 3.1.2 TOWARDS A MODEL OF NEUROCOGNITIVE AUTONOMY

In this chapter we propose an operational model of sensorimotor autonomy for exploring the role of the modulation of sensorimotor adaptive behaviour by the emergence and reconfiguration of homeostatically regulated phase-locking dynamics in the brain. Typically, neurodynamic approaches have focused mostly in understanding the kind of neural organization that can cope with the requirements of conscious, autonomous activity. However, there is a crucial aspect of cognitive dynamics not always present: the sensorimotor coordination that continuously feeds back into brain dynamics (from saccadic eye movements to proprioception; from perception to action; O'Regan and Noë, 2001; Aguilera et al., 2013a). The emergence of different 'cognitive acts' should not be limited to neural events in the brain, but need to be established through a continuous interaction between brain, body and environment. Thus, synchronizing and de-synchronizing mechanisms should also be studied as embedded in organismic, sensorimotor, and intersubjective regulatory cycles, whose interaction is responsible for the emergence of global organism-environment processes that define the cognitive acts that neural ensembles account for. Conversely, embodied robotic models focusing on sensorimotor interaction through dynamical systems coupling (e.g. the model presented in the previous chapter), typically are locked-in a particular mode of sensorimotor engagement and do not actively regulate this coupling with the environment, therefore failing to represent a key aspect of autonomous behaviour: the ability to break agent-environment symmetry and actively define the conditions of the sensorimotor interaction.

In this chapter we present some notions about sensorimotor autonomy, and develop and analyse in detail a minimal model of a plastic neurodynamic controller as a behaviour generating mechanism to test the role of phase-locking/scattering mechanisms and sensorimotor regulation in the emergence and maintenance of behavioural and neurodynamic relational invariants. We update previous models from autonomous robotics proposing phase relationships between oscillatory nodes as a better suited regulatory variable to maintain organizational invariances in the system (e.g. the maintenance of specific neural ensembles), finding that the proposed model presents some advantages over previous ones. To explore the potential of the model we analyse a minimal agent controlled by three oscillatory nodes with synaptic plasticity. The agent is evolved to perform behavioural preference task where two divergent possibilities of behaviour are presented (approach two different types of lights). By using tools of nonlinear dynamical systems analysis and information theory we are able to determine that some phase-locking relations between oscillators are dynamically linked to particular sensorimotor loops, and to describe in detail how the agent's behaviour emerges as a result of a double loop of interaction between sensorimotor and internal regulatory cycles. Comparing information flows in active situated agents with a *passively-coupled* agent that receives the same recorded signal as the situated agent but is not able to influence the world, we are able to assess the contribution of sensorimotor interaction to the resulting neural ensembles, supporting the idea that the creation and dissolution of transient neural patterns is dynamically intertwined to the sensorimotor coordination between the agent and its envi-

ronment. Moreover, we are able to describe the emergence of an autonomous regulation of the organisms in terms of the asymmetric information flows between different levels of description of the agent's activity, in the form of a double regulatory feedback loop of internal emergent oscillatory patterns and sensorimotor behavioural patterns.

### 3.2 NEUROCOGNITIVE AUTONOMY

We consider two relevant aspects for characterizing neurocognitive autonomy: (1) what kind of neural structures can support a type of organization that dynamically sustains a sensorimotor operational closure required for autonomous agency and (2) how neural, bodily and environmental processes can be coordinated so that this operational closure can be sustained across brain-body-environment boundaries. During the last few decades, interesting advances in both aspects have been developed in systems neuroscience, autonomous robotics, and ecological psychology. In this section, we aim to put them together into a common frame as a conceptual base for our model.

#### 3.2.1 RESONANT CELL ASSEMBLIES IN THE BRAIN

One of the pervasive problems in neuroscience has been that of the *binding problem*, which brings the question of how brains segregate elements in complex perceptive patterns so that they are allocated to discrete objects. It is typically exemplified by the situation of a person looking at a red square and a blue circle at the same time. In this case, the binding problem addresses what kind of mechanisms make sure that, from the discrete perceptions of red, blue, square and circle, the person perceives a red square and a blue circle. Classically, the solution to this problem was constructed in terms of a combination of feature-sensitive neurons organized in a tree-like hierarchical structure for associating different aspects of the environment into coherent percepts. Nevertheless, this kind of approach presents critical deficiencies, such as the combinatorial explosion of possible associations when the system scales up. Today, it is generally assumed that synchronization between neurons correlated with different sensory inputs is a more adequate mechanism for binding different related features of the environment (von der Malsburg, 1995). Purely theoretical formulations of the binding-by-synchrony hypothesis were proposed earlier by Milner (1974), Grossberg (1988), and von der Malsburg (1981), and the Singer lab was the first to obtain experimental evidence supporting the potential role of synchrony as a relational code (see Singer, 1993, 2007). Antonio Damasio proposed phase-locking between multiple sensory and motor brain regions, creating a temporally coherent entity that generates consciousness of the disaggregated representations of information perceived by the senses (Damasio, 1989). Inspired by such findings, Francisco Varela proposes a novel interpretation of these ideas, defending the idea that synchronized groups of neurons or (referred to as cell assemblies or neural ensembles, Hebb, 1952) play a crucial role in the generation of every conscious state (Varela, 1995). For Varela: (1) '[a] singular, specific cell assembly underlies the emergence and operation of every cognitive act' and (2) '[a] specific CA [cell assembly] is selected through the fast, transient phase locking of activated neurons belonging

to sub-threshold competing CAs' (Varela, 1995, p.82). These two ideas are reformulated as the *core hypothesis*, stating that '[m]ental-cognitive states are interpretations of current neural activity, carried out in reference to a transient coherency-generating process generated by that nervous system' (Varela, 1995, p.90-91). Thus, Varela proposes a universal mechanism (supported by the nervous system of most animals) for the generation of a primary consciousness. The emergence of cognitive states depends on the constitution of an assembly that incorporates or discards into a coherent whole different sources of neural activity. This coherence goes further than the binding of particular sensory attributes; it compresses 'the overall integration of all dimensions of a cognitive act' (Varela and Thompson, 2003). In contrast with previous descriptions of mental life, Varela's view does not lie in a topographical space in which different cognitive aspects lie in certain anatomical regions in a topological hypergraph of neural synchronization. The core hypothesis guarantees that mental events occur in a unitary space (since only one cell assembly can be dominant at a particular moment), and that the duration of mental states is transitory (since fast synchrony allows the successive emergence of different assemblies at different moments).

In parallel, an empirically tested theoretical framework to systematize the functioning of the synchronizing brain was being developed by Scott Kelso. This is the framework of Coordination Dynamics, which brings the notion of metastability as a fundamental feature for adaptive behaviour (Kelso, 1995). For Kelso the brain 'is fundamentally a pattern forming self-organized system governed by [...] cooperative interactions among neural clusters' (Kelso, 1995, p.257). But, phase-locking between neural areas is not sufficient to for explain brain dynamics, since 'a nervous system trapped even temporarily in resonant mode-locked states is too stable, and hence too rigid and inflexible' (Kelso, 1995, p.255). In contrast, a model of the brain should contain mechanism for (1) 'entering and exiting coherent spatiotemporal patterns of neural activity' and (2) 'flexibly engaging and disengaging participating subsystems' (Kelso, 1995, p.255). Instead of phase-locking, Kelso proposes the notion of metastable coordination between oscillatory systems (which we described in section 1.2.2). Metastable systems, rather than residing in attractors of a neural network, dwell transiently near attractive states, switching flexibly and quickly between them. The brain is depicted as a 'twinkling metastable system living on the brink of instability' (Kelso, 1995, p.200).

Soon after, Varela's core hypothesis was reformulated by Giulio Tononi and Gerald Edelman, focusing on two complementary properties of the brain: (1) conscious experience is integrated (we experience conscious states as unified and cannot be decomposed into independent components) and (2) at the same time it is highly differentiated (we can experience a vast number of different conscious states over a short time) (Tononi and Edelman, 1998). Tononi and Edelman formulated the *dynamic core hypothesis* as follows:

1. 'A group of neurons can contribute directly to conscious experience only if it is part of a distributed functional cluster that achieves high integration in hundreds of milliseconds'.
2. 'To sustain conscious experience, it is essential that this functional cluster be highly

differentiated, as indicated by high values of complexity’.

For Tononi and Edelman rapid integration or binding must be achieved through strong and rapid reentrant – ongoing, recursive and highly parallel feedback – interactions, creating a dynamic core as a functional cluster of neuronal groups that interact much more strongly among themselves than with the rest of the brain. Furthermore, they propose different mathematical measures to characterize this state sustaining this behaviour in terms of integrated information (Tononi and Edelman, 1998; Tononi, 2004). Subsequently, the notion of a dynamic core has become relatively widespread in large-scale neuroscience studies (Varela et al., 2001a; Seth et al., 2006; Werner, 2007a; Edelman, 2011; Ward, 2011; Le Van Quyen, 2011).

### 3.2.2 RADICAL EMBODIMENT & AUTONOMY

As we saw in the previous chapter, neurodynamic approaches are now well established in cognitive neuroscience but we cannot say the same for embodied dynamical perspectives. Different formulations of the dynamic core hypothesis have stressed the contribution of different parts of the brain to conscious processes: from distributed computation in the cortex (Tononi and Edelman, 1998), to synchronized activity in cortico-thalamic circuits (Llinás et al., 1998), or thalamic neural activity comprising an active blackboard displaying the results of cortical computations (Ward, 2011). Although these approaches do not deny the role of sensorimotor interaction, they are almost exclusively focused on the mechanisms for brain organization and information processing. In contrast, this chapter brings together some insights from both neurodynamics and situated and embodied cognition to offer a brain-body-environment interactive view of the constitution of neurodynamic autonomy. We argue that there is room to make progress in situated autonomous robotics integrating some of the insights and findings related to the dynamic core hypotheses and situated/embodied perspectives of cognition.

The enactive school, in the field of neurodynamics, has contributed with interesting ideas about how a dynamic core may be grounded in the body and the embodiment. For example, the prominent role of sensorimotor interaction has been highlighted in Thompson and Varela’s *radical embodiment* approach to neural dynamics and consciousness, in which they suggest that:

the processes crucial for consciousness cut across the brain-body-world divisions, rather than being limited to neural events in the head [...] Neural, somatic and environmental elements are likely to interact to produce (via emergence as upward causation) global organism-environment processes, which in turn affect (via downward causation) their constituent elements (Thompson and Varela, 2001).

Varela and Thompson propose that the integrated complexity required for conscious activity emerges through the combination of large scale synchronization in the form of phase-

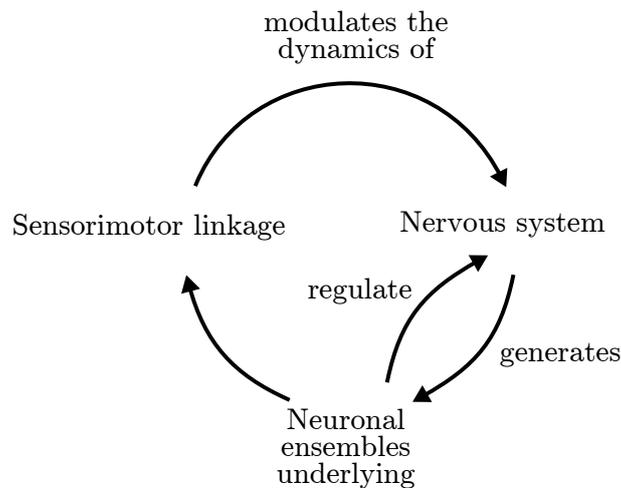


Figure 3.1: Organizational closure of neurocognitive agency. Adapted from Varela (1997).

locking of widely distributed neuronal sub-populations or regions, combined with active uncoupling or phase-scattering mechanisms that mediate the transition between different synchronous neuronal assemblies or cognitive acts (Varela and Thompson, 2003). Synchronizing and de-synchronizing mechanisms must be embedded in organismic, sensorimotor, and intersubjective regulatory cycles, whose interaction is responsible for the emergence of global organism-environment processes that define the cognitive acts that neural assemblies account for. An illustration of this depicted in Figure 3.1, where the coordinated activity of the neural system gives rise to neural ensembles in a circular manner, where the latter drives the behaviour of the organism, which in turn generates a sensory input into the neural system closing the double loop. However, Thompson and Varela do not explain how these cycles are constitutive for cognitive activity, nor what kind of interaction takes place between neural, bodily, sensorimotor and social cycles.

Following this idea, Thomas Fuchs has conceived of the brain as ‘a plastic system of open loops that are formed in the process of life and closed to full functional cycles in every interaction with the environment’ (Fuchs, 2011). In Fuchs’ view, the brain’s structure is shaped by experiential process through brain plasticity, where the ‘spatial, temporal, logical and symbolic structures of interaction processes are extracted and transformed into neural micro-structures that facilitate corresponding future interaction’ (p.211). These microstructures would just create neural dispositions in terms of ‘attractors’ or ‘resonances’. Fuchs argues that primitive organisms without a central nervous system were already coupled to their environment in afferent and efferent cycles. The central nervous system would appear inserted into this already existing system as a transforming and diversifying organ. Thus, the role of the brain is to mediate and modulate the cycles of embodied interaction that the organism engages in. The repertoire of possible behaviours is created interactively in the coupling between behavioural and neuro-plastic dynamics, which is sedimented in the form of organic habits, dispositions

and interactive schemes. Similarly, (Dotov, 2014) suggests that the main function of neural tissue does little more than exhibiting spontaneous synchronization and other behaviours common for nonlinear dynamical systems. For Dotov, the solution to complex behaviour does not lie in a configuration of a massive neural architecture with many independent functional modules, but in how neural tissue is embedded into a field of environmental and bodily activity. Dotov proposes the idea of a *metastable circuit breaker* as a fundamental functionality of the brain that is sensitive simultaneously to multiple threats and opportunities in the environment, and allows the brain to flexibly switch between different dynamic fields of agent-environment engagement. These views may be supported by neuroscientific theories like the *massive redeployment hypothesis* or the *neuronal recycling hypothesis*, which opposing of the idea of the brain as a highly selective and specialized architecture of localized brain functions, they propose that reuse of neural circuitry for various cognitive purposes is a central organizational principle of the brain (Anderson, 2010), extending into reuse of parts of the body and the environment (Anderson et al., 2012).

Along the same line, Xabier Barandiaran proposes habits as a basic building block for autonomous behaviour. He defines a habit as ‘a *self-sustaining pattern of behaviour* that is formed when the stability of a particular mode of sensorimotor engagement is dynamically coupled with the stability of the mechanisms generating it’ (Barandiaran, 2008, p.284). Thus, a habit emerges when stability of both behaviour and the neurodynamical structures of an agent is sustained by the coupling of the two of them through sensorimotor correlations. Habits depends on the plasticity of the neurodynamical structures supporting them, so that the recurrence of a habit reinforces it, and successive enactment of habits can create dynamic stability dependencies between their neurodynamical structures. The latter case gives rise to ‘a nested web of neurodynamic structures: an interdependent bundle of habits that continuously maintains itself through interactions’ (Barandiaran, 2008, p.293). By integrating this notion of habit bundling with large-scale brain mechanisms, Barandiaran is able to establish a notion of self-sustaining autonomous agency that is empirically plausible and sets up an operational framework to specify a simulation model and behavioural autonomy.

### 3.3 A SITUATED DYNAMIC CORE

In Chapter 1 we reviewed how cellular autonomy is proposed to be constituted through the self-production of the cell as a material unit enclosed by a membrane and the interaction between the cell and its environment as it regulates material flows through the same membrane. These theoretical claims were effectively illustrated by different models (e.g. Varela et al., 1974). Nevertheless, in the case of neurodynamic autonomy, although there are many models depicting large-scale neuronal coordination, there are almost no models explicitly targeting how to explore the constitution of an agent as an autonomous sensorimotor entity. Worse still, some of the most prominent models existing when Varela presented his views about neurocognitive autonomy contain serious theoretical pitfalls that make it impossible to use them to understand the interaction between neurodynamic and sensorimotor structures (e.g. the problems

of Bittorio as a model of sensorimotor autonomy; Barandiaran, 2015; Varela, 1988). We therefore consider that is crucial to propose models of neurocognitive identity to clarify some of the surrounding theoretical definitions and reconcile them with experimental evidence.

For example, (Varela, 1997, p.82) describes neurocognitive autonomy as a process in which:

the nervous system is organized by the operational closure of a network of reciprocally related modular subnetworks giving rise to ensembles of coherent activity such that:

- i. they continuously mediate invariant patterns of sensory-motor correlation of the sensory and effector surfaces; and
- ii. they give rise to a behavior for the total organism as a mobile unit in space [...]

The operational closure of the nervous system then brings forth a specific mode of coherence, which is embedded in the organism. This coherence is a cognitive identity: a unit of perception/motion in space, sensory-motor invariances mediated through the interneuron network.

Nevertheless, this definition (and others theoretical definitions of autonomy) remains unclear about specific aspects of the characterization of neurocognitive autonomy: What exactly is the relationship between neural ‘modular subnetworks’ and ‘invariant patterns of sensory-motor correlations’? What kind of ‘specific mode of coherence’ emerges from the operational closure of the system, and how is it ‘embedded in the organism’? How exactly is a ‘unit of perception/motion in space’ constituted? and so forth (for a more complete review of conceptual problems with definitions of sensorimotor autonomy see Barandiaran, 2015).

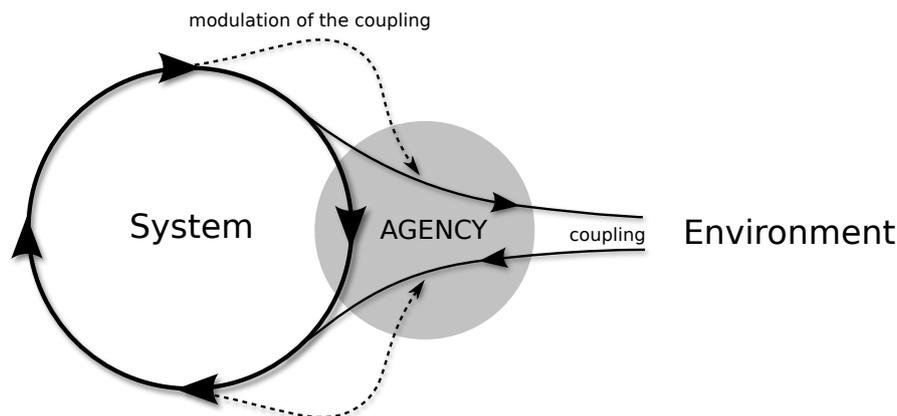
These questions are very difficult to answer from theoretical elaboration alone or from available experimental understanding of the brain. Our goal is to present a model to connect and reconcile theoretical and experimental notions about autonomy, trying to shed some light on how to operationalize neurocognitive autonomy.

We provide a tentative definition of neurocognitive autonomy providing a more specific definition on what are the generative phenomena that allow autonomous agency integrating the theoretical and experimental contributions reviewed so far. Inspired by Barandiaran et al. (2009), we consider an autonomous system *S* and an environment *E*, coupled through a sensorimotor loop *C*, if meets the following conditions:

- I. **OPERATIONAL CLOSURE** among a set of processes a system *S* can be distinguished as a network of interdependent processes whereby every process belonging to the network depends on at least another process of the network and enables at least another one.

2. **INTERACTIVE ASYMMETRY** in the interaction between S and E, which modulates the coupling C between S and E adaptively. Modulation indicates an alteration in the set of constraints that determine the coupling between A and E, and adaptive means that change in the coupling C contributes to the maintenance of some process that constitute S.
3. **PRECARIOUSNESS**. By virtue of the two properties above, the processes that constitute S are actively created and maintained by the system as a whole, so that if we were to isolate from the system any process of S it would tend to run down or extinguish.

This definition is illustrated through Figure 3.2, which illustrates this double loop of operational closure and regulation of such closure. However, without specific models implementing this kind of behaviour it is not easy to illustrate how such a system would operate and what kind of mechanisms might be underlying these processes. Thus, below we review recent advances in modeling in autonomous robotics as a departure point to propose a model for sensorimotor autonomy integrating the dynamic core hypothesis of neural organization.



**Figure 3.2: Autonomous Agency.** The figure illustrates the definition of agency: the system is constituted by a self-sustained network of processes (pictured as a circle, left) coupled to the environment; the systems exerts regulatory constraints over its coupling giving rise to agency. Adapted from Barandiaran et al. (2009).

### 3.3.1 OPERATIONAL MODELS

In the previous chapter we have explored a minimal gradient-climbing robotic model, showing how neurodynamic states of the robot are coupled (through the body) to the environment and thereby deeply grounded in an interactive space of perception/action. The existence of such couplings is essential for the emergence and persistence of structured behaviour. Nonetheless, we were addressing a quite simplified situation in which just one behaviour (gradient climbing) was available for the agent. An autonomous agent, in order to be so-called,

should be able to transit different models of dynamical engagement with its surroundings, learning to adapt to new unencountered situations, and should do so while maintaining its own ‘identity’.

During the last decade, some researchers have taken an interest in redefining the concept of autonomy in robotics. Autonomy in robotics is frequently used in a very loose sense, often applied to any kind of robot that is able to adapt, self-recharge or able to behave robustly in a real-world environment. Nevertheless, the functionalities a robot can display are often explicitly defined by the designer. Even when the robot is able to perform a wide variety of behaviours, they are frequently predesigned and the robot can switch from one another as easily as switching the value of a variable. Researchers as Ezequiel Di Paolo and Hiroyuki Iizuka have argued that these kind of robots cannot be autonomous in a strong sense. For example, Di Paolo and Iizuka (2008) describe an agent that can display two possible behaviours depending on the value of a binary variable. Imagine an agent that displays a preference for a type A of objects when its binary variable is 0, and type B objects when the binary variable is 1. Imagine also that the value of the variable is determined by an internal module of the robot depending on its own state and the state of the environment through a function. In some sense, we may say that the agent is self-determined, since it creates the rules that determine its behaviour. Most of the time we would intuitively say that the robot has some degree of autonomy. However, Di Paolo and Iizuka argue that we may imagine an equivalent situation in which the variable controlling the behaviour of the robot is not inside the robot but in another room, as an independent system (implementing the same function as the robot’s internal module) telling the robot what to do at each moment. Or we may imagine a human operator switching the value of the variable at her will. In those cases we would not think that it is appropriate to call the robot autonomous. The problem here is that the ‘autonomous robot’ in this case is a *system by convention*, and it is only integrated in a nominal sense. If a module behaving independently of the rest of the robot can determine its behaviour (even if it uses the state of the robot as an input) it cannot make the robot more autonomous in any sense. In Di Paolo and Iizuka’s words, ‘autonomy is not a function [...] autonomy is not something that a system does, it is a property of how the system is organized and re-organizes itself so as to channel its functionalities towards newly generated interactions’ (Di Paolo and Iizuka, 2008, p.15-16).

Instead, Di Paolo and Iizuka propose that the very structure of an autonomous robot should by itself select and regulate a sensorimotor flow based on the effects it has on the maintenance of an internal organization. For doing so, they propose a neural architecture based on continuous-time recurrent neural networks with a homeostatic mechanism inspired by Ashby’s Homeostat. They define the net’s neurons as having one (or sometimes more) homeostatic region, that triggers plastic changes in the connections between neurons when the activation rates of the neuron is outside of such zone. Neuronal activity out of the homeostatic bounds would trigger network connectivity changes that would bring the neuron back to its homeostatic activity range. Di Paolo and Iizuka propose that artificial evolution of an agent

with such homeostatic neural controllers is useful to generate a dynamical system that encapsulates the mutual constraining between higher levels of sensorimotor performance and lower neural dynamics.

This approach was first inspired by the studies of readaptation of visual inversion developed in the 50s and 60s by Ivo Kohler (Kohler, 1962), through the use of goggles built with mirrors so that the visual field of the subject had a right-left inversion and other distortions. Subjects wearing distorted glasses at first presented severe difficulties in coordinating behaviour. Although, after some weeks wearing inverted goggles, subjects started to behave coherently, and soon after subjects reported that the whole perceptive left-right regularities started to emerge again in the perceptual experience of the world, re-inverting the visual effect of the goggles. Interestingly, after goggles were removed subjects perceived their visual field as reversed again, and they only recovered ‘normal’ vision after a further process of readaptation. Inspired by such findings, Di Paolo (2000) evolved a robot controlled by a homeostatic neural network to simultaneously perform long-term phototaxis and maintain its internal variables within homeostatic bounds. After evolution, the visual field of the agents was inverted, provoking a breakdown of coordinated behaviour. This breakdown provoked a series of plastic changes in the controller of the agent, that in some of the agents ended up in normal phototactic behaviour after a process of habituation. The agents were also able to adapt to other severe disruptions that had not been experienced before. Di Paolo’s interpretation is that such agents develop a dynamical link between phototactic behaviour and internal homeostasis, so that a disruption of sensorimotor coordination results in internal instability, which triggers plastic internal changes until stability is regained. Di Paolo stresses the circular relationship between plastic neural dynamics and agent-level behaviour, where behaviour is generated and modified by neural activity and behaviour triggers plastic changes when it drives the agent outside of its homeostatic bounds. The result is a subset of behavioural patterns that produce dynamics that maintain the same set of patterns. This is what Di Paolo calls a habit. Furthermore, habits share the properties of operational closure shared by autopoietic systems, guaranteeing the autonomy of their formation.

An extension of this model was developed by Iizuka and Di Paolo to model behavioural preferences (Iizuka and Di Paolo, 2007). This model presented two homeostatic zones instead of one, allowing the agent to display two stable coordinated modes of behaviour. In it, the agent was evolved to perform phototaxis towards two different types of lights, developing self-sustained preferences and switching spontaneously from one light to the other<sup>1</sup>. That is, the agent tended to maintain a preference for the light visited previously, but this preference was transient and at some point the agent ended turning to the other light. An analysis of the agent behaviour revealed that these self-sustained preferences were generated by a flow of alternating high and low susceptibility to sensory influences (Di Paolo and Iizuka, 2008).

Another interesting model of robotic autonomy was developed in parallel by Takashi

---

<sup>1</sup>Note that during evolution the agent was selected to associate each behaviour with one of the two homeostatic regions

Ikegami around the concept of *embodied chaotic itinerancy* (Ikegami, 2007). Using a Fitz-Hugh-Nagumo neural network with delayed signal transmissions Ikegami modelled an agent that moved around a patterned space. The agent displayed a variety of motion styles and a spontaneous selection of different behavioural modes triggered by its embodied dynamics. Chaotic itinerant dynamics is typical of chaotic systems with high degrees of freedom and an itinerant behaviour among chaotic quasi-attractors, with the particularity that in Ikegami's model this dynamics emerge from the embodied agent as a whole. Interestingly, chaotic itinerancy was developed through a sequence of attachment-detachment from sensory influences similar to Di Paolo and Iizuka's homeostatic model. Thus, autonomy of the robot is modelled in terms of unstable coupled neural dynamics, in which the perception of the world is not based on a snapshot of sensory input but in its temporal flow. Ikegami connects this with the idea of autopoiesis as the interplay between motion and internal chaotic dynamics creating an interface between the inside and outside of the system. This dynamic boundary would play a role similar to the membrane of an autopoietic cell in Maturana and Varela's conception. This boundary is maintained dynamically by the internal network, and autonomy is maintained by regulating the flow of sensory input.

Although these models are greatly valuable, we consider that they still fall short of clarifying important aspects of a definition of neurocognitive autonomy. Although the models are designed to shed some light on the kind of mechanisms generating autonomous behaviour and how neurodynamic and sensorimotor structures interact, they do not characterize in depth how the dynamics of the models operate and, instead, the analysis of the robots is based largely on behavioural descriptions. We propose the design of an agent based on the robot modelled by Di Paolo and Iizuka (2008), as well as develop the analysis required to clarify the interrelations between different elements of the model and their relation with our theoretical characterization of autonomous behaviour.

### 3.4 MODEL OF A NEURODYNAMIC CONTROLLER WITH RELATIONAL HOMEOSTASIS EMBEDDED IN A ROBOTIC AGENT

In this chapter we have reviewed different notions about how an autonomous organization might be achieved in the brain. We follow the work of Di Paolo and colleagues on autonomy in evolutionary robotics (Rohde and Di Paolo, 2006; Iizuka and Di Paolo, 2007; Di Paolo and Iizuka, 2008), extending previous models of homeostatic adaptation. Unlike previous models, for which continuous time recurrent neural networks (CTRNN) controlled the robot, this model is implemented in a simulated mobile agent with a plastic Kuramoto network as a neural controller, and with an additional loop of homeostatic regulation by which the homeostatic zone favours some preferred phase relations between the network's oscillators. The simulated robot is presented with two lights of different colours that are perceived by two different pairs of sensors, and evolved using a genetic algorithm to develop transient but robust preferences to the two types of lights. All parameter values (except where otherwise specified) are fixed by a genetic algorithm within the indicated range.

### 3.4.1 NEURAL CONTROLLER: KURAMOTO OSCILLATOR NETWORKS WITH RELATIONAL HOMEOSTATIC PLASTICITY

A fully connected Kuramoto network (Acebrón et al., 2005) with three oscillators is used as the agent’s controller. The evolution of the state of each oscillator is defined by:

$$\dot{\theta}_i = \omega_i + I_i + \sum_{j=1}^N K_{ij} \cdot \sin(\theta_j - \theta_i) \quad (3.1)$$

where  $\theta_i$  represents the phase of oscillator  $i$ ,  $\omega_i$  is its natural frequency (range  $[0, 5]$ ),  $K_{ij}$  is the strength of the coupling between the oscillator  $i$  and the oscillator  $j$ , and  $I_i$  represents the sensory inputs, which are given only to sensory neuronal oscillators 1 and 2.

The model is not intended to represent the activity of individual neurons but, more generally, to capture the dynamics of neural oscillations at a mesoscopic level. Integration mechanisms in the brain are hypothesized to be based on phase synchronization processes between neuronal groups (Varela et al., 2001b), thus we aim to represent the large-scale synchronization of brain regions that are far apart in the brain.

Nevertheless, we hypothesized above that the coordination of neural ensembles alone cannot stand for autonomous behaviour. Instead, they should be modulated by plastic mechanisms shaping the neurodynamic patterns of the agent, coupling it to its sensorimotor interactions while maintaining constant the organization of the agent, i.e., maintaining some relational invariants between the different variables of the agent. (Maturana and Varela, 1980, p.78) describe an autopoietic machine as ‘an homeostatic (or rather a relations-static) system which has its own organization (defining network of relations) as the fundamental variable which it maintains constant [...] autopoietic machines are unities whose organization is defined by a particular network of processes (relations) of production of components, the autopoietic network, not by the components themselves or their static relations’. Following this formulation, if we model autonomous robots controllers as regulators of the relations between internal variables, we can design systems that conserve certain invariants of their organization as a whole, allowing the kind of local-global relationships needed for the maintenance of a unitary identity (see Figure 3.1).

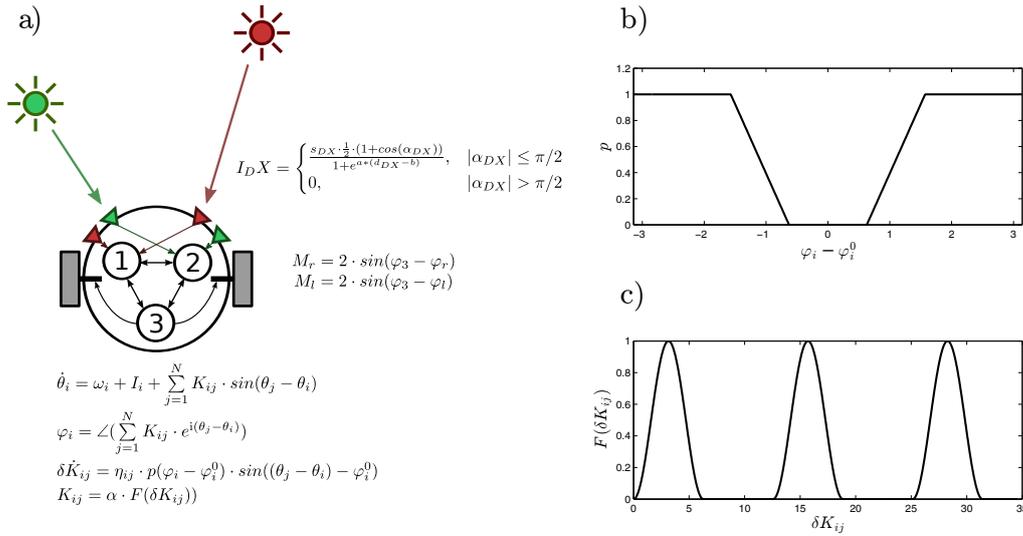
In order to represent relational variables able to preserve some organizational invariances of the system, we have extended the Kuramoto model to include the relations of one oscillator with respect to its neighbours. This new parameter  $\Phi$  is computed by the phase difference of one oscillator with respect to the sum of the oscillator cluster connected to it weighted by the strength of their connections. It is represented by:

$$\Phi_i = \angle \left( \sum_{j=1}^N K_{ij} \cdot e^{i(\theta_j - \theta_i)} \right) \quad (3.2)$$

where  $\angle$  denotes the phase of a complex value and  $i$  is the imaginary unit. Here  $\Phi_i$  represents

the weighted phase relation between oscillator  $i$  with respect to the other oscillators with it is connected.

We add plastic mechanism to the Kuramoto network to homeostatically modify the weights of the connections between neurons. Homeostatic regulation is defined by a stepwise function that determines synaptic plasticity as a function of the phase relations  $\Phi$  between an oscillator and its connected cluster,  $p(\Phi_i - \Phi_i^0)$ , where the homeostatic region is located around a preferred phase relation  $\Phi_i^0$  (range  $[-\frac{\pi}{2}, \frac{\pi}{2}]$ ). The function  $p(x)$  (Figure 3.3.b) is defined by two activation thresholds  $H_1$  and  $H_2$ . The value of  $p(x)$  is 0 when  $x < H_1$ , it increases linearly while  $H_1 < x < H_2$  and it is equal to 1 when  $H_2 < x$ . We have arbitrarily set the values of  $H_1 = 0.2\pi$  and  $H_2 = 0.2\pi$ . That is, neural oscillations connect to each other in such a way that they can show preferred phase relations with other oscillating clusters. This function determines the level of plastic change for all incoming weights. Within the homeostatic region the value of the plastic function is 0, which means that no plastic changes take place while the oscillator stands within the boundaries of its preferred phase relation with the cluster to which it is connected. Though the implemented homeostatic mechanism may be arbitrary, there is evidence to show how brain networks adjust their temporal relations with great precision by plastic mechanisms (Singer, 1999).



**Figure 3.3: The robotic agent with three plastic oscillatory units.** a) Schema of the agent, the environment, sensors and motors, and the neural controller b) Plastic function  $p(\Phi_i - \Phi_i^0)$ , in which plasticity depends on the difference between the weighted phase relation  $\Phi_i$  of the neural oscillator  $i$  with respect to other oscillators and the preferred weighted phase relation  $\Phi_i^0$  c) Mapping function  $F(\delta K_{ij})$  which transforms weight values  $\delta K_{ij}$  into the actual value of coupling strengths between oscillators  $K_{ij}$ .

Changes in the value of the incoming weights of oscillator  $i$  depend on the local plasticity of oscillator  $i$ ,  $p(\Phi_i - \Phi_i^0)$ , multiplied by an extra term which determines the level of synchronization between oscillator  $i$  and each incoming oscillator  $j$ , the plasticity being smaller for

oscillators with higher levels of phase synchronization:

$$\delta\dot{K}_{ij} = \eta_{ij} \cdot p(\Phi_i - \Phi_i^0) \cdot \sin((\theta_j - \theta_i) - \Phi_i^0) \quad (3.3)$$

where  $\delta K_{ij}$  are the connection weights which are initialized randomly at the beginning of each trial and,  $\eta_{ij}$  is the rate of plastic change (range  $[0, 0.9]$ ) of each connection.

Finally, the actual connection strengths are mapped by a modified sinusoidal function of the weights  $\delta K_{ij}$  (denoted by  $F(\delta K_{ij})$ ). A similar strategy was already used in Iizuka and Di Paolo (2008) in homeostatic neural networks, instead of using a Hebbian-like rule, to avoid saturation of synaptic strengths. This mechanism ensures that there is enough variability for plastic reconfigurations in order to explore the whole space of connection weights. In our case, we have modified the sinusoidal mapping multiplying it by a positive square wave with half its frequency in order to allow the possibility of total disconnection between oscillators. Thus, when plastic changes take place, connection strengths follow a continuous non-monotonic function  $K_{ij} = \alpha \cdot F(\delta K_{ij})$  (see Figure 3.3.c) capable of exploring the full configuration space, where  $\alpha$  is a constant (range  $[0, 5]$ ) that regulates how strong the oscillator coupling can be.

In a nutshell, this minimal model works under the assumption that large-scale neural oscillatory components try to maintain an invariant preferred phase relation with respect to other oscillatory components by means of plastically regulating the strength of their connectivity.

### 3.4.2 MINIMAL ROBOTIC EMBODIMENT: BEHAVIOURAL PREFERENCE TASK

The agent is modelled as a simulated wheeled robot with a circular body of radius 4 and two diametrically opposed motors (Figure 3.3.a). The motors can drive the agent forward and backwards. We assume that the agent's mass is small enough to be neglected (in order to avoid inertial resistance), so we can describe the speed of the agent as follows: the translational speed of the robot is calculated as the vectorial average of the motor velocities, and the angular speed as the difference of the motor velocities divided by the body diameter. The motor outputs are calculated from the phase relation  $\Phi$  of the effector oscillator (oscillator 3). The speed of the motor is obtained by multiplying the motor output by a gain parameter of value 2:

$$\begin{aligned} M_r &= 2 \cdot \sin(\Phi_3 - \Phi_r) \\ M_l &= 2 \cdot \sin(\Phi_3 - \Phi_l) \end{aligned} \quad (3.4)$$

where  $\Phi_r$  and  $\Phi_l$  (range  $[0, 2\pi]$ ) are bias terms which map the motor output into the actual motor activation.

The agent has two pairs of sensors (right and left) for each of the different light sources A and B. Each sensor points to a direction at  $\pi/3$  radians from the forward direction. Light A sensors are connected to oscillator 1 and light B sensors are connected to oscillator 2. The activation of the sensor depends on the angle between the sensor and the light, such that the maximal activation happens when the sensor faces the light. The effects of both the angle and

the distance on the sensor activation are represented by the following function:

$$I_{DX} = \begin{cases} \frac{s_{DX} \cdot 0.5 \cdot (1 + \cos(\alpha_{DX}))}{1 + e^{a \cdot (d_{DX} - b)}}, & |\alpha_{DX}| \leq \pi/2 \\ 0, & |\alpha_{DX}| > \pi/2 \end{cases} \quad (3.5)$$

where  $X$  can represent either light A or B,  $D$  stands for either right or left sensor,  $\alpha_{DX}$  is the angle of sensor  $DX$  to light  $X$ ,  $d_{DX}$  is the distance between sensor  $DX$  and light  $X$ , and  $a$  and  $b$  have the arbitrary values of 0.03 and 100 respectively. The light intensity received at each sensor is multiplied by a gain parameter  $s_{DX}$  (range  $[-8, 8]$ ). The resulting value is fed to the corresponding oscillator's input  $I_i$ . A full schema of the robot is represented in Figure 3.3.a.

### 3.4.3 ARTIFICIAL EVOLUTION: FITNESS FUNCTION

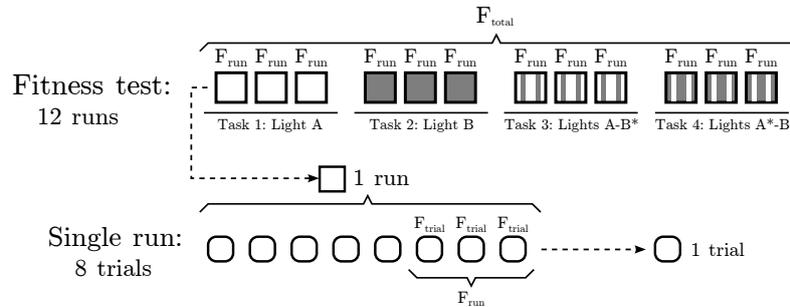
Interactions between coordinating neural units, plastic mechanisms and higher order behavioural patterns at the whole-agent level can take many complex forms that are not easy or intuitive to design. Frequently, these interactions cannot be reduced to one-to-one relations but many-to-many interactions. Evolutionary robotics has been a greatly productive tool to model systems exhibiting such properties and finding non-intuitive solutions to complex problems (Harvey et al., 2005; Nolfi and Floreano, 2000), therefore, we have chosen to use a genetic algorithm to model our agent.

A population of 20 agents is evolved using a rank-based genetic algorithm with elitism. Each of the agent parameters  $\omega_i$ ,  $s_{DX}$ ,  $\alpha$ ,  $\eta_{ij}$ ,  $\Phi_r$ ,  $\Phi_l$  and  $\Phi_i^0$  is encoded into a 5 bits string representing a real number within the specified range. For each generation, the best 4 agents (20% of the population) pass to the next generation without change. For the remaining slots, pairs of individuals are selected for crossover with a probability proportional to their fitness value, and new individuals are created mixing their genes (bit series) adding a mutation probability of 3% for each gene.

The agents are evaluated for 4 different tasks as in Iizuka and Di Paolo (2007): a single light A, a single light B, one light A and a blinking light B, one light B and a blinking light A. In the two first tasks, only one light (either A or B) is present, and the agent gains fitness by approaching the light. In the two latter tasks, two lights are presented (one of type A and one of type B) and one of them is blinking. The agent gains fitness by approaching the non-blinking light. The blinking light emits light only with a probability of 0.15 for each time step. The objective of this is to create a 'dummy' that encourages the agent to learn to ignore one of the lights while approaching the other. The lights appear at a random distance,  $[100, 150]$ . When two lights are present, they appear, from the agent's point of view, with a random separation within the range  $[\pi/2, 3\pi/2]$ . The length of each trial is 125s.

Each individual agent is tested for 12 independent runs (3 for each of the 4 tasks, see Figure 3.4). Each run consists of some trials where a light or a pair of lights are presented to the agents for a fixed time. At the beginning of each run, the synaptic weights  $\delta K_{ij}$  are reset to initial random values (note that this is different from (Iizuka and Di Paolo, 2007), where weights were reset to a fixed initial value obtained by evolution). Each run consists of 8 trials. For each

trial, one or two lights (depending on the task) are presented to the agent for a specified time. After the trial is finished, a new trial begins and two new lights are presented at new positions. Only the last 3 trials of each run are evaluated in order not to penalize slow plastic changes and bootstrap evolution.



**Figure 3.4: Structure of the task evaluated in the genetic algorithm.** Each agent is tested for 12 different runs (3 for one of the 4 tasks), averaging their fitness values to obtain the total fitness. Each run consists of 8 different trials. For each trial, two lights are presented to an agent during a period of 125s. Only the three last trials of a run are taken into account for computing the fitness of the run in order to avoid penalizing slow plasticity.

Fitness for each trial is calculated in three terms,  $F_{trial} = (F_D + F_p) \cdot F_H$ . The term  $F_D$  represents how close to the light the agent has got at the end on each trial. For each trial,  $F_D = 1 - d_f/d_i$ , where  $d_f$  and  $d_i$  respectively correspond to the final and initial distances to the target light.  $F_p$  indicates the proportion of time that the agent spends within a distance of less than 4 times its body radius (i.e. a distance of 16) to the target light during a trial. Finally,  $F_H$  represents the proportion of homeostasis in the system, computing the degree of homeostasis  $1 - p(\Phi_i - \Phi_i^0)$  (i.e., 1 minus the level of plasticity) for each oscillator and averaging the result over time and over the three oscillators. This way we select agents that remain as homeostatic as possible. This combination of fitness ensures that the agent develops a phototactic behaviour (through  $F_D$  and  $F_p$ ) and also that this behaviour is evolved to display patterns of activity that lie within the homeostatic bounds of the neural controller (through parameter  $F_H$ ). The total fitness is calculated and then averaged over all 12 runs <sup>2</sup>

### 3.5 BEHAVIOURAL ANALYSIS OF THE EVOLVED AGENT

Once the agent is defined and we have evolved its parameters to perform the behavioural preference task, we want to analyse in detail its behaviour, with special interest in the interaction between oscillator dynamics, synaptic plasticity and the agent's behaviour. Our hypothesis is

<sup>2</sup> The code simulating the behaviour of the agent and the parameters obtained from the genetic algorithm can be accessed from the following repository <https://github.com/IsaacLab/HNA-robotic-model/tree/master/minimal-preference-task>.

that the mechanisms of relational homeostasis embedded in the agent's controller can exemplify some characteristics of neurodynamic autonomy, in particular the operational closure of the system through the coupling of neurodynamic and sensorimotor structures in a web of self-sustaining behavioural patterns or habits.

After running the genetic algorithm we select the evolved agent from the last generation that obtained the best fitness value. In what follows we analyse the behaviour of this agent but the results can be extended to many other agents that we found to display similar dynamics (i.e. the agent under analysis in this section displays a typical behaviour of this and other evolutionary runs). The different agents are tested in an environment where the two lights have the same intensity (no dummies). Note that this task (two lights presented with the same intensity) was never experienced during evolution.

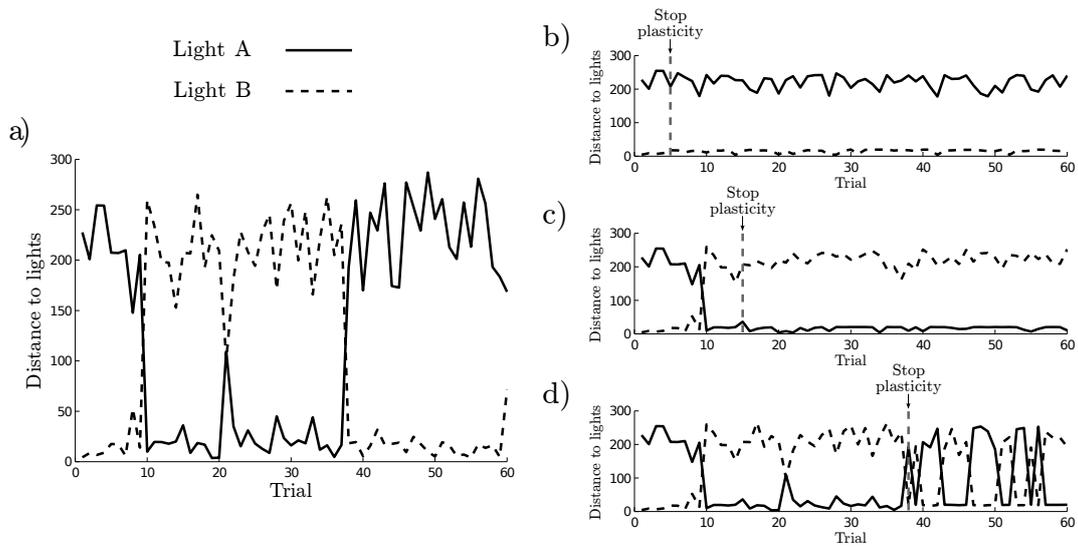
First, we describe the behaviour of the agent under different conditions, showing that habits emerge in the form of self-sustained behavioural preference patterns. Second, we describe the dynamics of the agent oscillatory and synaptic patterns. Finally, we perform a detailed analysis using information theory tools to understand the interplay between the different elements of the system for generating the emergent patterns of behaviour.

### 3.5.1 DESCRIPTION OF BEHAVIOURAL PATTERNS

The first step consists of analysing the behaviour of the agent in a descriptive manner. We present the agent with a series of 60 pairs of lights, and we provide the agent a time of 125s to choose and approach one of them. We measure the final distance to both lights at time  $t = 125s$ , obtaining the series represented in Figure 3.5.a. As shown, the agent is able to develop stable preferences towards one of the lights, maintaining it for several trials until the preference is changed. A video of the behaviour of the agent (including plastic mechanisms) can be found at <https://vimeo.com/53847420>.

We can test the role of plasticity in the emergence of a new preference by stopping plastic mechanisms at a particular point. We have taken the results of the simulation represented in Figure 3.5.a, and run the simulation again under exactly the same initial conditions but freezing the plasticity in the midpoint of a particular trial. We observe in Figure 3.5.b and 3.5.c that when plasticity is stopped the preference is *frozen* and the robot always chooses the same light. Thus, we can say that plastic mechanisms mediate *the creation and destruction of behavioural preferences*. These results show how preferences arise from a complex interplay between the agent's interaction with its environment and the neural mechanisms of synaptic plasticity. That is, preferences emerge and are maintained when a certain plastic configuration (e.g. sensitiveness to light A and blindness to light B) are met with certain environmental conditions that allow the agent to maintain its configuration (e.g. the stimuli received from the lights do not trigger plastic mechanisms that can destroy the current sensitiveness of the agent).

Moreover, if we try to stop plasticity during a transition trial (precisely at the moment in which the agent is about to switch its preference, as shown in Figure 3.5.d), we observe

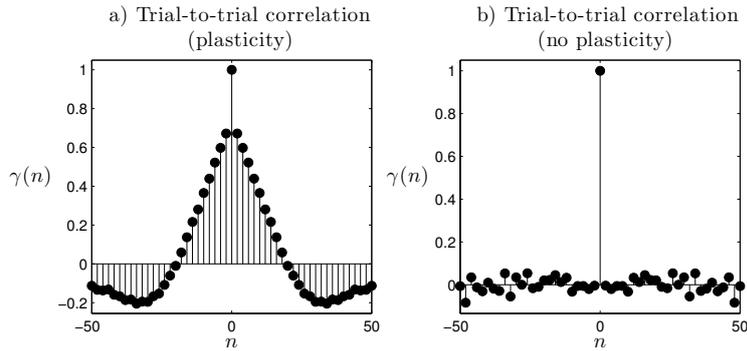


**Figure 3.5: Distance to the two lights at the end of each trial for different agents.** A normal agent a), and three agents in which plastic mechanisms have been frozen at different instants; b) in a situation of preference to light A (trial 5); c) in a situation of preference towards light B (trial 15), and d) at a moment in which the preference switches from B to A (trial 38).

that the agent continuously alternates between the two lights, being unable to hold a specific preference. When the weights are frozen at specific transition points (like the agent in Figure 3.5.d) the agent seems to go indifferently to either one of the two lights. We call these *undecided* agents. However, the question arises as to whether these agents are really indifferent for choosing one light or the other, or do they have preferences that simply last for a shorter amount of time?

To answer this question we simulate different runs of an agent with synaptic plasticity and several agents without synaptic plasticity of the undecided type. We run the simulations for 1000 trials. For each simulation, we create a time series with the result of the 1000 trials, having a value of 1 when the agent goes to light A, and  $-1$  when light B is chosen. The time series  $L(n)$  represents the *decisions* of the agent for all trials  $n$ .

First, we can analyse whether the sequences of consecutive trials choosing the same light constitute a consistent preference, or if the agent just chooses randomly and the sequences of same-light consecutive trials are merely coincidences. We hold the hypothesis that the agent we called ‘undecided’ (whose connection weights are frozen from an instant in which the agent was just switching preferences) does not really possess an internal bias for any light source and whose behavioural choice is merely the result of environmental contingencies. As the configuration of the environment at each trial is completely random, the series of decisions  $L(n)$  will be totally uncorrelated if there is no mechanism that makes subsequent decisions dependent on the choice of the agent in this trial. Thus, we can test the existence of consistent preferences simply by computing the autocorrelation of  $L(n)$ , obtaining  $\gamma(n) = \sum_{\tau=-1000}^{\tau=1000} L(\tau) \cdot L(\tau + n)$ .



**Figure 3.6: Cross-correlation in the series of agent's decisions.** We compute the cross-correlation between the time series of the decisions made by the agent at the end of each trial (where 1 means light A and  $-1$  light B). We observe how consistent trial-to-trial cross-correlations only arise for the agent with synaptic plasticity.

In Figure 3.6.a we observe how the agent with plasticity displays strong correlations in its sequence of decisions. The correlation function shows how decisions are positively correlated to the decisions taken in the 20 previous and posterior trials, suggesting that this is the average duration of preferences. In contrast, we see in Figure 3.6.b how the ‘undecided’ agent is genuinely undecided and presents no correlations between one decision and the next, meaning that decisions are determined by the randomly generated configuration of the environment. We observe how synaptic plasticity indeed plays a crucial role in the emergence of behavioural preferences, since it is a necessary element to produce correlations between one decision and the next.

#### DEGREES OF ENGAGEMENT WITH THE ENVIRONMENT

We have demonstrated that plasticity allows the agent to generate preferences that are maintained during a number of trials and then vanish to give place to a new preference. How do preferences arise from the interaction of plastic mechanisms and different modes of engagement with the environment? For testing this, we reinitialize the simulation with the same conditions as those shown in Figure 3.5.a, but this time we fix the initial distance of both lights to 125. For each trial, we perform the following experiment: we record the initial conditions of the trial (robot position, network state, etc.), and we simulate the same again except that we modify the angle of the position of the two lights, trying 800 different possible combinations, with the angle between the two lights between the interval  $[\pi/2, 3\pi/2]$  (this analysis was also performed for a different homeostatic agent by Di Paolo, 2000). In Figure 3.7.c we can see the behaviour of the robot for trials 30, 38 and 46, where light grey dots correspond to a situation where the final distance of the agent to light A is less than 50 (that is, we interpret that the agent has chosen a light when it reaches closer than 50 units to the light, considering that the initial distance is between 100 and 150 units), dark grey to the final distance of the agent to light B being less than 50 (that is, we interpret that the agent has chosen when the agent reaches closer than 50 units) and black dots to a situations where none of these condi-

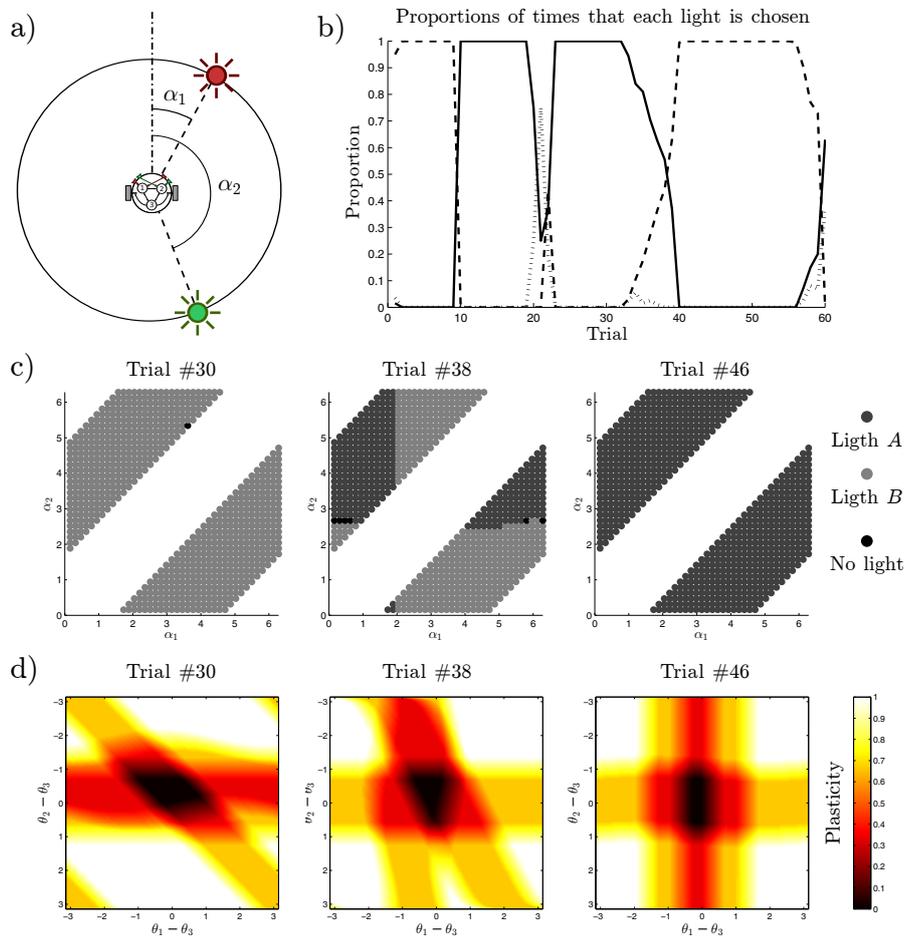
tions are met. For the whole simulation, we can plot how the proportion of light/dark grey and back dots changes from trial to trial (Figure 3.7.b), seeing how the behavioural preference is modulated over time. We can observe how there are periods where the preference for one of the lights completely dominates the other (no matter what the positions of the light are), and periods where the agent is *open* to external stimulus and they can influence the action developed. This suggests some kind of spontaneous intermittency in which the behaviour of the robot switches between states in which it is just driven by internal preferences and it is ‘blind’ to environmental or external factors, and states in which the agent is very sensitive to the configuration of its environment and the emergent preferences depend on the kind of interaction between agent and environment.

Also, we can analyse the homeostatic region of the agent at each moment of time. Since homeostasis is not defined for individual variables of each oscillator but for relational variables  $\Phi_i$  (which depend on the phase relations  $\theta_j - \theta_i$  and the weights of their synapses  $K_{ij}$ ), if we define the homeostatic region in terms of the phases of the oscillators  $\theta_i$ , what we see is that the homeostatic region for each individual variable changes with changes in  $K$ . We can define the global plasticity of the system as  $P = \sum_{i=1}^3 p(\Phi_i)$ , and plot the homeostatic region in terms of the phases  $\theta_i$  for a particular state of synapses  $K_{ij}$ . We can observe the homeostatic regions at the beginning of trials 30, 38 and 46 in Figure 3.7.d (note that  $\theta_1 - \theta_3$  and  $\theta_2 - \theta_3$  represent all possible phase combinations since  $\theta_1 - \theta_2 = (\theta_1 - \theta_3) - (\theta_2 - \theta_3)$ ). Furthermore, the homeostatic box containing the viable region for phase relations changes when the preference of an agent is changed, from a horizontal box that triggers plastic changes when  $\theta_2$  and  $\theta_3$  lose their synchronization, to a vertical box that triggers plastic changes when  $\theta_1$  and  $\theta_3$  lose their synchronization. In the middle, we find a state of transition where the shape of the box is between the other two. Whereas in the original model from Di Paolo and Iizuka (2008) different preferences were defined a priori forcing the behaviour to be constrained by two separated homeostatic regions, our model presents a unique box that changes its shape when synaptic changes take place.

### 3.6 DYNAMICS AND INFORMATION FLOWS IN THE SYSTEM

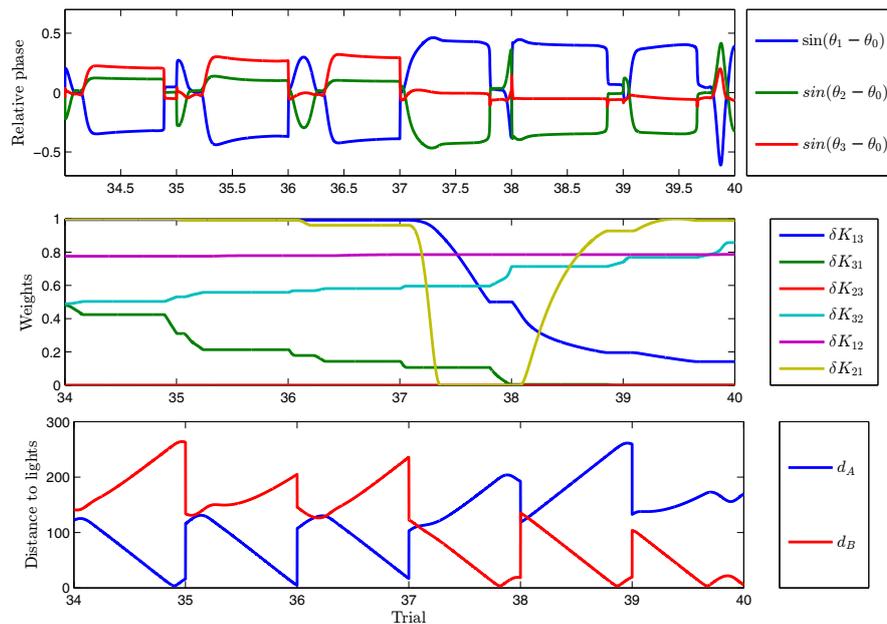
So far, just analysing the behaviour of the agent under some controlled conditions, we have observed the interplay between the emergence of different behavioural patterns and the plastic changes triggered when the relational homeostasis of the oscillator network is challenged. This interaction generates self-sustained patterns of behaviour characterized by preference towards one of the lights. Moreover, in some cases, these patterns are not closed to attend or approach a single light source, but they are open to perform one or other action depending on the position of the light. Nevertheless, in order to get a better understanding about how these interactions operate between the different elements of the neural controller and the agent’s behaviour we need more sophisticated tools. Thus, we make use of dynamical systems and information theory analysis to get a better understanding about the agent’s behaviour.

We can observe any particular moment of time, and observe how conduct is correlated



**Figure 3.7: Openness of the agent to each type of light.** For each trial in the simulation, we reset the agent with the same initial parameters and change the angles of the two lights. In b) we can see the behaviour of the robot for trials 30, 38 and 46, where light grey dots correspond to a situation where the final distance of the agent to light A is less than 50, dark grey to the final distance of the agent to light B being less than 50 and black dots to situations where none of these conditions are met. In c) it is displayed how the proportion of light/dark grey and back dots changes from trial to trial. In d) we can observe the shape of the homeostatic 'box' in which the behaviour does not trigger plastic changes.

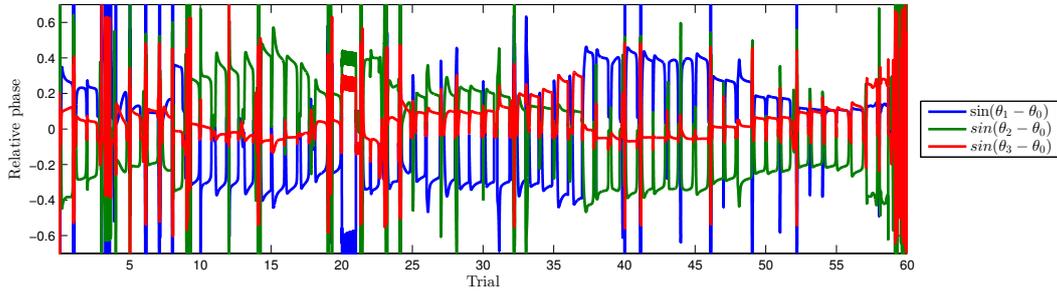
with behavioural changes in different variables of the system. For example, in Figure 3.8, we can observe how a specific preference switch is correlated with changes in both the oscillatory patterns that are stabilized and the change in the synaptic weights values. The oscillatory patterns are defined as the relative phase of the oscillators  $\theta_i$  with respect to an average phase  $\theta_0 = \angle(\sum_{j=1}^3 e^{i\theta_j})$ . The question to answer here is what the relation between behavioural, oscillatory and synaptic patterns is. Does one of them determine the behaviour of the other? Are they co-regulated in any way? Here, our task is to disentangle the dynamic and causal relations between the different levels of the system.



**Figure 3.8: Correlation between behavioural and neural patterns.** All patterns are defined for 6 specific trials in which a change of behavioural preference takes place. Oscillatory patterns (top) are defined as the relative phase of the oscillators  $\theta_i$  with respect to an average phase  $\theta_0 = \angle(\sum_{j=1}^3 e^{i\theta_j})$  (top). Weight values (middle) are represented as a function between 0 and 1. Behavioural patterns (bottom) are defined as the distances from the agent to each one of the lights.

For example, we can analyse the full series of oscillatory patterns. In Figure 3.9 we can observe the relative phase of the oscillators  $\theta_i$  with respect to  $\theta_0$ . Comparing the results with Figure 3.5.a, we can observe that particular configurations of the oscillators are sustained in coordination with different modes of behaviour. We can consider that each possible configuration of the phases of the oscillators correspond to a particular cluster of neurons that emerges sustaining particular cell assemblies in the neural controller of the agent. But, how are these cell assemblies generated? What is their relation with plasticity, and with behaviour? Throughout the rest of this chapter we will unveil these intricate relations between these non-linearly coupled elements.

Although we can visualize the evolution of phase relations in our oscillatory controller, it is much harder to analyse the evolution of their synaptic connectivity (since with 6 variables the



**Figure 3.9: Oscillatory patterns in the neural controller.** Sine of the relative phase of the oscillators  $\theta_i$  respect to an average phase  $\theta_0 = \angle \sum_{j=1}^3 e^{i\theta_j}$  during 60 trials for a normal agent. We observe that particular configurations are maintained in coordination with the behaviour of the agent depicted in Figure 3.5.a.

possible relations grow exponentially). In order to analyse the role of synaptic connectivity, we need to find useful descriptions of the patterns that arise for synaptic plasticity.

### SYNAPSEMBLES

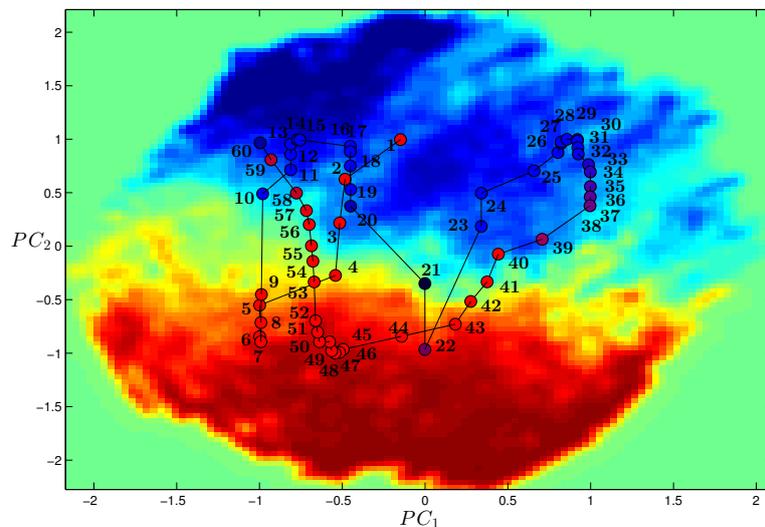
Although we have observed the evolution of neural ensembles in Figure 3.9, if we analyse just the state of the oscillators in the system alone we are just capturing a small space of the variables controlling the agent. Although the agent is composed of just three oscillators, its behaviour it is also controlled by 6 extra differential equations corresponding to the change of the synaptic weights between the oscillators. Similarly, although activation of neurons is generally regarded as the main source of information in the brain, short-term synaptic plasticity can also serve different behavioural functions (Abbott and Regehr, 2004; Mongillo et al., 2008). Moreover, since there are orders of more synapses in the brain than there are neurons, they are being postulated as a richer source of information. Inspired by this idea György Buzsáki has focused on the notion of *synapsemble*, in which analogously to cell assemblies ‘a particular constellation of synaptic weights in a defined time window can be conceived of as an assembly of synapses or “synapsemble”’ (Buzsáki, 2010, p.12). Buzsáki hypothesizes that synapsembles may be critical for (1) building up and dissolving cell assemblies, limiting their lifetime to timescales no longer than a few seconds, and (2) linking together sequences of cell assemblies despite the cessation of spiking activity.

Although experimental evidence supporting the role of synapsembles is hard to find in the brain (because of the lack of tools to directly measure synaptic connectivity in a behaving animal), our model allows us a detailed analysis of the evolution of synapse ensembles since we have access to the state of the weights at any moment. Nevertheless, since there are 6 different synapses we cannot display the dynamical field of synapses in three dimensions. Thus, we have performed a PCA analysis for the states of the six weights. We have simulated an agent for 1000 trials of 125s. For each pair of neurons  $i$  and  $j$ , we have fed the values of  $\sin(0.5 \cdot \delta K_{i,j})$  (since the values of the weights are  $4\pi$  periodic) to the PCA algorithm. In order to display the behaviour of the system, we take the first two components of the PCA analysis, which stand

for 54.2% of the variance (which means that a significant proportion of the variance cannot be captured). For each value of the agent weights, we computed the proportion of the trials this value showed when the agent was approaching a type A light or a type B light. The difference between the two calculations gives us a value  $L$  between  $-1$  and  $1$  ( $1$  for all A lights, and  $-1$  for all B lights).

### SYNAPSEMBLES VS. BEHAVIOURAL MODES

To compute the landscape of the agent's behaviour for each configuration of weights, we mapped the values of  $L$  over the first two components of the PCA, discretizing their values using 20 bins and convolving the result with a bidimensional sigmoid kernel with a standard deviation equivalent to one bin. We plotted the result in figure 3.10, where the red areas correspond to behaviour of the trials where the agent goes to light A and blue areas to light B. On top of the landscape, we plotted the sequence of states at the beginning of the trial of the sequence displayed in Figure 3.5, where we can observe how the synapsemble configuration of the agent evolves for sixty trials. The colour of the dots corresponds to the proportions each light is chosen during that particular trial on different initial conditions as we did in Figure 3.7.



**Figure 3.10: Dynamical landscape of synapsemble configurations.** The values of the synaptic configurations are collapsed into the first two dimensions of a PCA algorithm. Colour of the map represents the percentage of times a particular state is associated with a trial where the agent goes to a particular light. The dots correspond to the trajectory of the weights at the beginning of the trials of the run represented in Figure 3.5.

We can see how this landscape of synapsemble configurations has areas that are clearly linked with one or other behavioural mode. For example, trials 10 to 20 are clearly connected

with preference towards light B, or trials from 41 to 53 are clearly connected with preference towards light B. Nevertheless, there are moments in which the agent enters a different area of the landscape but keeps the previous behaviour for some trials as we can see in trials 54 to 59. This suggests that synapse configurations do not completely determine the behaviour of the agent, so a more complex interplay between cell assemblies and synapse configurations may be taking place here. However, this is not clear since we are missing part of the picture since the PCA analysis only allows us to analyse a 54.2% in the synapse configuration. Thus, we need a visualization of the behaviour of the agent that shows a better representation of all the dimensions of the agent controller.

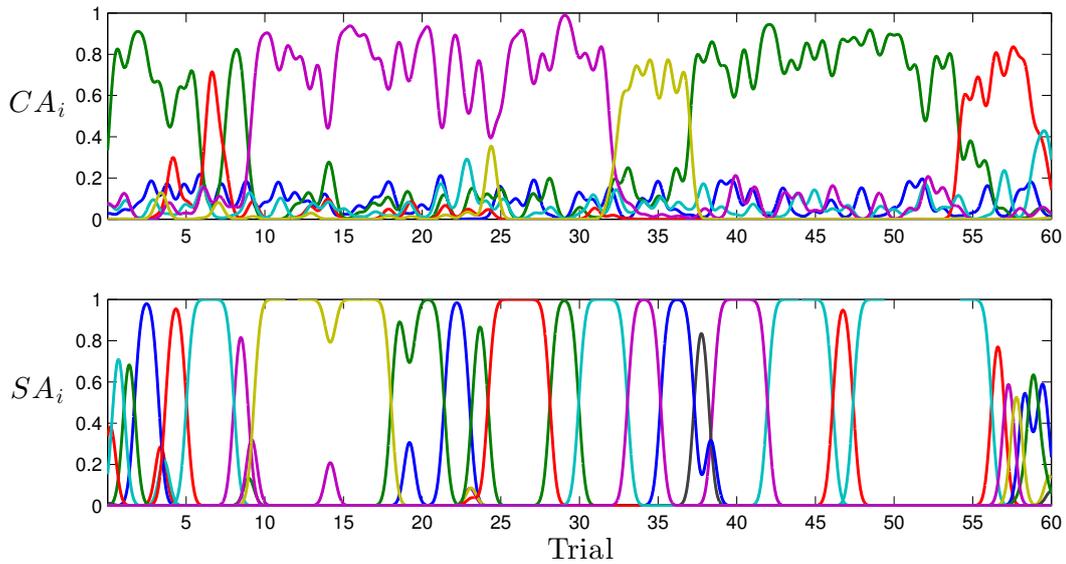
#### COORDINATION BETWEEN CELL AND SYNAPSE ASSEMBLIES

Analysing the coordination between neural ensembles and synapse configurations is not an easy task due to the moderately high dimensionality of the system (9 dimensions of the neural controller, plus the dimensions of body and environment). For this reason, we have simplified the analysis by reducing both the state of neural ensembles and synapse configurations to two discrete values representing the state of a network. To accomplish this, we define different discrete states based on the configuration of cell and synapse assemblies. Since Figure 3.9 suggests that the configuration of the oscillators in the assembly is the relevant variable for generating one or other type of behaviour, we define cell assemblies depending on the relative phase of the oscillators. We can codify the state of the cell assembly with a string of three bits  $CA$  (where bit 1 encodes whether oscillator 1 precedes oscillator 2, bit 2 whether oscillator 1 precedes 3, and bit 3 whether oscillator 2 precedes 3; note that from all the three combinations of these conditions only 6 are non-contradictory and therefore possible). As for the synapse assemblies, we define each synapse as active or inactive if the value of the synapse is higher than  $\frac{\alpha}{4}$ <sup>3</sup> (recall that  $\alpha$  is the maximum possible value for the weights). The state of the synapse configuration  $SA$  is codified with a string of 6 bits (where each bit codifies whether one of the synapses is active).

We define whether a particular cell or synapse ensemble is active using a set of binary variables  $CA_i$  and  $SA_i$ , which represent if a specific ensemble is active at a particular moment of time. For improving the visualization, instead of binary activation values, we filtered the resulting signal using a Gaussian kernel with a standard deviation of 12.5s to represent the average activation over a window of time. In Figure 3.11 we can observe the evolution of  $CA_i$  and  $SA_i$ . The figure shows that, although cell assemblies may be connected with particular behaviours of the agent, at different moments there are always different cell assemblies struggling to emerge. Also, the interaction between cell ensembles influences the change of synapse configurations at a lower timescale, creating and destroying preference patterns in the behaviour of the agent as the state of synapses travels through the complex landscape we characterized in Figure 3.10. On an even longer timescale, some cell assemblies are sustained for even longer than the duration of particular synapse configurations, emerging over all the others for several trials.

---

<sup>3</sup>Since  $\alpha$  is the maximum possible value of the weights,  $\frac{\alpha}{4}$  is an arbitrary threshold for capturing when weights are active and sufficiently strong. Different values have been tested without altering the results significantly.



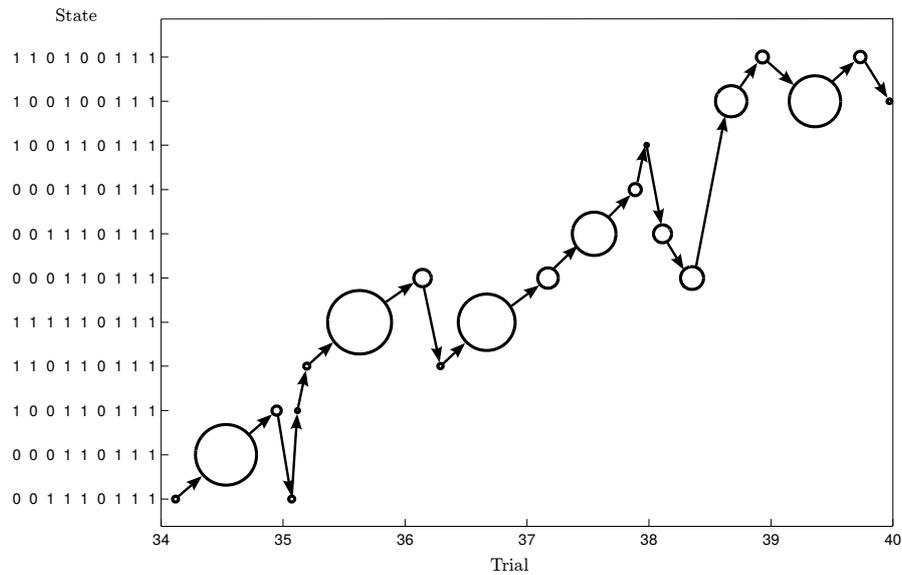
**Figure 3.11: Evolution of neural ensembles and synapse ensembles.** Each line of the plot represents the activation of a particular cell or synapse ensemble, filtered by a Gaussian kernel with a standard deviation of 12.5s to represent the average activation over a window of time.

## NETWORK VISUALIZATION

How can we map these complex relations at different timescales? In a first approach we can make use of a network visualization of the state of the system. Since the PCA analysis has problems for capturing the dimensionality of the system just in the case of the synapse dynamics, the discrete description of our agent will allow us to depict how the system travels through different patterns.

Having defined the patterns of the system, we can define a network in which each state corresponds to a node, and transitions between patterns are represented as edges of the network (Figure 3.12). In addition, we map the amount of time the system spends in any one state into the size of the node. Finally, for each state, we compute the relative amount of time that state appeared in trials in which the agent was approaching light A and light B into a variable  $L$ , whose values lie between 1 and -1. A trial was assigned to a particular light when at the end of the trial the distance from the agent to the light was less than 50 units. We map the values of  $L$  into the network into the colour of the nodes, which ranges from blue to red proportionally with the value of  $L$ .

Now that we have codified the oscillatory, synaptic and behavioural patterns of the agent, we can easily represent the evolution of the system as a discrete system. For example, if we may represent the state of the neural controller for the series represented in Figure 3.8. We can join the patterns  $CA$  and  $SA$  into a string of 9 bits to describe the state of the controller in a particular moment (Figure 3.8). In such a way, we can describe the system as a state diagram.



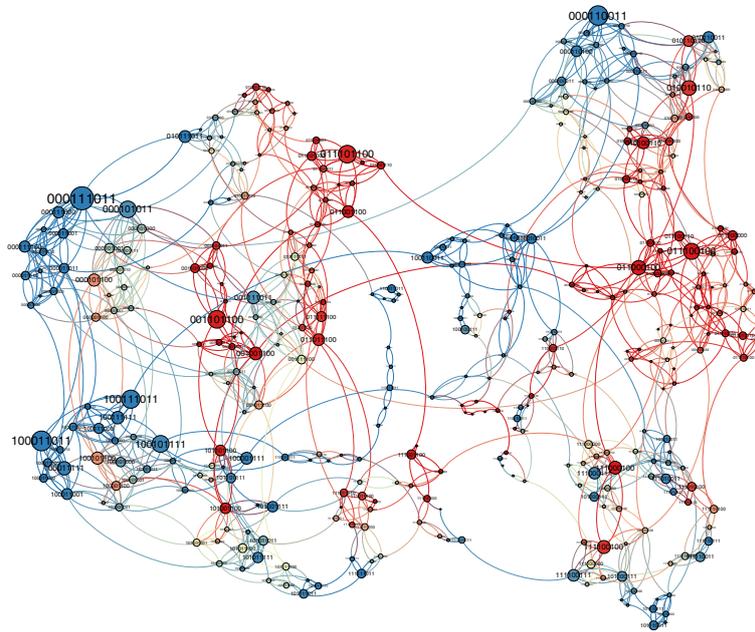
**Figure 3.12: State diagram of the discretized states of the agent.** The series of states represent the transitions of the evolution of CA (3 former bits) and SA (6 latter bits) in the series depicted in Figure 3.8.

We know that neural ensembles in the agent emerge displaying different phase-locked relations made possible or constrained by their synaptic configuration, that is in turn triggered when the oscillator network is not able to maintain some relational homeostatic invariances. Moreover, the stability of patterns of behaviour is tightly coupled to the behaviour of the agent itself. How exactly do these three levels of behaviour relate to each other? Is there a role for situatedness and motility in the generation of the agent behaviour, or it is just reacting to particular inputs in a non-autonomous way? Here we analyse in detail the behaviour of the agent to understand this complex loop of interaction. As in Chapter 2, in order to discover the influence of the sensorimotor loop in behaviour-generating mechanisms we compare a situated agent with a disrupted sensorimotor loop. To achieve this we compare:

1. A situated agent with normal sensorimotor interaction.
2. A passively-coupled agent, in which the input fed to the agent is recorded from a real interaction from the situated agent, thus decoupling the received input from the activation of the motors but maintaining the structure of an input generated by a situated agent.

In Figure 3.13 we depict the resulting network. We can observe that the network is structured in different clusters, which in general correspond to the same behavioural pattern (all nodes have the same colour, see e.g. the bottom left cluster and the middle right cluster). Also, some clusters present nodes of both colours (e.g. top right cluster). Moreover, we can observe that, in general, nodes belonging to a particular cluster share their first threebits, meaning

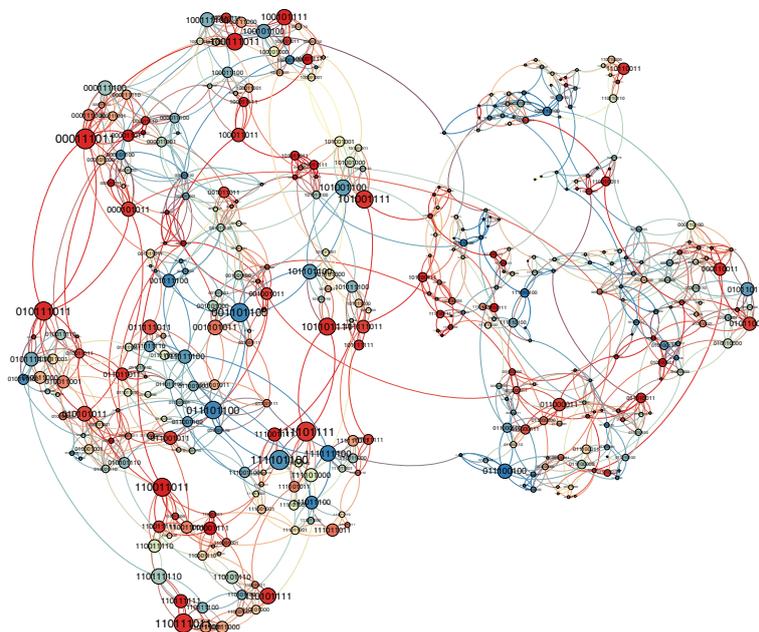
that they correspond to the same cell assembly. With respect to the second string of six bits of each cluster, they generally comprise a subset of the possible synapsembles (there is a total of 64 possible synapsembles), suggesting that each possible cell assembly is sustained by a set of possible synapsembles, and that these synapsembles may change within the same set maintaining the resulting assembly. This is also what we observed in Figure 3.11. The global picture is a map of the mental life of the agent, with clusters of dynamical patterns corresponding with particular modes of behaviour, and clusters corresponding with states in which the agent may choose different behaviours. In these clusters the stability of a set of neurodynamical states is coupled with the stability of a behavioural pattern, as in the definition of habit by Barandiaran (2008) we described in section 3.2.2. The different clusters and the paths between them display a web of neurodynamic structures that is tightly coupled with the behaviour of the agent.



**Figure 3.13: Network of state transitions of the situated agent.** Each node corresponds to a discrete state in which the first three bits codify the cell assembly active in the agent and the six following bits codify the active synapses. The size of the nodes codify how frequently a state is visited and its colour the behaviour (preference for light A or preference for light B) in a span of tones from blue to red.

We can also observe what happens when we decouple these neurodynamic structures from behaviour. As in chapter 2, we define a passively-coupled agent as a copy of our agent that is fed with a recording of the input of the original agent instead of a real input generated by its own behaviour. By doing so, we are decoupling the internal mechanisms generating neurodynamical structures and the behaviour generating mechanisms. We can generate a network as in Figure 3.13, although in this case the values of  $L$  should be defined from the behaviour of the situated agent that actually generates the input (since passively-coupled agents do not move

in the environment), and the rest of the values from the passively-coupled agents. In Figure 3.14 we can observe that, although both agents receive the same input, the resulting network is quite different. Although clusters of nodes are generated, these clusters are not organized any more by the behaviour that generates the input that is fed into the agent. We observe that red and blue nodes are distributed among clusters in an apparently random distribution. Also, whereas in the previous network the vast majority of the nodes were clearly corresponding to red or blue, we find a greater number of nodes with intermediate colours, meaning that they are not completely ascribed to one particular behaviour. Note that the colour of the nodes does not arise from the real behaviour of the agent (which does not move), but from the behaviour of the agent that generates the input that we fed into the passively-coupled agent.



**Figure 3.14: Network of state transitions of the passively-coupled agent.** Each node corresponds to a discrete state in which the first three bits codify the cell assembly active in the agent and the six following bits codify the active synapses. The size of the nodes codify how frequently a state is visited and its colour the behaviour (preference for light A or preference for light B) in a span of tones from blue to red.

We can quantify the difference between the two networks dividing the network in two communities (the nodes that correspond mostly to behaviour of approaching light A and the nodes corresponding to light B) and analysing what percentage of links in the network connect nodes inside the same community and nodes between the two different communities. By doing so, we obtain the result that in the case of the situated agent 80.16% of the links connect nodes within the same community, and in the case of the passively-coupled agent only 61.01% of the link do so (note that 50% would be the case of a uniformly random distribution of links).

## INFORMATION FLOWS

The analysis above describes how the different scales of the system are intertwined and deployed in a coordinated manner, but it does not show what the causal relations are between the different elements. Here we quantitatively define the relation between behaviour, cell assemblies and synapsesemblies, applying information theory tools for characterizing nonlinear relations between elements of the system.

We previously defined the different cell and synapse assemblies as discrete units of behaviour. All possible relationships between oscillators give rise to 6 possible states for the cell assemblies, and all possible activated synapses give rise to 64 possible combinations or synapsesemblies. Therefore, we can codify the active cell assembly or the active synapseassembly with two discrete variables CA and SA ranging from 1 to 6 and 1 to 64 respectively. Similarly, for each trial we can codify the emergent behaviour of the agent. For each trial we define a variable L, which is equal to 1 if the agent's final distance to light A is less than 50 units, -1 if the final distance to light B is less than 50 units, or 0 otherwise.

Having defined these macro-variables of the system in a discrete manner, we can use information theory tools to determine the relation between variables. Information theory is built over a set of general measures for quantifying the uncertainty associated with individual variables and characterizing the nonlinear relationships between them (Cover and Thomas, 1991). In our model, we can use these measures to quantify relationships between the state of different elements in the neural controller of the agent, or between such components and features of the environment surrounding the agent.

Information is quantified in terms of *entropy*, i.e. the information contained in a random variable, which is generally defined as:

$$H(X) = - \sum_{x \in X} p(x) \log(p(x)) \quad (3.6)$$

where  $X$  is a set of states of a variable.

The information shared between two random variables  $X$  and  $Y$  is given by their *mutual information*:

$$I(X; Y) = H(Y) - H(Y|X) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (3.7)$$

By definition,  $I(X; Y) = I(Y; X)$ , thus no causal relationship can be detected through mutual information. Thus, to analyse causal relationships between variables typically transfer entropy measures are employed. The decrease of uncertainty in the state of a variable by knowing its uncertainty from the past history of other variables is defined as *transfer entropy* between two variables:

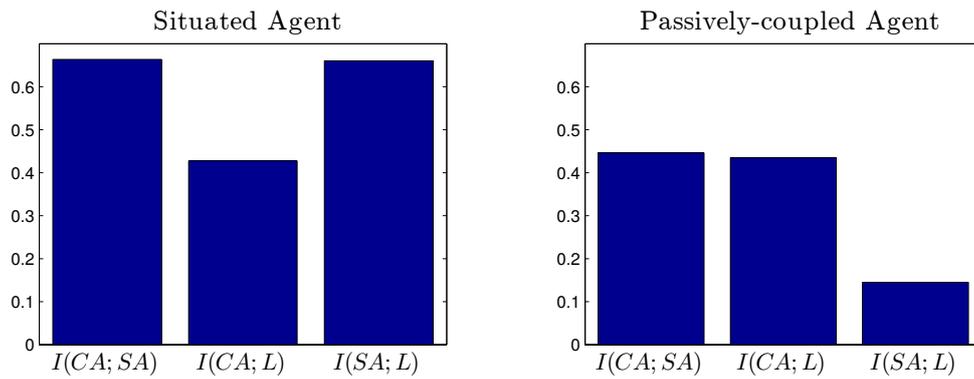
$$\begin{aligned}
\text{TE}(X \rightarrow Y) &= H(Y_{t+\tau} | Y_t^{(d')}) - H(Y_{t+\tau} | Y_t^{(d')}, X_t^{(d)}) = \\
&= \sum_{x_{t+\tau}, x_t \in X} \sum_{y_t \in Y} p(x_{t+\tau}, x_t^{(d)}, y_t^{(d)}) \log \frac{p(x_{t+\tau}, x_t^{(d)}, y_t^{(d)}) p(x_t^{(d)})}{p(x_{t+\tau}, x_t^{(d)}) p(x_t^{(d)}, y_t^{(d)})} \quad (3.8)
\end{aligned}$$

where  $X_t^{(d)}$  denotes the past history of  $x$  counted from time  $t$  and length  $d$  (i.e.  $x_t, x_{t-1}, \dots, x_{t-d}$ ).

With these methodological tools, we analyse the information flows for a situated and a passively-coupled agent simulated during 1000 trials of 125s. Although the analysis is just performed over one simulation for each agent.

### MUTUAL INFORMATION

We first analyse what information is shared by the emergent cell assemblies and synapse assemblies and the light chosen by the agent at the end of each trial. For doing so, we can measure mutual information along the time series of values of each variable. In Figure 3.15 we can observe how the three variables CA, SA and L share an important amount of information. The entropy of L (which is the variable with the lowest entropy) is 0.86, thus the shared entropy is, in most cases, in the same order of magnitude. In the case of the situated agent, we can observe in Figure 3.15 (left) that all variables share a relevant amount of information. However, in Figure 3.15 (right) we observe that the information shared between SA and L decreases dramatically, suggesting that the interaction between the two variables is lost.



**Figure 3.15: Mutual Information.** Values of mutual information between cell assemblies (CA), synapse assemblies (SA) and behaviour (L).

## TRANSFER ENTROPY

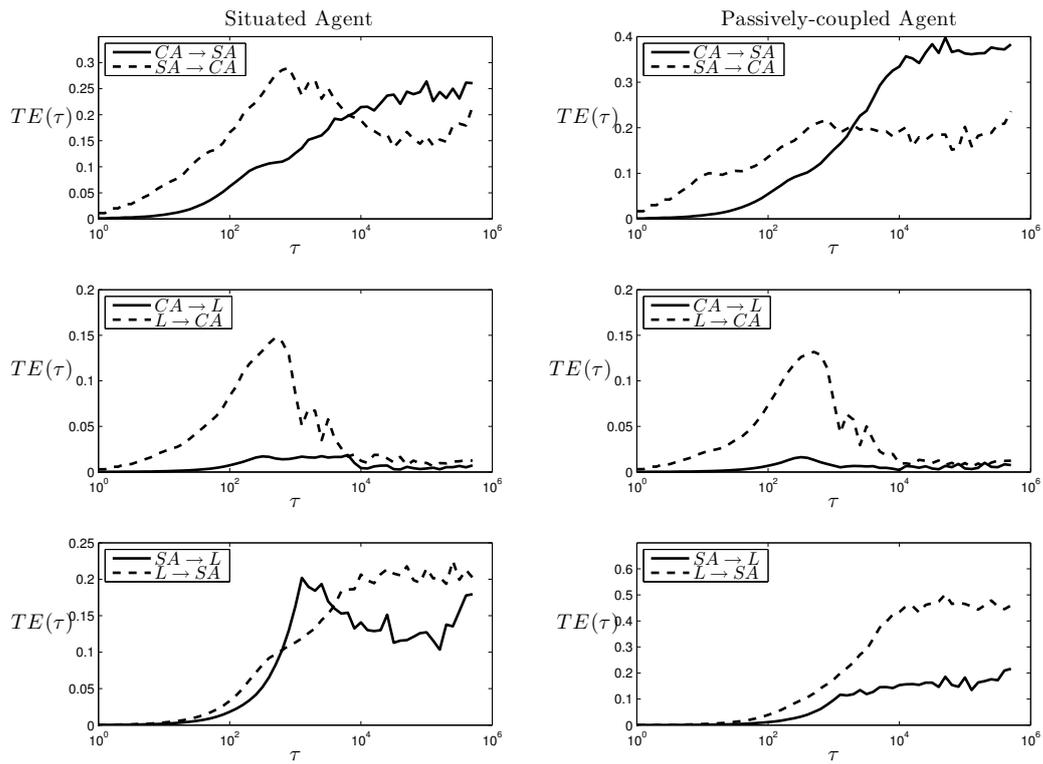
We can get a clearer picture by measuring transfer entropy between variables. We do so by using equation 3.8, with  $d = d' = 1$  as the length of the past history that we consider<sup>4</sup> and a logarithmically distributed series of values of  $\tau$  from 0.1s to 62500s (half the length of the 1000 trials), with multiplicative intervals of  $10^0.1$ . In Figure 3.16 (left) we can observe a complex chart of information flows for the situated agent. For example, we can observe the interaction between variables CA and SA where at small and medium values of  $\tau$  the information flows mainly from CA to SA, whereas at longer values of  $\tau$  the flow of the information is reversed. This suggests a circular causal chain in which, at short timescales, the structure of the current synapsembles determines the cell assemblies that can emerge, but at long timescales it is the self-sustainment of particular assemblies during different trials that determines the stability of the possible synapsembles (note that the duration of a trial corresponds to a value of  $\tau$  of 125s). Also, we observe that while there is an important information flow from L to CA, the flow does not exist in the opposite direction (Figure 3.16 middle-left), suggesting that the light chosen by the agent influences the cell assembly that emerges, but that the assembly that is active at a particular moment of time is not decisive for the behaviour that the agent will choose. In turn, there is an important bidirectional exchange of information between SA and L. This suggests that SA is the variable that determines the behaviour that will be chosen by the agent. Also, we can observe that  $TE_{L \rightarrow SA}$  and  $TE_{CA \rightarrow SA}$  are very similar in value and shape (if we integrate the area of the difference between  $TE_{L \rightarrow SA}$  and  $TE_{CA \rightarrow SA}$  and divide it by  $TE_{CA \rightarrow SA}$  the result is 0.11, showing that both functions coincide with almost 90% of accuracy). This is supported by the fact that there is no functional dependency from L to SA, since change in the weights  $\delta K$  is only a function of K and  $\theta$ . As we can easily check, all the other information flows present in Figure 3.16 correspond to actual functional dependencies depicted by the equations defining the systems<sup>5</sup>. This suggests (since L influences CA and not otherwise), that L causally determines CA which in turn influences SA in their circular mutual interaction<sup>6</sup>.

Putting together the information flows above, we may summarize them in the schema at Figure 3.17 (left). The behaviour of the agent L generates an input that determines the emer-

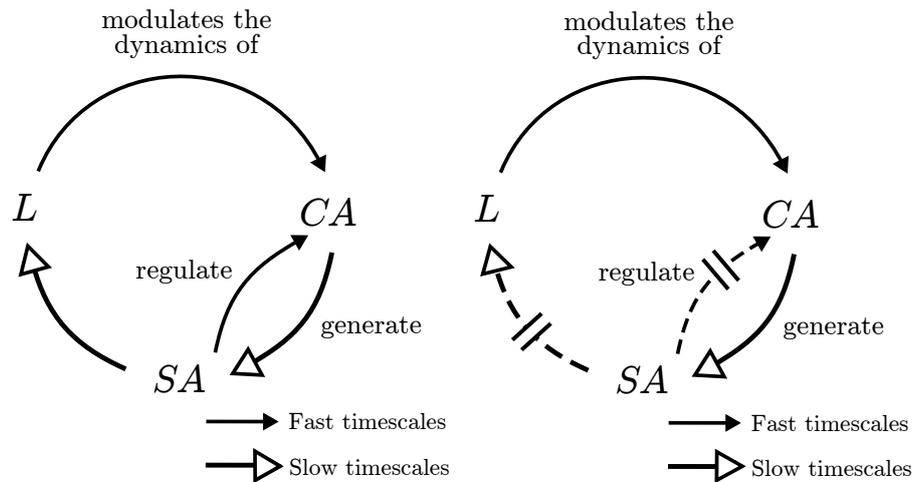
<sup>4</sup>we tested different values of  $d$  and  $d'$  up to 5 for a series of representative values of  $\tau$  with similar results, therefore we used  $d = d' = 1$  to reduce the computational cost, since just varying the value of  $\tau$  seems to be enough for capture the different timescales of information flows.

<sup>5</sup>Since  $\dot{\theta}$  is a function of  $\theta$ , K and  $s_{DX}$ , and  $\dot{s}_{DX}$  is a function of  $\theta$ , K and  $s_{DX}$ . The only functional dependency that presents no informational content is  $TE_{CA \rightarrow L}$ .

<sup>6</sup>It is worthy to note that in the experiments depicted in Figure 3.16 the variable L is defined as the macroscopic behaviour of the agent, whereas CA and SA are patterns of low level variables of the agent. To test that this difference did not distorted in any way the result of our analysis, we repeated the analysis substituting L by a variable S defined as a string of 4 bits (one for each sensor of the robot), in which each bit was equal to one if the activity of its corresponding sensor is higher than its average activation. The results of transfer entropy measures were strikingly similar to those displayed in Figure 3.16, indicating that the series of sensory activation are strongly correlated with particular behavioural macroscopic patterns of the agent. Since the use of L simplifies other analyses performed in this chapter, we have chosen to use this variable instead of S.



**Figure 3.16: Transfer Entropy.** Values of transfer entropy at different timescales among cell assemblies (CA), synapses (SA) and the behaviour (L).



**Figure 3.17: Operational closure.** Simplified information flows for the situated (left) and passively-coupled (right) agents.

gence of cell assemblies *CA*. A circular relation between the emergent cell assemblies and their underlying synapse ensembles *SA* generates a particular behaviour *L* which is determined by the state of *SA*. This schema is surprisingly similar to the one proposed by Varela (1997) and depicted in Figure 3.1

Moreover, if we ‘cut’ the sensorimotor loop by feeding the agent a recorded input, the information flows from *SA* to *L* and from *SA* to *CA* are disrupted, whereas information flows from *CA* to *SA* (which are almost equivalent to flows from *L* to *SA*, supporting our suspicion that they represent the same thing), thus dissolving the self-sustained neurodynamic structures that generate a coordinated behaviour and reinforcing unidirectional influence from *CA* to *SA*, as we depict at Figure at Figure 3.17 (right). Thus, these results strongly suggest that the generation of complex and integrated neurodynamic structures is a product of a double loop that strongly couples (1) neural mechanisms and higher order parameters (synaptic ensembles in this case) and (2) these neurodynamic circular structures with the behaviour they generate.

#### TEMPORAL INFORMATION FLOWS

The informational analysis above only shows a static picture of information flows on average, but does not tell us much about how these flows unfold on time (even when it gives information about the different timescales of the system). To overcome this limitation, Beer and Williams (2015) have proposed a framework combining information flow and dynamical analyses, exploring how a simulated model agent in a relational categorization task integrates information at different moments of time about a cue used for solving the task. Instead of analysing information as an average of the dependences between variables along a time series, they run the same task several times for different initial conditions and compute information measures for each time instant. Instead of using a series of temporal values of a variable, they

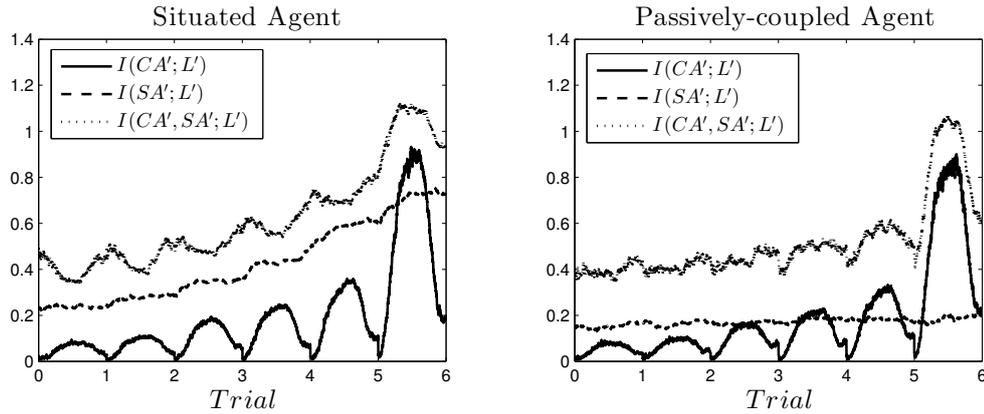
use a series of values of a variable on each instant along different starting conditions. Similarly, we can fold our time series into 1000 time series in which e.g.  $CA'(i, t) = CA(t + (i - 1)125)$ , where 125 is the duration of a trial. Now, for each value of  $t$  we have series of 1000 values for  $CA$ ,  $SA$  and  $L$ .

For six trials (i.e.  $t = 0, \dots, 6 \cdot 125s$ ) we compute the mutual information between the neurodynamic variables of the agent  $CA'$  and  $SA'$  and the light chosen at the end of the sixth trial  $L'$ . Also, we can compute the joint mutual information that  $CA'$  and  $SA'$  share with  $L'$ . In Figure 3.18 (left) we can observe the result for the situated agent. We observe how the mutual information  $I(CA'; L')$  increases during the middle of the trial, fading out at the beginning and the end. Also, from one trial to the next the information at the peak increases, being maximum at the sixth trial. We can interpret this as a self-sustaining behaviour of cell assemblies: when a cell assembly emerges, it reinforces itself and has more probabilities to emerge again the next trial. Similarly,  $I(SA'; L')$  steadily increases until a cell assembly is activated at the middle of the sixth trial. However, mutual information between  $SA'$  and  $L'$  is steadily accumulated and does not decrease, thus we can hypothesize that the configuration of the synapse ensembles 'stores' information about the behaviour that the agent will develop. Furthermore, when we analyse the joint contribution  $I(CA', SA'; L')$ , we observe that it is always higher than the individual contributions. Moreover, it increases when  $I(CA'; L')$  decreases at the transitions between one trial and another. Also,  $I(CA', SA'; L')$  decreases when at the middle of the trial a cell assembly is activated, except in the last trial. This portrays an interesting picture, where information is flowing back and forth between the emergent cell assemblies and the collective cell-synapse assemblies, until at the sixth trial an assembly emerges producing behaviour  $L'$ .

If we analyse the passively-coupled agent we observe a quite different picture. Although  $I(SA'; L')$  is quite similar in both situations (its values are slightly smaller in the passively-coupled condition),  $I(SA'; L')$  does not integrate any information. This suggests that even when the input produced by behaviour  $L'$  is able to influence the cell assemblies that emerge, coordination between behaviour and the stabilization of synapse assemblies does not take place. Furthermore, the joint information  $I(CA', SA'; L')$  does not integrate much information either, and the anticorrelation between  $I(CA'; L')$  and  $I(CA', SA'; L')$  disappears. This suggests that the passively-coupled agent loses the struggle between information flows through individual and collective variables, indicating that the important moments for generating the behaviour of the agent are not only synchronizing moments of emergence of cell assemblies, but most information is built mostly at desynchronizing instants corresponding to transition from one assembly to the next.

## RECAPITULATION

The different analysis present a picture in which for the situated agent a closed network of interaction at different scales is circularly organized. On one hand, sensory activity and synaptic configurations can trigger the emergence of cell assemblies at fast scales, provoking different conducts in the agent. On the other, synaptic weights integrate information about the oscil-



**Figure 3.18: Information flows.** Mutual information at different instants between cell assemblies ( $CA'$ ), synapses ( $SA'$ ) and behaviour ( $L'$ )

latory patterns that are being homeostatically maintained, allowing to emergent behavioural patterns that build a repertoire of habits for an agent. If the sensorimotor loop is disrupted (e.g. when the agent is passively-coupled, and probably for more severe sensorimotor disruptions), this circular closure disappears and autonomous organization of the agent vanishes. In our model, the constitution of an agent as such depends on different neural, synaptic and behavioural processes intertwined into a circular organization, from where asymmetries necessary for differentiating between agent and world emerge from the recurrent interaction of the different levels of activity.

### 3.7 DISCUSSION

In this chapter we have proposed a characterization of neurocognitive autonomy based on (1) the integration of transient assemblies underlying behaviour through nonlinear coupling of neural clusters generating specific conducts in the agent, and (2) the coordination between sensorimotor and neurodynamic structures into a self-maintaining web of behavioural patterns (habits). The integration of these two level of activity gives rise to a sensorimotor dynamic core which allows the emergence of moments of coherence sustaining an agent's identity and the possibility of flexibly engaging and disengaging into different available sensorimotor loops.

In order to explore this hypothesis, we have presented a minimal model of autonomous behavioural preference, in which a robot is able to create self-sustaining and transient phototactic patterns when presented with an environment with two different lights. The robot's controller is based on a network of three Kuramoto oscillators with a plastic homeostatic mechanisms designed to maintain constant some relations among the oscillators by triggering changes in synaptic connectivity. The result is that behavioural preferences emerge from the interaction between oscillator cell ensembles, ensembles of synaptic weights or 'synapses', and the robot's behaviour.

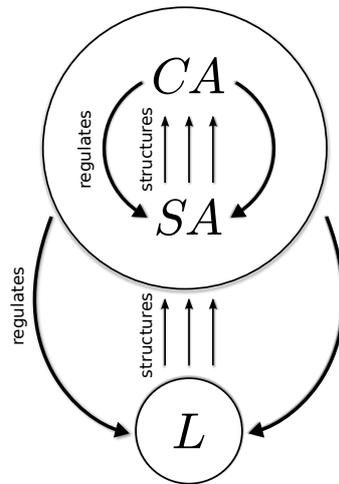
The picture drawn by our agent's behaviour, although oversimplified, provides us with in-

interesting notions for characterizing neurocognitive autonomy. In a world with two types of stimulus, the coupling between the neural oscillators controlling the agent is able to generate a basic form of coherence when a cell assembly emerges maintaining a specific sensorimotor invariance (e.g. approaching light A). This self-maintained cell assembly however, quickly dissolves when the agent reaches the light or when another behaviour emerges. This loss of stability provokes plastic changes modifying the organization of the agent. Nevertheless, some cell assemblies generate behaviours that reinforce the synaptic structures sustaining them, creating a circular interaction in which a cell-synapse neurodynamic structure is self-sustained together with the behaviour generating it. This is what we have defined above as a habit or a self-sustaining pattern of behaviour coupling neurodynamical and sensorimotor structures. At this longer timescale, a coordination takes place among microscopic neural activity, mesoscopic configurations of synaptic organization, and the macroscopic level of behaviour. On an even longer timescale, sequences of self-sustained patterns of behaviour create self-reinforcing paths through a web of habits or behavioural preferences, as depicted in Figure 3.13, characterizing the long-term organization of the agent.

Our model shows how different scales of a self-organizing neural system (neural ensembles, an underlying layer of mechanisms of relational homeostatic plasticity and the agent's behaviour) generates a web of habits regulated in a distributed manner. Our model presents a concrete implementation of different schemas of autonomous behaviour that have so far been mostly based on theoretical elaborations. Through our conceptual robotic model, the experiments carried out with it and the information theoretic analytic methods applied, we are now in a position to provide a more detailed operationally explicit and empirically testable formulation of neurocognitive autonomy. In our definition of autonomy, we defined three necessary properties for autonomous agency: operational closure, interactional asymmetry and precariousness. Our model allows us to examine each one of the properties.

**OPERATIONAL CLOSURE.** As we can observe in Figure 3.17 the information flows we analysed in our agent create a closed network between the agent's internal processes and sensorimotor interactive processes. This network takes the form of a double coupling loop of self-sustained assemblies and its plastic generative structure on the one hand and behavioural sensorimotor patterns and the neural activation structure that generates them on the other. The interaction between the two takes the form of a double feedback loop of (1) a circular causal dependence between the emergence of cell ensembles and synapsembles, and (2) a causal chain in which synapsembles influence behaviour and behaviour influence the emergence of cell ensembles as in Figure 3.17 (left). This double feedback loop is broken when we analyse an agent in a passively-coupled condition, being fed with an input recorded from a regular agent, suggesting that the existence of a sensorimotor loop is essential from the emergence of coherent behaviour.

**INTERACTIVE ASYMMETRY.** Moreover, as we represent in Figure 3.19, the operational closure of sensorimotor interactions creates a clear asymmetry between agent and environment, in which the directionality of interactions is reversed at different scales (Figure 3.16). This al-



**Figure 3.19: Autonomous agency.** Characterization of the information asymmetry in the interaction between an autonomous agent and its environment.

allows us to identify the agent as a unit which is affected by bottom-up causal flows from the variability of in the sensorimotor loop, and at the same time is able to develop a downward causation over its sensorimotor interaction. This multiscale nature of the interactive asymmetry between agent and environment potentially adds a novel element in our definition of autonomous agent. Our model suggest that the interactional asymmetry that defines an agent against this world is one in which the agent is structured by the microscopic regularities of the environment, and in turn is able to exert a macroscopic regulation over them. The nature of this asymmetry will be further explored in the next chapter.

**PRECARIOUSNESS.** Also, the operational closure of the system provides a level of ‘sense-making’ into the activity of the agent, since now actions have a meaning proper to the perspective of the self-sustaining coherence of the agent: e.g. a particular event may help to maintain a behavioural preference or to destroy it. This meaning is not rooted in any genetic encoding or the design of the agent but in the history of interactions that generates or perturbs a particular self-sustaining web of neurodynamic and sensorimotor structures. Thus, closed regulatory loops modulate the maintenance and dissolution of neurodynamic structures necessary to imprint autonomy into neurodynamic self-regulation. The possibility for autonomy is created precisely in these breakdowns of the stability of the agent’s behaviour. In Varela’s words, ‘[t]he key to autonomy is that a living system finds its way into the next moment by acting appropriately out of its own resources. And it is the breakdowns, the hinges that articulate microworlds, that are the source of the autonomous and creative side of living cognition’ (Varela, 1992). As we observed in Figure 3.18, it is the breakdowns from one trial to the next where information about the agent’s behaviour is generated. This is when, using Varela’s ter-

minology, microidentities and their corresponding microworlds are generated – the A-light seeking robotic agent, or the cyclist that enjoys a sunny ride through the park in our example at the beginning of this chapter – in the form of an operational closure that sustains the stable, specific forms of activation of the web of habits that constitutes the macroidentity of the agent – the autonomous preference switching agent, or the cyclist that shapes her way to work in an autonomous manner.

Here, we have examined the behaviour of the macroscopic variables of behaviour of our agent, characterizing the transition between states and the information loops between collective patterns of groups of variables. Still, we may ask ourselves how these variables interact and regulate themselves at different scales of description. We know that our model adaptively maintains the state of the phase relations between oscillators (variables  $\phi_i$  which trigger plastic changes when they are not kept within a viability space), but, what is the relation between low-level relational homeostasis and macroscopic operational closure? What kind of invariances in the dynamical organization of the agent are preserved? Are the relational homeostatic variables in our agent a proxy for more fundamental aspects of the organization of the system or the preservation of a dynamic core? In the next chapter we will investigate some of these aspects taking a look at some indicators of the organization of the system at different temporal and spatial scales.

# 4

## Situated Autonomy: complexity and self-organized criticality

**ABSTRACT:** Using the model in previous chapter, we characterize the interactions between different spatial and temporal scales of autonomous organization, using different tools from complex systems science. We address the role of plastic mechanisms and sensorimotor coupling in the emergence of autonomous organization.

**CONTRIBUTIONS:**

- We characterize the role of different variables in the generation of information flows in the system using TSE complexity measures under different conditions of the agent.
- We assess the emergence of long-range correlations between different temporal scales by studying the presence of  $1/f$  patterns of the agent under different conditions.
- Based on the results of our experimental analysis, we are able to describe how different notions of our definitions of autonomy are met in the mechanisms of the system: interactive asymmetry and precariousness in the system operational closure.

**RELATED PUBLICATIONS:**

- Aguilera, M., Barandiaran, X. E., Bedia, M. G., and Seron, F. (2015). Self-Organized Criticality, Plasticity and Sensorimotor Coupling. Explorations with a Neurorobotic Model in a Behavioural Preference Task. *PLoS ONE*, 10(2):e0117465.

In the previous chapter we have described and analysed in detail the behaviour of a model of an autonomous robotic agent. We have observed how a neural controller embedded into an agent through a double regulatory loop combining sensorimotor interaction and systemic homeostatic regulation is able to display autonomous behaviour as it generates the invariances and equilibriums that govern its behaviour. In the previous chapter we described autonomous organization in terms of the macro-variables that drive the behaviour of the agent (phase relationship patterns and connectivity patterns), but we do not know much about how

these patterns emerge from the individual variables of the agent's controller and what kind of interaction takes place between micro, meso and macro levels. According to our definition of autonomy, a simple double feedback loop is not sufficient for autonomous behaviour. Instead, these loops play a quite particular role in an asymmetric organizational structure, they are able to display high levels of differentiated activity at the micro level in interaction with external inputs, in order to cope with the complexity of the environment, at the same time that they generate emerging integrated patterns at the macro level, in order to effectively act over the world surrounding the agent. In this chapter we focus on the kind of relations between different temporal and spatial scales that allow the emergence of the double regulatory loop required for cognitive autonomy. If in the previous chapter we characterized the circular relations between macroscopic variables of the system, treating the mechanisms generating them as a black box, here we analyse the inner structure behind the emergent properties of the system. First, we analyse the low level patterns arising in our system and the presence of metastability in the switching of those patterns, finding that more complex methods are required for unveiling the temporal and spatial structures underlying the metastable patterns displayed by the system. We use TSE complexity to understand the spatial distribution of information flows among the different variables of the system, and  $1/f$  fractal scaling of different variables to characterize the distribution and coordination of different temporal scales. This allows us to characterize with more detail some properties of autonomy we described in the previous chapter, such as interactive asymmetry and precarious operational closure.

#### 4.1 EXPERIMENTAL SETUP

In the previous chapter we have simulated our agent in a quite controlled environment, in which the position of the agent and the lights were reset after a given time. This was useful for delimiting the agent's behaviour into different trials and analysing how behavioural preferences evolve. However, this is far from the kind of environment that is natural for an autonomous agent, since it does not allow our agent to explore large scale regularities of its world that may emerge from a less constrained behaviour. For the following simulations, we create an environment with two types of lights as before. This time, whenever the agent is close to one of the lights (closer than 25 units), the light starts being consumed. Each light starts with an energy of 1, and whenever the agent is near the light it consumes at a linear rate of 0.2 units per second. When a light is consumed, it disappears and a new light of the same type appears at a random position. When a new light appears, it starts with a value of energy of 0, which grows progressively during its 1 first seconds of life at a linear rate of 0.2 unit per second. The intensity perceived by the agent is now multiplied by the amount of energy of the light. Thus, we can model the evolution and renewal of the environment without needing to reset the agent and the lights' positions. We now have a more natural environment without abrupt transitions where the agent can explore its environments without further restrictions than the lights fading away when consumed by the agent. This allows us to explore larger timescales of the behaviour of the agent (especially relevant for the analysis of the fractal

structure of the agent's behaviour). Agents are simulated for a period of  $1.25 \cdot 10^5$  seconds, in which agents are able to consume several hundreds of lights.

We compare different types of agents to address the effect of neural plasticity and embodiment in the generation of the agent's behaviour. However, in the first part of the experiment we want to analyse the contribution of embodied interaction in a more precise way. As we saw in Chapter 2, we can study the contribution of embodiment without removing the structure of the input by analysing a partially decoupled agent.

We study four types of agents:

1. A normally functioning agent with both synaptic plasticity and a normal sensorimotor interaction (S-P).
2. A passively-coupled agent, that is fed an input generated by a situated agent, but is not able to influence its world (PC-P).
3. An agent without synaptic plasticity, where the values of the weights are obtained randomly from a simulation of a situated agent<sup>1</sup> (S-NP).
4. A passively-coupled agent without synaptic plasticity, which is fed an input signal generated by a situated agent also without synaptic plasticity (PC-NP).

#### 4.2 METASTABLE ORGANIZATION

In Chapter 3 we discussed in detail the idea of cell assemblies as an organizational model supporting neural autonomy. Originally, Donald (Hebb, 1952, p.xix) developed the notion as a 'diffuse structure comprising cells [...] capable of acting briefly as a closed system, delivering facilitation to other such systems and usually having a specific motor facilitation'. This view is still alive in modern neuroscience in different ways. For example, it has been hypothesized that synchrony of neural oscillations is relevant for the encoding and maintenance of rules (Engel, 2012). The role of neural assemblies seems to be reduced to the mediation between a learned stimulus and a response. Although it is perfectly plausible that neural assemblies generate the possibility of rule-based behaviour, we think it is unlikely that their role is reduced to simple input-output relations in the form of logical rules.

The notion of metastability allows us to conceive of a different picture, where the emergence of synchronized patterns is not the result of logical rules but the action of a network of loose interaction between dynamic coalitions of neural areas. Karl Friston (2000b) has stated that the ability of a system to create *dynamical instabilities* (i.e. metastability) is key for the generation of the diversity necessary for adaptive behaviour in the brain. A metastable regime is characterized by transients periods of stability, thus, the frequency composition of the time

---

<sup>1</sup>60 combinations of weight values were extracted from the value of the weights at the midpoint of the trials of the simulation represented in Figure 3.5

series of the system must change with time. As a consequence, Friston proposes a simple measure of metastability in terms of the instability or entropy of the spectral density of a signal. If we characterize the spectral density of a signal at different moments of time (for example through a wavelet transform) as  $g(\omega, t)$ , a measure of the level of metastability of the system may be obtained as:

$$MS = H(g(\omega, t)) \quad (4.1)$$

where  $H()$  computes the Shannon entropy of a signal. However, if many frequencies are taken into consideration, it is costly to compute the density function of the resulting multidimensional signal. Therefore, Friston proposes computing it under Gaussian assumptions in the form  $H = \log((2\pi e)^m \det\{\text{Cov}\{g(\omega, t)\}\})/2$ .

For Kuramoto oscillators we can use another measure of metastability based on the order parameter of the oscillator. Typically, the order parameter in a Kuramoto network is computed as:

$$r = \left| \frac{1}{N} \sum_{j=1}^N e^{i\theta_j} \right| \quad (4.2)$$

Murray Shanahan (2010) proposes that the level of metastability can be defined by the coalition entropy, in which coalitions are defined as binary variables that arise when  $r_c(t)$  is bigger than a threshold, where  $c$  is defined as a subpopulation out of all the possible subdivisions  $C$  of the network in which the order parameter is computed. A coalition  $s$  is said to arise at time  $t$  if  $\theta_c > \gamma$  if for all  $c \in s$ , where  $\gamma$  is a synchronization threshold. We define the coalition entropy as the joint entropy of the order parameter of the different populations as:

$$H_C = -\frac{1}{\log_2|S|} \sum_{s \in S} p(s) \log_2(p(s)) \quad (4.3)$$

where  $S$  is the set of distinct coalitions the system can generate and  $p(s)$  the probability of coalition  $s$  arising in any given time point.

The diversity in the possible coalitions that emerge out of the Kuramoto network is an indicator of the level of metastability in the system. We use this measure in this chapter as an indicator of the level of metastability in the Kuramoto controller of a robotic agent.

#### 4.2.1 EMERGENCE OF METASTABLE PATTERNS IN AUTONOMOUS BEHAVIOUR

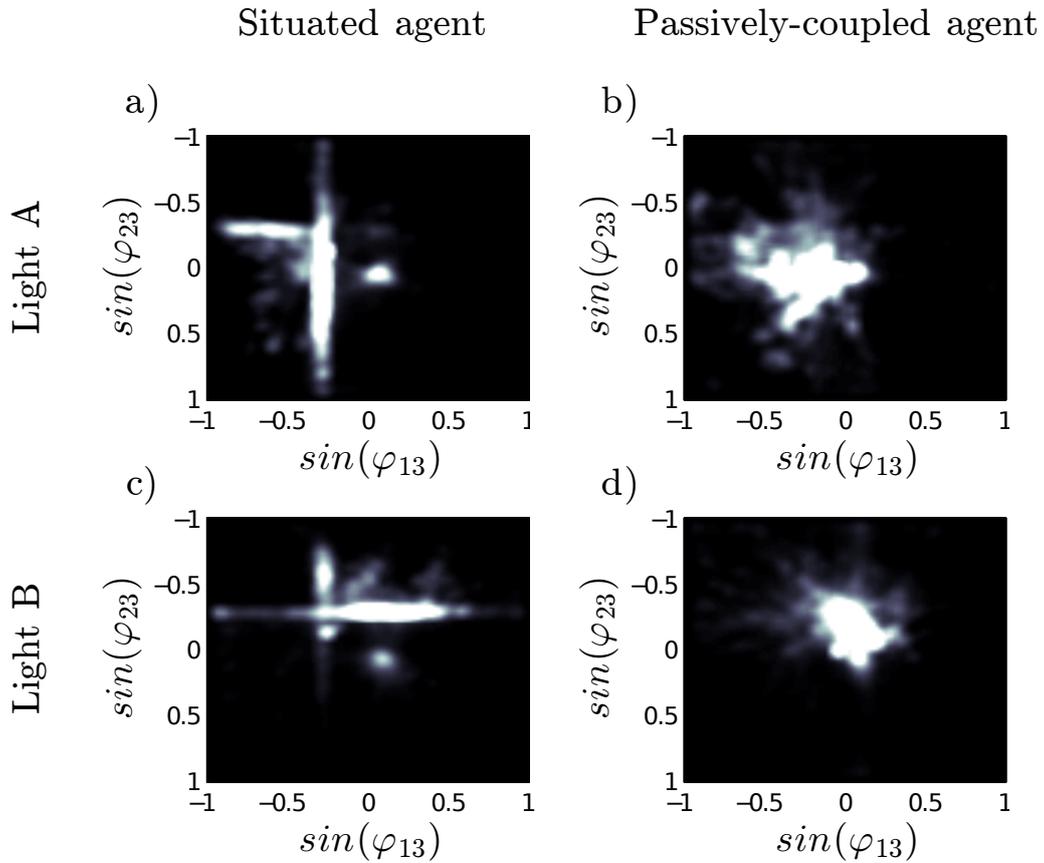
In Chapter 3 we analysed how certain patterns were associated with the behaviour of the agent, and we characterized a circular regulatory relation between behaviour and high-order patterns in a double feedback loop. We also observed how this circular regulatory organization disappeared when sensorimotor loops were disrupted in a passively-coupled agent. However, in the previous chapter we explored high-order patterns of joint phase-relationships and weight activations, which do not explain much about how these patterns are created from low-level mechanisms and how these mechanisms are regulated.

What patterns does our agent display and how are they organized? Are there metastable mechanisms regulating the emergence and dissolution of these patterns? What kind of measures can unveil the underlying structure of the system? Since the main behaviour of the agent is to develop behavioural preferences of approaching either of the two lights, we study whether these preferences correspond to a particular neurodynamic pattern. For doing so, we identify the moments in which the agent is approaching one light and not the other by computing the relative velocity towards one light with respect to the other. We compute this relative velocity as  $v_{AB} = \frac{d(d_A - d_B)}{dt}$ , where  $d_A$  and  $d_B$  are respectively the distances to light A and B. Then we extract the moments when  $v_{AB}$  is positive or negative in windows of 50 seconds. We consider that the agent is approaching light A when  $v_{AB}$  is positive for at least 75% of the 50 seconds window interval, and that the agent is going to light B when the  $v_{AB}$  is negative for 75% of the interval. With this index we find that 21.0% of the time the agent is going only to light A and 24.3% of the time the agent is approaching only light B (note that there may be a percentage of the time when the agent is approaching both lights at the same time and it is not considered here).

We analyse what kind of patterns arise when the agents are going to either of the two lights for a trial of duration  $1.25 \cdot 10^7$  seconds. For doing so, we compute the joint density distributions of variables  $\varphi_{13}$  and  $\varphi_{23}$  for both the situated and passively-coupled situated agents. We do it by computing an averaged shifted histogram (Scott, 2012) of the joint distribution with 100 bins and 8 shifts. In the case of the situated agent, we find that well defined patterns are formed when the agent approaches one of the lights (Figure 4.1.a). When the agent approaches light A,  $\varphi_{13}$  is stabilized and  $\varphi_{23}$  moves freely, and the opposite happens when the agent goes to light B. This does not happen for the passively-coupled agent, since the patterns arising are much more ‘blurred’ (Figure 4.1.b). This suggest that situated agents are able to coordinate precise neurodynamic patterns when they enter in different modes of engagement, thus generating finely coordinated complex patterns that display a high degree of metastability and complexity. On the other hand, passively-coupled agents are able to stabilize sensorimotor patterns and thus not generate moments of transient stability.

Agents without synaptic plasticity however do not create such dramatic distinctions between dynamical signatures of situated and passively-coupled agents. Although signatures are not identical, they present similar shape and attractor locations (Figure 4.2), suggesting that agents with plasticity actively regulate the emergent sensorimotor patterns through modification of their synaptic weights. In turn, for agents without plasticity, it is the structure of the received input that determines the patterns being stabilized, and not the metastable interactions between the agent’s internal and sensorimotor dynamics.

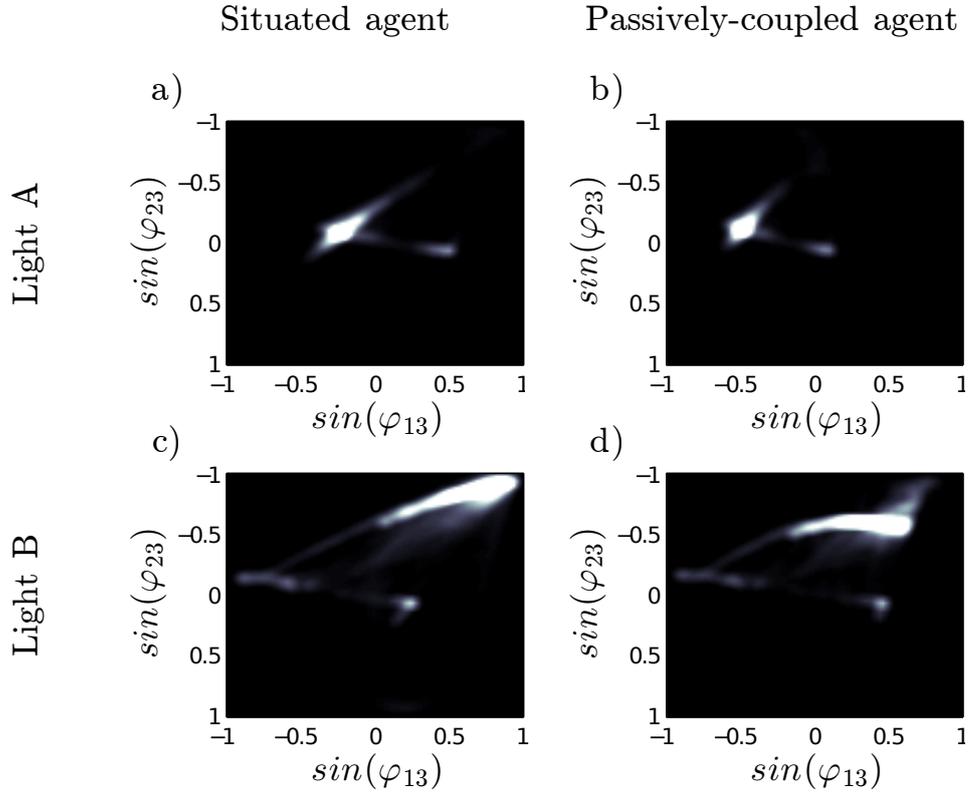
To test this hypothesis, we analyse the degree of dynamical instabilities (i.e. metastability) in the system using the coalition entropy measure defined in Equation 4.3, using a threshold of  $\gamma = 0.95$ . This measure computes the joint entropy of the order parameter of the possible coalitions that may arise in the system. We compute the values of metastability for 50 trials of duration  $1.25 \cdot 10^7$  seconds for each type of agent. As displayed in Figure 4.3, all agents present



**Figure 4.1: Dynamical signatures of agents with synaptic plasticity.** Averaged shifted histogram of the phase relationships between oscillators for each type of behaviour for situated and passively-coupled agents with synaptic plasticity.

in some cases significant values of metastability (note that the maximum value of entropy for four binary variables is 4). We observe that, for agents with normal synaptic plasticity, the situated agent always displays a higher value of metastability, showing more variety in the combination of coalitions that may arise in the system.

This is interesting since the situated agents seem to present both the ability to stabilize coherent patterns of phase-locking for particular behaviours, together with a more metastable state, therefore having a higher ability to dissolve these patterns and switch between them. For the case of the agents without synaptic plasticity, we find a broad range of metastability values. Being the weights selected randomly from a variety of cases, it seems reasonable that some configurations of the agent present a highly stable behaviour and others a metastable one (as we saw in the previous chapter, some states of the agent were more open to the world and others were ‘blind’ to the influence of certain environmental stimulus for the commitment or choice towards one of the light sources). Although situated agents without plasticity present a higher mean value of metastability, if we compare the distributions between situated

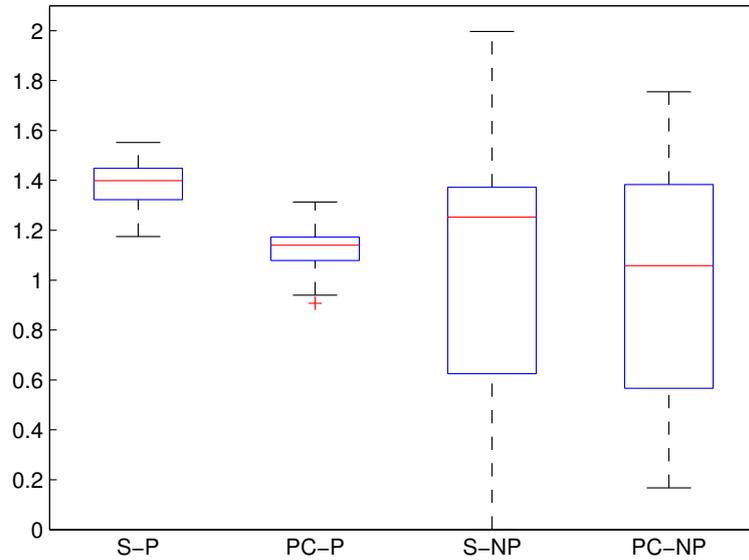


**Figure 4.2: Dynamical signatures of agents without synaptic plasticity.** Averaged shifted histogram of the phase relationships between oscillators for each type of behaviour for situated and passively-coupled agents without synaptic plasticity.

and passively-coupled agents with and without plasticity using ANOVA, we find that agents with plasticity present statistically significant different distributions [ $F(1, 98) = 233.13$ ,  $p = 1.17 \cdot 10^{-27}$ ], while agents without plasticity do not present significantly different distributions [ $F(1, 98) = 0.22$ ,  $p = 0.64$ ]. Thus, situated agents only increase their levels of metastability when they have synaptic plasticity.

Finally, we observe that the agent with synaptic plasticity presents values of metastability that are in the upper part of the wide range of the values presented by the agents without synaptic plasticity. This suggests that among all possible combinations of weights, plastic mechanisms operate in a way that maintain the agent with high levels of metastability.

From these results, we find that (1) when agents are situated, they present stable differentiated patterns and a more diverse repertoire of patterns (i.e. it is highly metastable) at the same time, and (2) plastic mechanisms modulate the dynamics of the agents to maintain a high level of metastability among a very wide range of possible embodied behaviours, indicating that plasticity might be playing a meta-regulatory role for the agent (i.e. not only regulatory behaviour but higher-order equilibriums in the system as the levels of metastability).



**Figure 4.3: Metastability.** Distributions of the indices of metastability for distribution for 50 agents of each type: situated (S) and passively-coupled (PC) agents, with synaptic plasticity (P) and without synaptic plasticity (NP).

These findings require a finer grained analysis of how the agent’s activity is distributed and coordinated in time and space, and what the role of the different mechanisms is. In the next sections we make use of more sophisticated tools to gain a better understanding of the agent’s behaviour.

#### 4.3 COMPLEX SYSTEMS TOOLS FOR CHARACTERIZING AUTONOMOUS ORGANIZATION

Characterizing autonomous organization is not an easy task. We have depicted autonomous organization of the nervous system in which coalitions of neural units synchronize transiently, creating the underlying dynamical structure that allows the emergence of specific cognitive acts. This description brings some requirements for neural organization, for example, it is required to display metastable modes of coordination in which the system is never permanently trapped into an attractor but always switches among a set of quasi-attractors. Moreover, neuronal systems must present a virtuous combination of integration (the system presents a coherence as a whole) and differentiation (the system displays a vast set of possible behavioural patterns). Inspired by recent work in autonomous robotics and notions from radically embodied neurodynamics we posed the question about whether different modes of neural organization are restricted to the brain or also cut across brain-body-environment boundaries. Doing so demands the use of tools for characterizing these modes of complex organization.

In a recent manuscript, Tognoli and Kelso analyse how information flows in the brain must be reconciled with synchronization and mass action in the brain (Tognoli and Kelso, 2013). They suggest that *metastability* may be a dynamical regime where both concepts can coexist, displaying an interplay of integrative and segregative tendencies that are expressed in space and time during the normal course of brain and behavioural function. In particular, the ability of the brain to create information flows must be reconciled with its propensity for synchronization and mass action. Typically, approaches to complex systems are divided into those that derive from information theory and those that come from deterministic nonlinear dynamical systems (Friston, 2000b). In this chapter, we use a representative example of each one of them to analyse the correlations between different temporal and spatial scales.

As we reviewed during chapter 1, tools and concepts of complex systems research (often forged in the realm of physics) are increasingly permeating cognitive science and the way we understand the mind and ourselves. Since early cybernetics (Ashby, 1954) and system's theory (Bertalanffy, 1969, 1950), what might currently be called the complex systems approach to cognition, has put the emphasis on self-organization, the irreducible non-linearity of cognitive processes and circular causality (Kugler et al., 1980; Kelso, 1995; Thelen, 1996). This approach contrasts with some classical assumptions behind the computational theory of the mind: modularity, functional decomposition, perception-action dichotomy or agent-environment 'decoupling' (sharp analytic separation between the dynamics of the agent and the environment). The complex systems approach to cognitive science seeks to move beyond the formalism of computational theory and information exchange and to place cognition within the broader formalism of dynamical systems (Van Orden et al., 2003; Dixon et al., 2012; Tognoli and Kelso, 2013).

Paradigmatic examples of the conceptual tools of the complex systems approach to neuroscience and cognitive science, with applications in several domains (in particular in cognitive science and neuroscience), are the concepts of TSE complexity (Tononi et al., 1994; Tononi and Edelman, 1998), and self-organized criticality (SOC hereafter) and  $1/f$  noise (also known as scale-free or pink noise) (Bak et al., 1987; Jensen, 1998). We believe that these notions are useful for characterizing our definition of autonomy, specifying the relationships among scales in the spatial and temporal domains. The goal of this chapter is to analyse the behaviour of the robotic agent presented in the previous chapter, and to understand the structure that sustains the circular closed loop that we depicted to characterize autonomous organization.

#### 4.3.1 GENERATIVE MECHANISMS OF AUTONOMOUS BEHAVIOUR

Sometimes, research in complex systems approaches to cognitive science has been divided between advocates of two apparently irreconcilable views. The discussion has been explicit for research in  $1/f$  patterns: while nomothetic perspectives on  $1/f$  noise promote general explanations of  $1/f$  noise (Van Orden et al., 2003; Dotov et al., 2010), and mechanistic perspectives on  $1/f$  noise focus on modelling specific processes underlying particular phenomena or calling for physiologically detailed mechanistic models (Torre and Wagenmakers, 2009; Rubinov

et al., 2011). Some of the latter accuse the former of proceeding ‘by identifying a mysterious phenomenon (i.e.  $1/f$  noise) and explaining this phenomenon by verbal reference to a series of other mysterious phenomena (e.g. SOC), without ever making contact with latent cognitive processes’ (Wagenmakers et al., 2012, p.91). This debate could be extended to many other metrics of complexity.

It is the tendency to ignore specific mechanisms that demands attention, since the alleged lack of explanatory power of nomothetic perspectives (making verbal references to other mysterious phenomena) is motivated by a lack of reference to constituent components of the explanation. Following Kaplan and Bechtel’s call for complementarity between nomothetic and mechanistic explanations (Kaplan and Bechtel, 2011) we believe that much is to be learned by attempting to connect and explore the relationship between behaviour generating mechanisms (at different levels of abstraction) and the regularities displayed by such mechanisms in operation. It is here that conceptual models (Di Paolo et al., 2000; Barandiaran and Moreno, 2006; Barandiaran and Chemero, 2009) can be useful. Conceptual models, such as the robotic model presented here, need not model specific target systems (i.e. a natural system, its behaviour and its physiological processes) but might bring into the model abstract mechanisms, general principles or generic processes, in order to study the relationship between different levels of explanation (e.g. between mechanistic and nomothetic explanations), between different types of phenomena or between different measurements and the principles or properties they capture providing valuable insights for future empirical experimentation and modelling.

In this sense, only measuring exponents of TSE complexity or  $1/f$  scaling is not useful for understanding autonomous behaviour if the mechanisms generating complex behaviour are not characterized. There are multiple examples of non-autonomous (and even trivial) systems generating high values of TSE complexity or  $1/f$  scaling. What is important is the role of different mechanisms in generating the complex structures captured by these indicators. By applying complexity and SOC indicators over a particular model (in which we can analyse in detail internal mechanisms in normal conditions or under different modifications), we can get a clearer picture about their origin and how they might be related to the notion of autonomy we are proposing. Thus, it is necessary to analyse what is the relation in our model of different complexity indicators and the mechanisms of sensorimotor coupling and synaptic plasticity generating them.

Frequently, measures of complexity like TSE complexity or  $1/f$  scaling have been connected with notions related to neurocognitive autonomy such as intentionality (Juarrero, 1999; Van Orden and Holden, 2002) consciousness (Tononi and Edelman, 1998). Although there is a wide consensus about the potential of such metrics to understand complex mechanisms, it is necessary go beyond their understanding and their dependence on specific, underlying, generating mechanisms. In this chapter we analyse whether these indicators are present, and what the role of internal mechanisms and sensorimotor regulation is for their generation:

- If an autonomous system has to be capable of spontaneously transiting among different dynamic regimes without these transitions being localized at any specific module, then

measures of simultaneous segregation and integration become valuable candidates to identify the degree to which a system might be considered autonomous. We make use of TSE complexity to analyse spatial segregation and integration among different variables in the agent’s controller and the environment to analyse the compromise between integration and segregation at different scales of systemic organization.

- Similarly, an autonomous agent faces a similar challenge in the temporal domain, it needs to coordinate particular sensorimotor contingencies at fast scales, but it also needs to create long-range correlations in order to create coherent patterns of macroscopic behaviour. Measures of fractal temporal scaling are suited to understand how multiscale correlations are generated within the agent’s controller and agent-environment interactions.

#### 4.4 INTEGRATION AND SEGREGATION AMONG THE SYSTEM’S COMPONENTS: TSE COMPLEXITY

Entropy-based complexity measures have been proposed to capture the balance between the integration of different brain areas and the local preservation of information that produces a rich variety in the patterns displayed by the brain. Specifically, *TSE complexity* was proposed by Tononi, Sporns and Edelman as a measure to capture the dialectic between the preservation of regionally specific dynamic and the maintenance of a global coherence in the system (Tononi et al., 1994). Later, Tononi and Edelman (1998) proposed that TSE complexity is a suitable measure for characterizing the kind of neural organization of their dynamic core hypothesis. There are different equivalent formulations of how TSE complexity is computed. One of them is to define TSE complexity in terms of the average mutual information shared between different bipartitions of the system summed over all bipartition sizes:

$$C(X) = \sum_{k=1}^{n/2} \langle MI(X_j^k; X - X_j^k) \rangle \quad (4.4)$$

where  $X$  is a system composed by  $n$  elements, and  $X_j^k$  is a subset of the system of size  $k$ . Thus,  $C$  computes the ensemble average of mutual integration between subsets of a given size and the rest of the system. Mutual information is a measure of the statistical dependence between two subsets of neural elements, as we described it in equation 3.7. In what follows we use TSE complexity as a measure of the joint maximization of integration and segregation predicted by the dynamic core hypothesis<sup>2</sup>.

---

<sup>2</sup> More recently, Tononi developed his information integration theory of consciousness (Tononi, 2004), claiming that consciousness corresponds to the capacity of a system to integrate information, and proposing a novel measure for the ‘level’ of consciousness generated by a system ( $\Phi$ ) as the amount of causally effective information that can be integrated across the weakest link of a system. As opposed to TSE complexity,  $\Phi$  is not a measure of complexity but irreducibility, and is much more computationally costly.

Generally, TSE complexity has been analysed for brain activity without taking into account bodily or environmental factors. However, there is no reason for not applying TSE or other information measures to variables placed in the body or the environment of the agent. Moreover, for measuring the type of sensorimotor autonomy we have defined, which does not take place exclusively in the head, we must take into account also those variables throughout the sensorimotor loop.

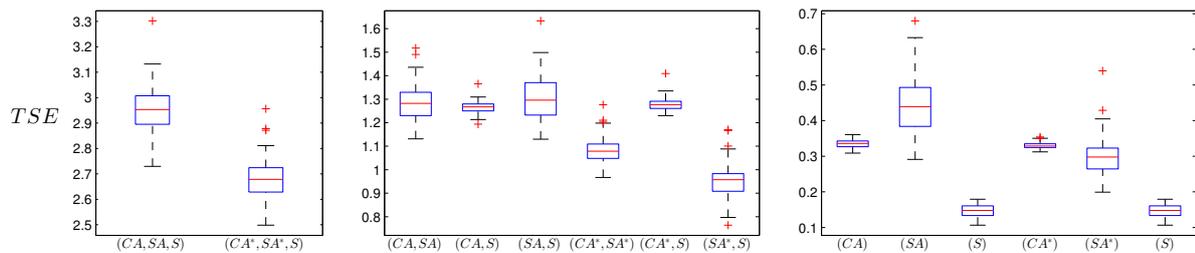
To our knowledge, there is so far no analysis of TSE complexity that explicitly takes into consideration an agent-environment coupled system as such. There are instances in which TSE complexity is analysed in a system as it is fed an input (Tononi et al., 1996), but they do not feature a strongly coupled system in which agent and environment influence each other. Nevertheless, there has been work characterizing information flows for embodied agents. In Beer and Williams (2015), a framework is proposed combining information flow and dynamical analyses, exploring how a simulated model agent in a relational categorization task integrates information about a cue used for solving the task. They evolve different kinds of agent, distinguishing between *passive* agents which store information about the cue in one of their interneurons, and *active* agents who store the information about the cue in the variable depicting the agent's position. This is an example of how information analyses may not only take into account the 'brain' of an agent but also its whole sensorimotor interaction, discovering counterintuitive processes going on in the agent-environment interaction.

#### 4.4.1 INFORMATION FLOWS BETWEEN OSCILLATORS, SYNAPTIC MECHANISMS AND SENSORIMOTOR COUPLING

We know that the agent presents both phase-locking mechanisms, which reduce the informational content in favour of a coordinated behaviour, and metastable states (which increases the information flow within the system). Above we hypothesized that the autonomous organization of the agent maximizes the compromise between integration and segregation. Here we analyse the extent to which the neural controller of our agent combines segregation and integration in its structure and how it is generated. Moreover, we want to understand the effect of sensorimotor and plastic regulatory loops, thus we analyse how integration and segregation takes place between agent and environment for the four types of agents.

We analyse TSE complexity in the agent's controller for different situations and different variable sets. Each situation is tested for an agent in 50 different trials of duration  $1.25 \cdot 10^5$  seconds. To reduce computational costs and similarly to what we did in Chapter 3, we simplify the variables of the system into binary variables. We take three binary variables  $CA_{ij}$  to characterize the phase relations between oscillators, which are equal to one if  $\theta_i - \theta_j > 0$  and zero otherwise. Similarly, we can define 6 binary variables  $SA_{ij}$ , which are equal to one if  $K_{i,j} > \frac{\alpha}{4}$ . Moreover, to take into account the state of the environment, we can define four variables  $S_i$  which encode whether each of the four sensors is activated above its mean value during the trial. Similarly, we define the variables  $CA_{ij}^*$  and  $K_{i,j}^*$  for the passively coupled agents (note that  $S_i$  is the same for both the situated and passively-coupled agents). For every agent, we compute

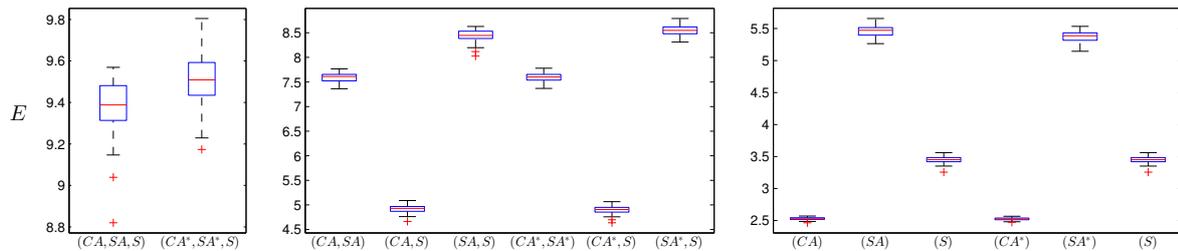
the values of TSE complexity for different subsets of variables for both situated and passively-coupled conditions. The result of analysing TSE complexity in the agent's controller with synaptic plasticity can be observed in Figure 4.4. The first thing we observe is that, for the set containing the three variable groups (CA, SA, S) (4.4, left), values of TSE complexity are higher for the situated agent than the passively coupled agents (note that we cannot compare here the values for the agents without plasticity since SA does not change with time). Situated agents with synaptic plasticity have around 10% more informational complexity than the passively coupled in similar conditions although they are being fed with the exact same signal. Moreover, we observe that values of TSE complexity are similar for (CA, SA) for the situated and passively-coupled agents, whereas the values of TSE complexity decrease again for variable sets (CA, SA) and (SA, S), indicating that the increase of complexity is only taking place when the synaptic weights are involved (4.4, middle). For groups of variables of the same type, the values of TSE complexity are similar for (CA) and (S), while they decrease again for the passively coupled agent for (SA) (4.4, right). This suggests that, although the individual complexity of CA and S are the same, there is a higher level of integration between those two groups of variables when synaptic plasticity is involved. Also, the complexity of SA increases when the agent is situated, suggesting that more complex patterns of weight connectivity emerge when the loop of sensorimotor configuration is operating. If we compare numerically the increase in value of complexity between situated and passively-coupled agents, we observe that while the mean increase of TSE values for (SA) is of 0.11, mean TSE complexity increases more for join sets of variables (SA, CA) (0.21), (SA, S) (0.36) and (CA, SA, S) (0.28). This means that although individual complexity of SA augments, the main increase in complexity takes place in the interaction between synaptic plasticity and oscillatory coupling or sensorimotor regulation, being especially intense for the interaction between synaptic plasticity and sensorimotor coupling.



**Figure 4.4: TSE complexity in agents with synaptic plasticity.** We measure TSE complexity for 50 trials for situated and passively coupled agents with normal synaptic plasticity. Variables belonging to passively-coupled agents are indicated with an asterisk (\*).

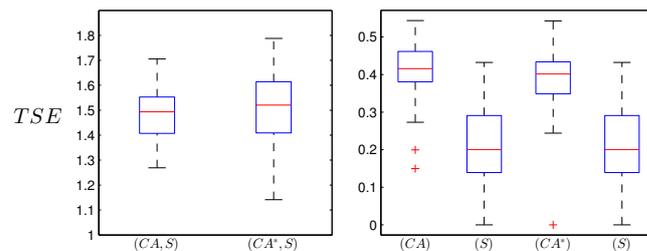
Moreover, if we analyse the amount of entropy of each set of variables (Figure 4.5), we find that the levels of entropy do not decrease for the passively-coupled agents (indeed, they even slightly increase for some cases as (CA, SA, S)). Thus, the increase of TSE complexity

for the situated agent cannot be explained in terms of an increase of entropy (segregation) in the system, but an increase of the integrative capacity of the situated agent. This suggests that the enhancement of neural integration achieved through sensorimotor coupling only happens when the agent has plastic mechanism that may engage or disengage different modes of coupling with the environment.



**Figure 4.5: Entropy in agents with synaptic plasticity.** We measure Shannon's entropy complexity for 50 trials for situated and passively coupled agents with normal synaptic plasticity. Variables belonging to passively-coupled agents are indicated with an asterisk (\*).

For the agent without synaptic plasticity (Figure 4.6), we observe that all the values of TSE complexity are similar in all cases between the situated and passively coupled agent. This suggests that synaptic plasticity has an essential role for enhancing integrated information of the agent's sensorimotor loop. A passively-coupled agent without synaptic plasticity has the same levels of integration and segregation of its variables than its corresponding situated agent. In this case, the closure of the sensorimotor loop on its own (without plasticity) does not add any extra complexity to the behaviour of the agent.



**Figure 4.6: TSE complexity in agents without synaptic plasticity.** We analyse the measure of TSE complexity for 50 trials for situated and passively coupled agents without synaptic plasticity. Variables belonging to passively-coupled agents are indicated with an asterisk (\*).

Together, these results indicate that the combination of a sensorimotor loop and synaptic plasticity boost the levels of complexity (understood as combined integration-segregation functions) of the agent-environment system. This is especially interesting since this increase

of complexity does not only take place in the internal variables of the agent's neural controller (CA, SA) but in a space of interaction between internal and sensorimotor variables (SA, S) and (CA, SA, S). Additionally, this increase of sensorimotor complexity only takes place for agents with synaptic plasticity, suggesting that enhanced complexity depends on the coupling between synaptic and sensorimotor regulatory loops: it is the active regulation of the sensorimotor loop by plastic controllers what allows a higher level of integrated information.

#### 4.5 FREQUENCY DOMAIN OF METASTABILITY: $1/f$ NOISE

Evidence suggests that the metastable regime of the brain is achieved by the cross-scale interactions of many levels of organization through the temporal coordination of neuronal oscillations that operate at multiple frequencies and at different spatial scales (Le Van Quyen, 2011). The existence of many spatial and temporal scales within the brain ensures that neural oscillations are never locked into a particular frequency. Instead, different oscillators are loosely coupled with adjacent frequencies in a regime of perpetual fluctuation. This creates a structure where no scale is favoured over the others. In terms of the frequency spectrum of brain activity, this is translated into a scale-free or  $1/f$  distribution of the frequency spectrum (Buzsaki, 2006, p. 119). This regime describes dynamical system where the amount of power the signal outputs over a certain bandwidth (its power spectral density) is inversely proportional to its frequency. If we think about it in logarithmic terms, it means that if we integrate the amount of power contained at each temporal scale (e.g. 1 – 10Hz, 10 – 100Hz, 100 – 1000Hz) the result is the same no matter what the scale is. Thus, the system does not display a preferred timescale of activity. This behaviour is typical of physical systems in a critical state where the correlation length tends to infinity (Reatto, 2007). An example of this is the liquid-vapor critical point, which is achieved under some specific conditions of temperature and pressure. However, there are certain kinds of systems that evolve spontaneously towards a point displaying critical behaviour. These systems are said to be in a state of self-organized criticality (see Section 1.2.3). The concept of SOC was initially proposed by Bak et al. (1987) to define certain classes of dynamical systems which have a critical point as an attractor, displaying critical behaviour without any significant 'tuning' of the system from the outside. Critical systems have very interesting properties, the most characteristic of which is that they spontaneously display a lack of a dominant scale of activity. They show complex dynamical responses and their statistical properties have to be described by power laws. Thus, critical systems typically display temporal and spatial scale invariance in the form of fractals and  $1/f$  noise, reflecting the process of propagation of long-range interactions based on local effects. The notion of SOC is appealing because it allows us to imagine systems that are able to self-regulate coordinated behaviours at different scales in a distributed manner and without a central controller, which are all properties that have been argued to be necessary for autonomy.

It has been shown that the brain is in a continuous state of SOC. Experimental evidence and artificial neural network models support this hypothesis. Criticality in the brain is suggested by power law scaling in degree distributions of functional brain networks or avalanche

size distributions (Chialvo, 2006). Long-range correlations in the form of  $1/f$  patterns have been extensively found in EEG/MEG measurements of brain activity in alpha, mu, and beta oscillations (Linkenkaer-Hansen et al., 2001). Scale-free neocortical dynamics has also been ascertained by Freeman in EEG measurements in rabbits (Freeman, 2005). In addition, there are many instances of critically self-organized behaviour in artificial network models, which provide some interesting insights (Rybarsch and Bornholdt, 2012). For example, it has been shown that a broad (robust) critical regime is favoured by the presence of strong synaptic interaction and high synaptic time-dependent plasticity (Rubinov et al., 2011).

SOC has been suggested as a candidate for characterizing a state allowing the system to remain in a self-organized state that is at the same time stable (or metastable) while also extremely sensitive to small microscopic perturbations, and with the capacity to develop long-range correlations at all the scales of the system without requiring a costly fine tuning of local parameters (Werner, 2007b). In turn, the presence of  $1/f$  noise has been proposed as an indicator of the long-term evolution of a dynamic core in EEG measurements (Le Van Quyen, 2003).

#### SELF-ORGANIZED CRITICALITY AND EMBODIED AUTONOMOUS BEHAVIOUR

The extensive finding of  $1/f$  noise patterns does not only apply to the brain but also extends to behaviour. SOC and  $1/f$  have also been extensively found in cognitive science and psychology. For example,  $1/f$  noise is present in time series showing how performance fluctuates over time (Gilden, 2001). More recently, Van Orden et al. (2003, 2005) used  $1/f$  noise measurements in different tasks to gather evidence to defend the idea that certain systems are not modular and decomposable but are ‘softly assembled’ systems sustained by *interaction-dominant dynamics* (IDD hereafter) as opposed to *component-dominant dynamics* (Van Orden et al., 2003). That is, IDD systems do not consist of additive interactions of their components, but of multiplicative interactions that imply coordination between the different timescales in the system. Silberstein and Chemero suggest that  $1/f$  measurements can be an indicator of brain-body-environment systems of nested self-organization (Silberstein and Chemero, 2012). This hypothesis is inspired by measurements of  $1/f$  noise in a simple human-tool interaction with and without perturbations in the interface between *human and tool* (Dotov et al., 2010). Self-organized criticality thus becomes an interesting candidate for explaining the holistic and integrated nature of mental life; and its  $1/f$  noise manifestation a signature or quantitative indicator of this holistic integration (Silberstein and Chemero, 2011).

Whereas evidence for the presence of  $1/f$  noise in neural dynamics and in certain experimental psychology tasks is apparent (mostly behavioural data in the latter case), there is very little understanding of how both neurodynamic and behavioural levels relate to each other and to  $1/f$  noise. Moreover, most models proposing generative mechanisms for the emergence of  $1/f$  noise are focused on either a behavioural (Torre and Wagenmakers, 2009) or a neural level (Usher et al., 1995; Rubinov et al., 2011). Although insightful, these models fall short of providing in-depth explanations about the nature of neural and behavioural organi-

zation in embodied subjects in a softly assembled brain-body-environment system, because in such models the systems are never shown to operate embedded in a specific environment.

#### THE CONTROVERSY OVER THE EXPLANATORY CAPACITY OF $1/f$ NOISE MEASUREMENTS

Despite the enthusiasm aroused by the widespread finding of  $1/f$  noise, some authors remain cautious about its specific interpretation. Van Orden et al. claim that ubiquitous pink noise is not sufficient evidence of self-organized criticality but rather a necessary consequence (Van Orden et al., 2003, p. 343). They argue that while it is possible to interpret  $1/f$  noise in some contexts as a signature of a softly assembled system (self-sustained by IDD and SOC) this is not always necessarily the case. Moreover,  $1/f$  is not a unique and exclusive property of SOC systems (see (Wagenmakers et al., 2004, 2012)) since it can be displayed by a linear superposition of random components acting on multiple time scales (Hausdorff and Peng, 1996). To avoid the uncertainty about the true origin of  $1/f$  noise some authors have suggested complementing it with a measure of multifractality as a quantitative indicator of interaction-dominant dynamics, providing a means to ensure the nonlinear nature of the ongoing interactions that build the self-organized process (Ihlen and Vereijken, 2010).

#### 4.5.1 SELF-ORGANIZED CRITICALITY IN AUTONOMOUS ORGANIZATION

So far, we have quantified for different types of agents the dynamic creation and dissolution of patterns through metastable mechanisms and through the integration of information loops through the sensorimotor loop. All these analyses deal with how activity is coordinated between different variables of the system and the environment, displaying a picture of how different spatial scales are integrated into a complex but coherent unit. In this section we analyse interaction between scales in the temporal domain. We analyse  $1/f$  patterns for the three different agents (situated, decoupled and without synaptic plasticity) to find out if they can be characterized as SOC systems, being especially careful to rule out false positives of  $1/f$ -like patterns that are not produced by SOC. We first use two different methods to characterize  $1/f$  patterns (fractal and spectral methods). Then, we measure multifractal exponents for characterizing nonlinear correlations between scales .

#### FRACTAL AND SPECTRAL METHODS

For the same four conditions (situated or passively-coupled, with and without plasticity) we analyse 50 simulations during 125000s (same as above) and looked for long-range correlations in different variables. A preliminary analysis shows that long-range correlations are found in series of sensory data and the instantaneous frequencies of the oscillators (i.e. the derivatives of  $\theta$ ). For other variables such as modification of synaptic weights, long range correlations did not appear or disappeared for longer timescales. Thus, we measure  $1/f$  correlations in two

different signals

$$\Omega = \frac{1}{3} \sum_{i=1}^3 \dot{\theta}_i \quad (4.5)$$

where  $\dot{\theta}_i$  is the first derivative of the phase of oscillator  $i$  (therefore being  $\Omega$  the mean instantaneous frequency of the oscillator ensemble), and

$$S = \frac{1}{4} \sum_{DX} s_{DX} \quad (4.6)$$

where  $s_{DX}$  are the four sensors of the agent (D standing for right or left sensor and X for light A or light B sensor). Both of these signals present correlations in the form of  $1/f^\beta$  patterns, at least in some cases, extending up timescales of the same magnitude of the duration of the simulation.

Simply identifying a linear slope in a logarithmic representation is not enough for characterizing  $1/f^\beta$  noise. Van Orden et al. (2005) propose the use of fractal and spectral methods in tandem in order to avoid the mistake of taking transient correlations for scaling relations. We therefore use two different methods for characterizing  $1/f$  dynamics: the discrete Fourier transform (DFT) and detrended fluctuation analysis (DFA). DFT allows us to decompose a signal into its different frequency components. Then, we use Welch's method (Welch, 1967) to estimate the signal's power spectrum. Welch's method divides a signal into different overlapping sections (we choose eight sections with 50% overlap), and each section is windowed with a Hamming window and eight modified periodograms are computed and averaged. If the resulting power spectrum has the form of a  $1/f^\beta$  function, we say that the signal exhibits fractal dynamics. Usually, processes with  $\beta$  around 1 are considered to display SOC (Jensen, 1998).

DFA (Peng et al., 2000) is a method for determining the statistical self-affinity of a signal. In a nutshell, the DFA algorithm removes the mean and integrates (cumulatively sums) the analysed time series  $x(i)$  into  $y(k)$  and then divides it into segments of equal length  $n$  (i.e. of a certain time scale). For each segment, the least squares line (the trend of the signal within that segment) is fitted to the data obtaining a local linear approximation  $y_n(n)$ . The characteristic size of the fluctuation  $F(n)$  is computed as the root mean square deviation between the integrated signal and its trend in each segment. This computation is repeated for every value of  $n$ .

$$y(k) = \sum_{i=1}^k x(i) \quad (4.7)$$

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^N [y(k) - y_n(k)]^2} \quad (4.8)$$

where  $N$  is the total length of  $x(n)$ . Typically,  $F(n)$  increases with  $n$ . A linear relationship on a log-log plot with slope  $\alpha$  indicates the presence of fractal scaling in the analysed signal,

where  $\alpha$  is a generalization of the Hurst exponent, and is related to the scaling in the Power Spectrum of the Fourier analysis being  $\beta = 2 \cdot \alpha - 1$ .

DFA has some advantages compared to DFT analysis. While DFT is only well suited for stationary signals, DFA has been reliably used for non-stationary signals. Moreover, results from DFT are often noisy and sometimes not absolutely reliable for determining linear relationships in logarithmic scales (e.g. see Wagenmakers et al., 2004). We focus our analysis in a temporal scale ranging from  $10^2$ s to  $10^{4.5}$ s<sup>3</sup>.

#### DETRENDED FLUCTUATION ANALYSIS

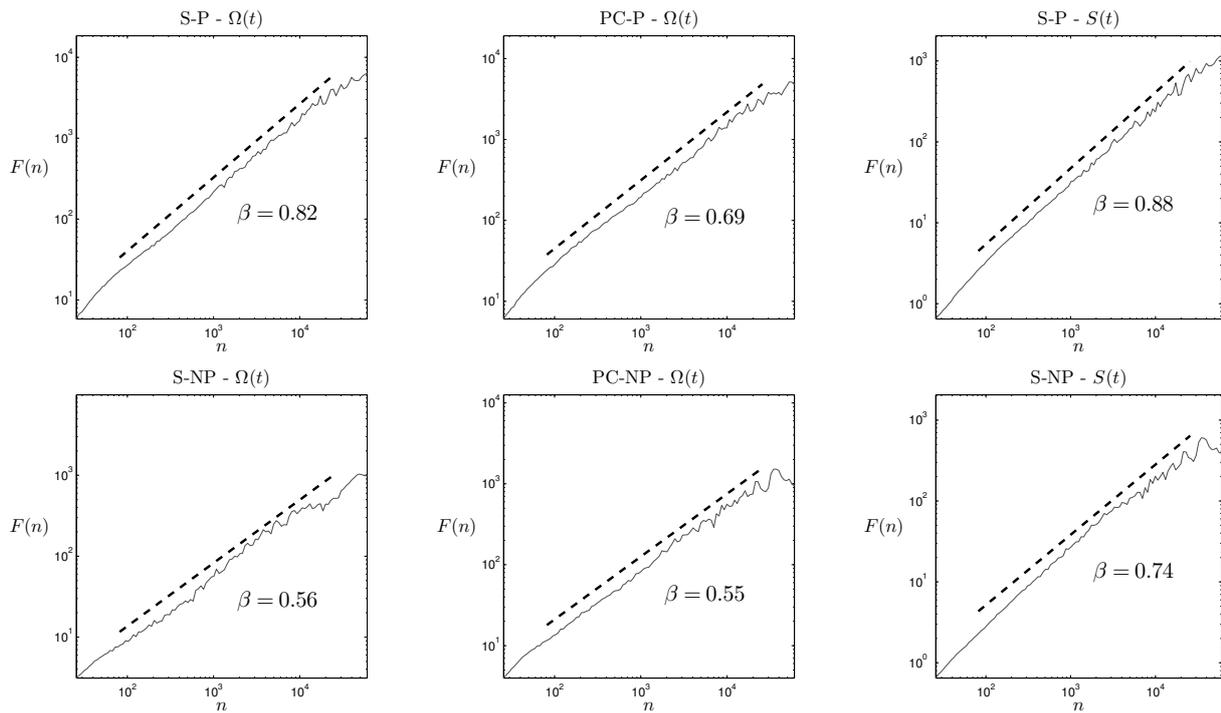
If we take different series of the variables  $\Omega$  and  $S$  from equations 4.5 and 4.6 for the situated agent and apply the DFA algorithm to them, we find that the signal presents linear relationships on a logarithmic scale from  $10^{2.0}$ s up to around  $10^{4.5}$ s. Note that in previous chapters we observe that  $10^2$ s was the approximate time that the agent needs to reach a light, therefore making long-range correlations belong to the different timescales of the agent's behaviour. If we compute the  $\beta$  coefficient for each variable and agent (see representative examples at Figure 4.7) we find that it is quite close to pink noise for both variables in the case of the situated agent. We observe that the values of the  $\beta$  exponents are typically close to 0.8 for  $\Omega$  and 0.9 for  $S$ . For the passively-coupled agent we observe that the value of  $\beta$  decreases with respect to the situated agent for  $\Omega$  (note that the value of  $S$  is the same in the situated and passively-coupled agents, since the input is recorded from the former and fed into the latter). In the case of the agents without synaptic plasticity, the coefficients of  $\beta$  vary widely compared to agents with plasticity. Moreover, in most cases the correlations of  $\Omega$  cut-off around  $10^4$ s, not reaching higher timescales. Further, we observe that in many cases the value of  $\beta$  decreases again for passively-coupled agents. We can run the simulation several times for different initial values and we always find quite similar results, with  $\beta$  around 0.77. This result suggests that the process of the emergence of complex patterns analysed above is generated by SOC.

#### DISCRETE FOURIER TRANSFORM

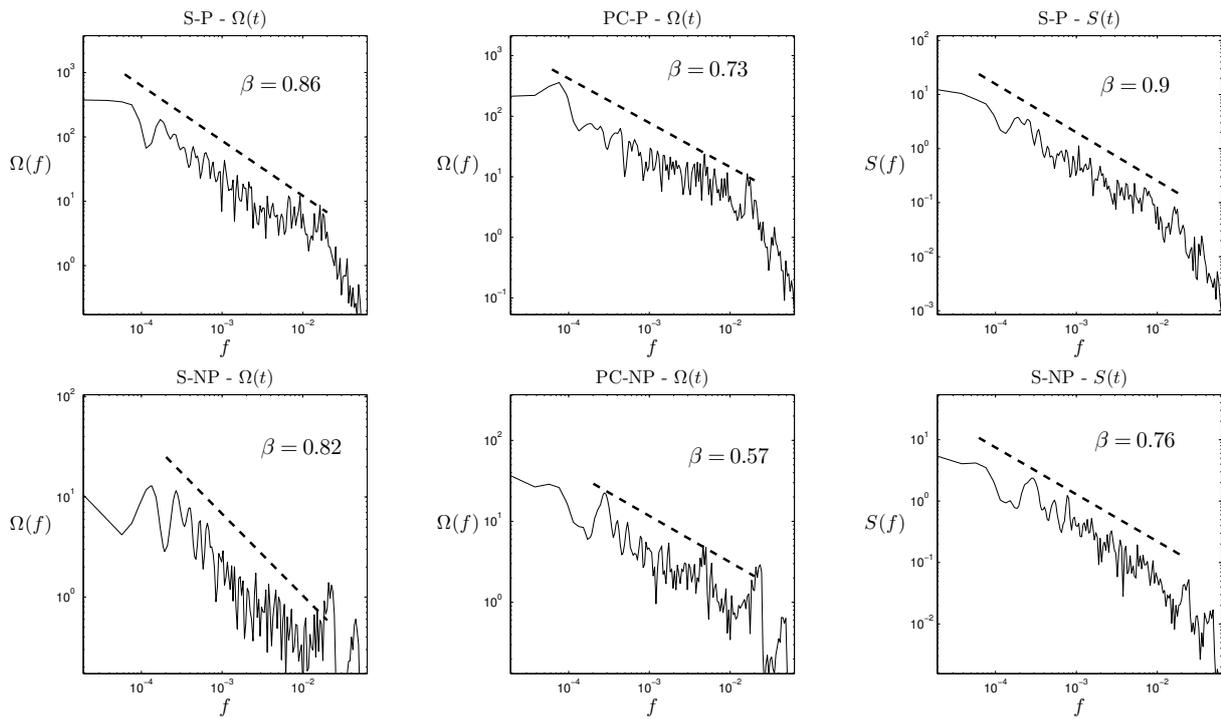
We have seen that sensorimotor coupling and synaptic plasticity play an important role in the generation of SOC dynamics. To further test this result, we have computed the equivalent DFT spectrum of the signal. For the situated signal, we find a region with a fractal scaling for frequencies between  $10^{-2}$ Hz and  $10^{-4.5}$ Hz very close to pink noise for both variables  $\Omega$  and  $S$ , confirming the results of fractal scaling for the situated agent with synaptic plasticity. The passively-coupled agent also presents similar results, although we observe how fractal scaling of variable  $\Omega$  is cut-off at  $f = 10^{-4}$ Hz. We observe that the same happens for the agents without plasticity (both situated and passively-coupled). Thus, from now on we restrict the frequency range of analysis of variable  $\Omega$  to  $[10^{-2}$ Hz,  $10^{-4.5}$ Hz].

---

<sup>3</sup>Although the figures show a broader spectrum, calculations of log-linear relations are made within this range or subsets of this range



**Figure 4.7: Detrended fluctuation analysis.** We observe that situated agents present values of  $\beta$  close to 1 for both variables  $\Omega$  and  $S$ , indicating the presence of long-range correlations. The value of  $\beta$  decreases when the agent is passively-coupled, removed of synaptic plasticity or both at the same time.



**Figure 4.8: Discrete Fourier transform.** We observe the results from the DFA algorithm are consistent with the spectral analysis, confirming the presence of  $1/f$  patterns shown by Figure 4.7.

## FRactal Scaling and Generative Mechanisms

We have shown the particular case of one agent and how long-range correlations are reduced when strong sensorimotor coupling or synaptic plasticity are lost. In figure 4.9 we show the distribution of fractal exponents computed with the DFA algorithm of 50 different agents for each condition, confirming the generality of the results. As we observe, fractal exponents are very close to 1 for situated agents with synaptic plasticity ('S-P') for both internal ( $\Omega(t)$ ) and interactive ( $S(t)$ ) variables.

### INTERNAL DYNAMICS

For passively-coupled agents ('PC-P') the fractal exponents of  $\Omega(t)$  are reduced and present a different distribution from the situated case according to the results from the ANOVA [ $F(1, 98) = 301.05, p = 1.22 \cdot 10^{-31}$ ]. This indicates that long-range correlations in internal variables of the oscillatory components are not built internally nor induced from an external input, but built in dialectical interaction between the agent's brain and its environment. It is the process of agent-environment interaction where long-range dependencies are generated.

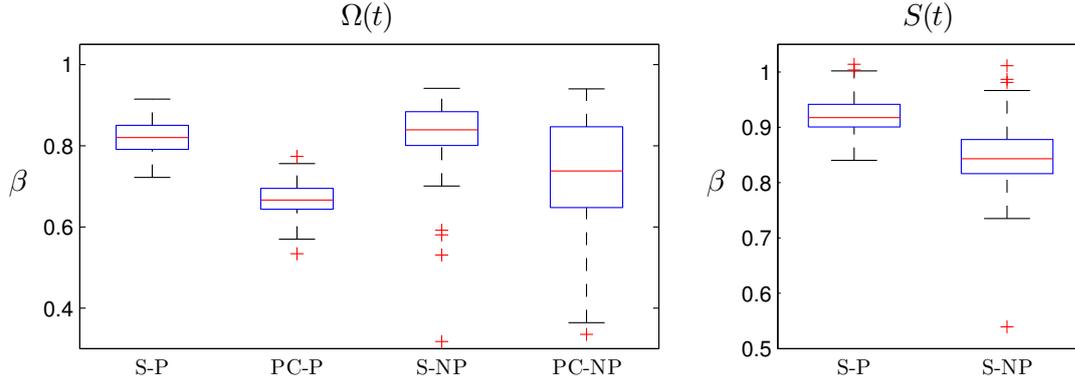
Moreover, for the agents without synaptic plasticity ('S-NP' and 'PC-NP'), we obtain similar scaling values for  $\Omega(t)$ , showing that agent-environment interaction generates scale-free patterns of behaviour in our agent. Agents without plasticity also show a decrease in  $\beta$  exponents when situatedness is lost, although in this case it is not clear that the difference between the two distributions is statistically significant [ $F(1, 98) = 3.19, p = 0.0772$ ].

### SENSORIMOTOR DYNAMICS

If we analyse fractal scaling in  $S(t)$ , we find a different distribution between agents with and without synaptic plasticity [ $F(1, 98) = 32.56, p = 1.23 \cdot 10^{-7}$ ]. The  $\beta$  coefficients decrease for agents without synaptic plasticity. Thus, agents without plasticity present regular long-range correlations in their internal variables but not in the interactive variables of agent-environment interaction. This supports our hypothesis that the role of plastic mechanisms is to regulate sensorimotor coupling, therefore generating complex correlations not only in internal variables but cutting through brain-body-environment divisions. Although not showed in the figure, results applying DFT analysis display a similar picture, and the ANOVA give similar results for the comparisons of the distribution of  $\Omega(t)$  between 'S-P' and 'PC-P' [ $F(1, 98) = 115.1, p = 3.20 \cdot 10^{-18}$ ], the distribution of  $\Omega(t)$  between 'S-NP' and 'PC-NP' [ $F(1, 98) = 2.73, p = 0.1015$ ] and the distribution of  $S(t)$  between 'S-P' and 'S-NP' [ $F(1, 98) = 7.39, p = 0.0077$ ].

### Interaction-Dominant Dynamics and Multifractality

The 'constructive' definition of SOC coined by (Jensen, 1998, p.126) emphasizes that SOC systems are interaction-dominated threshold systems driven by inherently nonlinear coordinative mechanisms. Ihlen and Vereijken have proposed that  $1/f$  noise is neither necessary nor sufficient evidence of interaction-dominant dynamics and that multifractal analysis is a

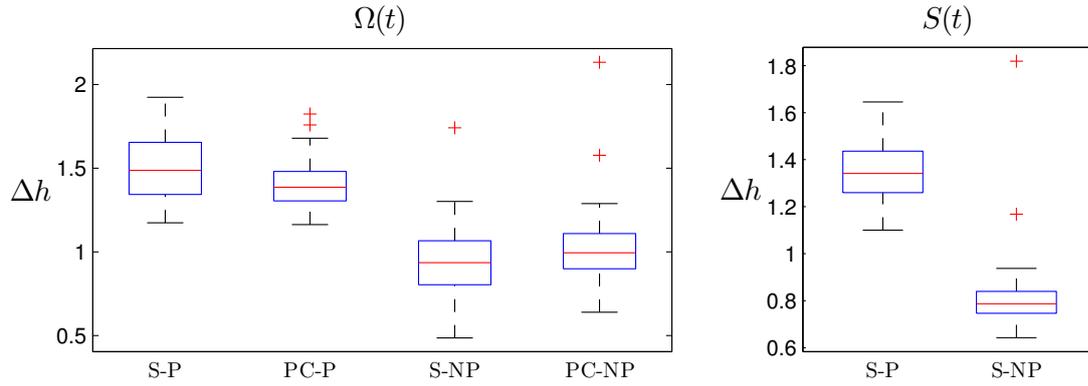


**Figure 4.9: Fractal exponents for variables  $\Omega(t)$  (left) and  $S(t)$  (right) for the four types of agents.** Only situated agents with synaptic plasticity present indices of  $\beta$  close to 1 for both variables  $\Omega$  and  $S$ . See the text for statistical comparisons between the distributions.

quantitative framework suitable for the analysis of interaction-dominant behaviour (Ihlen and Vereijken, 2010). The multifractal spectrum quantitatively defines the presence of multiplicative interactions between temporal scales that are responsible for the emergence of intermittent, emergent or coherent periods of large fluctuations within the response series of a system.

We computed the multifractal spectra of the different signals analysed above using the multifractal DFA algorithm (MFDFA), a variation of DFA in which the squared exponent of the root mean squares deviation becomes a variable  $q$ , therefore allowing calculations outwith the standard euclidean norm defined by the root mean square. Following this procedure, positive  $q$  values describe the scaling behaviour of the segments with large variance because the large deviations from the corresponding fits dominate the average  $F(n)$ . On the contrary, negative  $q$  values describe the scaling behaviour of the segments with small variance because the large deviations from the corresponding fits are largely attenuated on the average  $F(n)$  (Kantelhardt, 2008). This decomposition allows us not only to analyse the scaling of the variance of a signal (as with the DFA algorithm) but also the entire probability density function defined by all  $q$ -order statistical moments through a series of local exponents  $h$  associated with each  $q$ . The multifractal spectrum is represented by an arc defined as the difference between the maximum and minimum values of the local Hurst exponent for each scale ( $D(h)$  vs.  $h$ ). Thus, the width of this spectrum is a measure of the degree of multifractality and will be zero for a monofractal series. The higher the value of the width the more multifractal the spectrum is. The width of the multifractal spectrum  $\Delta h = h_{\max} - h_{\min}$  defines the amplitude difference between the variability in the intermittent and laminar periods of the response series, quantifying the influence of the multiplicative interactions between the multiple time scales of the time series.

We have computed both the  $\beta$  coefficient from DFA analysis and the multifractal spec-



**Figure 4.10:** Width of the multifractal spectra ( $\Delta h$ ) of variables  $\Omega(t)$  (left) and  $S(t)$  (right) for the four types of agents. Situated agents with synaptic plasticity presents the higher indices of  $\Delta h$  for both variables  $\Omega$  and  $S$ . See the text for statistical comparisons between the distributions.

trum of the variables  $\Omega$  and  $S$  for 50 different runs of the four types of agents. We have used the same temporal ranges for the MF DFA analysis than we used in the DFA analysis ( $n = [10^2s, 10^{4.0}s]$  for  $\Omega$  and  $n = [10^2s, 10^{4.5}s]$  for  $S$ ). Due to computational reasons, and since the smallest timescales are not considered in the MF DFA analysis, we have decimated the time series by a factor of 20, although similar results were found for individual runs without decimation. The multifractal spectrum was computed using a range of  $q$  of  $[-15, 15]$  with a step of 0.5.

The results of the multifractal analysis are shown in Figure 4.10. We can observe how agents with plasticity present a broader multifractal spectrum than agents without plasticity for both  $\Omega$  and  $S$ , suggesting that interaction-dominant dynamics are more present in the former types of agents. Also, situated agents with plasticity present a slightly higher index of multifractal spectrum width. Also, agents with plasticity present different distributions when they are situated or passively coupled in the multifractal spectra of  $\Omega$  (ANOVA [ $F(1, 98) = 8.87, p = 0.0037$ ]), being higher the width of the multifractal spectrum when the agent is situated. This difference is not clear for agents without synaptic plasticity (ANOVA [ $F(1, 98) = 2.00, p = 0.1603$ ]). Also, the multifractal spectrum of  $S$  is significantly higher for agents with plasticity than agents without plasticity (ANOVA [ $F(1, 98) = 326.80, p = 5.63 \cdot 10^{-33}$ ]). The presence of broad multifractal spectrum confirms the hypothesis that the dynamics of the agents are generated from nonlinear threshold-driven interaction between different scales, and are not the result of finely tuned additive interactions. Also, the increase of the levels of metastability and TSE complexity for situated agents with synaptic plasticity is accompanied by an increase in the width of the multifractal spectrum, supporting the relation between these indices.

## RECAPITULATION

In this section we have analysed fractal and multifractal patterns in internal ( $\Omega$ ) and sensory (S) variables for different types of agents. We have observed how situated agents with synaptic plasticity present indices of  $\beta$  close to 1 and the highest levels of multifractality for both internal and sensorimotor variables. This suggests that the system is self-organized to maintain certain temporal invariances across its different scales of operation, operating in a regime of self-organized criticality, driven by multiplicative interaction between its components. Furthermore, we observe how (1) the fractal and multifractal indices of internal oscillatory variables decrease when the agent is passively-coupled to the world, and (2) the same happens for sensorimotor variables when synaptic plasticity is removed. Interestingly, disrupting the sensorimotor loop weakens internal long-range correlations, and removing internal plasticity weakens sensorimotor long-range correlations. This further supports our hypothesis about internal regulatory plastic loops and adaptive sensorimotor loops being intimately entangled in a co-regulatory loop. If some enabling conditions between internal and sensorimotor process is to be disrupted, the invariances generated at higher levels fades away. The behaviour of the agent as a whole is generated in this circular regulation and cannot be reduced to particular components of the system.

It is worth noting that passively-coupled agents present significant differences respect to situated agents despite a passively-coupled sensorimotor loop presents only a ‘subtle’ distortion of the agent-environment interaction (since the structure of the input is maintained). More dramatic distortions would have more visible effects. For example, in [Aguilera et al. \(2015\)](#) we tested the same agent presented here decoupled from its environment and being fed a Gaussian white noise signal, finding that  $1/f$  patterns are not robust as in the situated case and are not maintained for some parametric configurations. Also, multifractal exponents decrease radically for decoupled agents.

## 4.6 DISCUSSION

In Chapter 3 we characterized internal and sensorimotor regulatory loops for an agent controlled by a plastic neural controller which performs a behavioural preference task, choosing alternatively between two lights and generating stable but transient preferences for them. These emergent regulatory circuits took the form of a double feedback loop, one regulating the interaction between the coupled oscillatory neural controller and the plasticity of synaptic weights between oscillator pairs, and the other regulating the interaction between the neural controller and the sensorimotor loop (see [Figures 3.16 and 3.19](#)). We mapped how the internal regulatory loop involved a bidirectional connection between oscillators and plastic mechanisms, in which synaptic configurations influence the emergent synchronized pattern in fast timescales, and synchronized patterns influence the change of synaptic connectivity patterns in slow timescales. In the sensorimotor regulatory loop, sensory stimulation influences the oscillator network at slow timescales while synaptic connectivity patterns determines the agent’s

behaviour at longer timescales. Thus, we have a nested system with the coupled oscillator network generating emergent patterns at the top of the hierarchy (slow timescales), sensory inputs at the bottom (fast timescales), and synaptic connectivity mediating between both.

What kind of spatial and temporal structure supports the deployment of these regulatory loops? In the previous chapter we have analysed the regulatory activity of the autonomous agents as a system of three macroscopic components (synchronizing oscillatory patterns, synaptic connectivity patterns and sensorimotor patterns), but all these are composed by groups of variables presenting different spatial and temporal regularities. In this chapter we have analysed how these regularities are organized and what is the role of plastic and sensorimotor mechanisms in their emergence and coordination. As we reviewed, there is an ongoing debate between nomothetic and mechanistic approaches respect to the validity of different methodological perspectives in order to reveal the origin and significance of regularities as scale free  $1/f$  dynamics. While the former try to find general explanations for the emergence of regularities, the latter seek to find specific models of the particular processes underlying the emergence of regularities. An account of autonomous behaviour should combine both approaches in order to understand the relations between behaviour-generating mechanisms and the general properties of emergent autonomous behaviour. In this chapter we have used our model as a mediator between nomothetic and mechanistic approaches and how they can inform future experimental research where regularities as information integration or self-organized critically arise from a brain-body-environment system. The proposed models aim at exploring a middle point between the two approaches cited focusing on the role of the sensorimotor loop to connect macroscopic behaviour with neural self-organized dynamics.

#### EMERGENT PATTERNS AND METASTABILITY

For understanding the relation between low level mechanisms and macroscopic emergent patterns we compared four different types of agents: regular situated agents, passively-coupled agents (whose sensors are fed with the recorded input of a situated agents), agents without synaptic plasticity, and passively-coupled agents without synaptic plasticity. As a first step, we analyse the formation of patterns in the neural oscillator network and the levels of metastability of these patterns. We observe two important differences between situated and passively-coupled agents (with synaptic plasticity): situated agents are able to create coherent patterns of synchronization and they display higher levels of metastability. In contrast, passively-coupled agents present a fuzzy set of synchronization patterns in comparison and present lower levels of metastability. Situated agents were at the same time able to present more structured patterns and display a more metastable structure facilitating switching among different patterns. These differences between situated and passively-coupled agents did not take place for agents without synaptic plasticity, indicating that plastic mechanisms are which allow the agent to present higher degrees of metastability while maintaining internal coherence of its synchronizing patterns.

## TSE COMPLEXITY

The next step was to analyse information integration and complexity among the different low level variables of the agent. For agents with synaptic plasticity, we find that measures of TSE complexity increase in situated agents for all sets of variables incorporating the agent's plastic weights compared to passively-coupled agents. Moreover, we find that this increase of TSE complexity is not due to an increase of the entropy of the variables, indicating that changes are due to level of integration between variables. Also, increases of TSE complexity cannot be explained only in terms of an increase of complexity of synaptic connectivity, since the greatest increases of complexity take place in the coordination of synaptic connectivity with other variables of oscillatory coupling and sensorimotor coordination. This increase of plasticity does not take place for agents without synaptic plasticity. All this indicates that plasticity in the neural controller mediates for an increase in the dynamic integration among different variables of the oscillatory controllers and sensorimotor coordination. This mediation allows the creation of macroscopic brain-body-environment synchronized loops enhancing a bidirectional flow of information between sensory streams across the environment and particular neural ensembles.

## SELF-ORGANIZED CRITICALITY

Finally, we have analysed the temporal macroscopic behaviour of different variables, analysing the presence of self-organized criticality (SOC) in the dynamics of the system. SOC refers to the phenomenon whereby a dissipative dynamical system with many degrees of freedom operates near a configuration of minimal stability, i.e. a critical configuration, and does so without any fine tuning by an external driving influence. For SOC to emerge, the dynamics of a system must be dominated by mutual interaction of the many degrees of freedom comprising the system, interaction-dominant dynamics being a necessary condition for the emergence of SOC. Also, SOC systems usually display robust  $1/f$  patterns without fine parameter tuning, being resistant to alterations in the system's parameters. In this chapter we have used scale-free exponents (fractal and spectral) and multifractal exponents as indicators of SOC, finding that robust  $1/f$  noise and high values of multifractality are only present in variables of the oscillatory controller and sensorimotor activation when the agent is situated in its world and plastic mechanisms are active.  $1/f$  patterns arise in the mean instantaneous frequency of the neural controller oscillators and in the mean activation of the sensors. In the case of the oscillators mean instantaneous frequency the exponents of  $1/f$  noise and the width of the multifractal spectrum decrease when the agent is passively-coupled. Also, the exponents of  $1/f$  noise for the mean sensor activation level decreases when the agent loses its plasticity. Interestingly, SOC in internal neural coordination patterns is only maintained when the agent is actively coupled to its environment and, in turn, SOC in sensorimotor activation patterns only arises when the agent has internal plastic regulatory loops. Interestingly, SOC appears to emerge only when a subsystem is regulated by other parts of the system (oscillatory patterns by sensorimotor coordination, and sensorimotor activity by plastic adaptive mechanisms). This adds

an interesting picture to our characterization of a double regulatory loop, in which both loops are entangled with other elements outside of the loop, generated a unified system in which the interaction between different parts of the system enables the processes sustaining the system's activity.

#### 4.6.1 A MINIMAL MODEL OF NEURODYNAMIC AUTONOMY

Note that the model presented here, being abstract and not empirically driven, does not prove that the properties just mentioned apply to natural behaving organisms, but rather shows that it *could* be the case, how it could occur, the measurement tools and analytic procedures that might capture it and which are the levels of observation that should be considered. In this sense the model works as a proof of concept that can inform and drive future experiments operating as what Chemero calls 'a guide to discovery' (Chemero, 2009; Barandiaran and Chemero, 2009). Thanks to our model we can hypothesize that, at least for some cases, TSE complexity and 1/f patterns in behaving systems *might* be the result of: (1) non-linear interaction dynamics capable of generating stable collective patterns, (2) internal plastic mechanisms that allow self-sustained criticality through a continuous modulation of sensorimotor flows, (3) strong sensorimotor coupling with the environment that induces transient metastable regimes and, (4) a small number of behaviour generating components or variables. The last three elements are relatively novel, in relation to previous models, and deserve detailed discussion.

So far, complex systems approaches to cognitive science face the problem that they are not able to cope with the definition of a cognitive system as such. Perfectly homogeneous systems for example rice piles are able to display SOC, thus the finding of SOC in the brain might provide us some insights into the way in which the brain operates, but without concrete mechanistical models it tells us little about the mechanisms that are operating behind it and how these construct a cognitive subject (Wagenmakers et al., 2012). Similar concerns have been expressed about integrated information approaches like TSE complexity, since many non-agential systems like low-density parity-check algorithms are able to produce high values of integrated information (potentially higher than human brains; Aaronson, 2014).

In Chapter 3 we proposed a definition of neurocognitive interactive autonomy stating that the nervous system, body and environment are organized in an operationally closed network in which:

1. **OPERATIONAL CLOSURE** among a set of processes a system  $S$  can be distinguished as a network of interdependent processes whereby every process belonging to the network depends on at least another process of the network and enables at least another one.
2. **CAUSAL ASYMMETRY** in the interaction between  $S$  and  $E$ , which modulates the coupling  $C$  between  $S$  and  $E$  in an adaptive way. Modulation indicates an alteration in the set of constraints that determine the coupling between  $A$  and  $E$ , and adaptive means

that change in the coupling  $C$  contributes to the maintenance of some process that constitute  $S$ .

3. **PRECARIOUSNESS.** By virtue of the two properties above, the processes that constitute  $S$  are actively created and maintained by the system as a whole, so that if we were to isolate from the system any process of  $S$  it would tend to run down or extinguish.

In the previous chapter we provided a conceptual model of a robotic agent with a plastic neurodynamical controller in a behavioural preference task. Our results illustrate how these three conditions are accounted for in the model, serving as a minimal example of how different properties of autonomous organization can be analyzed and quantified.

#### OPERATIONAL CLOSURE

In the previous chapter, we characterized the operational closure of the system, i.e. the ability of an agent to appear as a unified system in itself, without the need of an arbitrary division between agent and environment by an external observer (Figure 3.19). Describing the behaviour of the model using three mesoscopic groups of variables, we are able to define circular dependencies between the different variables of the system delimiting its closure (Virgo et al., 2011). Importantly, this closure is not restricted to the internal dynamics of the agent, but it extends to interactive variables defining the engagement of the agent in different sensorimotor loops.

The specific constitution of this closure, involving interactions at different timescales of the system, has important implications for the other two conditions for autonomy. The results presented in this chapter provide a characterization of agent-environment asymmetrical interaction and the constitution of the agent as a distinct entity from its environment.

#### CAUSAL ASYMMETRY

An autonomous agent, to be considered as such, requires the ability to act over its environment by itself. This has been described as the causal asymmetry condition for autonomous agency (Barandiaran, 2008, p.85-85). As we represented in Figure 3.16, regulation loops in our agent present such causal asymmetry, in which interaction is directional at different scales, and while agent is influenced by the environment at small timescales, at larger timescale the directionality is inverted and it is the agent which causally acts on the environment.

Some information theoretical approaches have been recently proposed to measure autonomy as the causal asymmetry between a system and its environment (Bertschinger et al., 2008; Seth, 2007) defined in terms of two properties: (1) self-determination, i.e. the system is dependent on its own history (2) non-heteronomy, i.e. the system is not completely determined by external factors. Nevertheless, although these measures try to compute the asymmetry between agent-environment dynamics, they do not take into account a self-referential operational closure of the system (Bertschinger et al., 2008, p.14). Instead, only the degree of

self-determination of the system is quantified. For example, Bertschinger depicts a simple automata with two states  $\{A, B\}$  that change its state if the state of the environment is 0, or keep it if it is equal to 1. This simple automata is self-determined and therefore would be autonomous according to this measure, although it is not able in any sense to act on its environment.

A different perspective on causal asymmetry has been depicted by Tononi et al. (1996) as the matching complexity of a system, defined as the increase of TSE complexity of a system receiving an input above the sum of the complexity values of the decoupled agent and the input. Although the system depicted by Tononi et al. was only passively fed an input, without the existence of any sensorimotor loop, the presence of matching complexity reveals and increase in the intrinsic flows of mutual information between subsets of different sizes in the system, by which different units correlate their responses in correlation with incoming stimulus, reflecting an increase of both integration of the system and functional specialization as a response to the external input. Each of the subsets responding to a given stimulus has a specialized relation to the rest of the brain, so that redundant inputs from the stimulus can lead to different functional consequences. In this sense, the response to an external input is not strictly passive, but it has the ability to literally go ‘beyond the information given’ by the external input.

In our model we extend this idea to sensorimotor coupled interactions. With the idea of a passively-coupled agent that is fed a copy of the input generated by a situated agent, we analysed how TSE complexity increases for different sets of variables of our system. We found that agents with synaptic plasticity increase the flows of mutual information coordinating different subsets and increasing the complexity of the agent when coupled to the world. This does not happen for agents without plasticity. According to the measure proposed by Tononi et al., both agents with and without synaptic plasticity present more complexity and information when you add together sensory and neurodynamic variables (we can extract this information by comparing the sum of TSE complexity of sets  $(CA, SA)$  plus  $(S)$  with that of  $(CA, SA, S)$  in Figures 4.4 and 4.6). It is probably not hard for nonlinear coupled systems to generate synergistic flows of information when two systems are in interaction. What is interesting is that the exact same input coupled to the same system produces less complexity when the directionality of the sensorimotor loop is disrupted (inputs are received but actions over the input are not performed). The issue here is not whether the response to an input is ‘passive’ or ‘active’ (as in Tononi et al.’s case), but whether the response is merely reactive or proactive. Not only going beyond the information of the external input, but overstep the agent-input interplay through an active coupling. We consider that this fundamental asymmetry of the system represents one key requirement for autonomous behaviour.

## PRECARIOUSNESS

Another fundamental aspect of autonomy is that of the precariousness of the system. Di Paolo and Thompson (2014) characterize an autonomous system as an operationally-closed and pre-

carious system. For them, operational closure is a property by which the enabling conditions of network of processes is woven in a way that all the processes in the operationally closed systems are enabled by other processes in the system. The precariousness of the system implies that in the absence of the enabling relations established by the operationally closed network, a process belonging to the network would stop or run down.

We find that the situated agent presents a form of coherence in the coordination of its scales in the form of a  $1/f$  dynamics in its neurodynamic and sensory variables. In *Aguilera et al. (2015)* we also observed that  $1/f$  patterns are robust to parametrical changes and actively maintained when the agent is coupled to the world. Interestingly, the span of the scale free spectrum coincides with the ranges of information transfer in Figure 3.16, where sensory patterns transfer information at fast timescales, neurodynamic patterns at slow timescales, and synaptic plastic mechanisms seem to transfer information at intermediate scales. Interestingly,  $1/f$  correlations spreads from fast to slow timescales for both sensory patterns and neurodynamical patterns, but do not appear for synaptic plasticity. It may be that synaptic plasticity mediates between fast changes of environmental inputs and higher order neurodynamic assemblies, allowing a temporal coherence in the structures of both of them displaying long-range correlations in multiple scales. In any case, our analysis suggest that  $1/f$  processes at sensory and neurodynamic variables are being maintained by a network of enabling relations of information transfer at different scales, in which a prominent role is being played by mechanisms of synaptic plasticity. As we depicted in Figure 3.17, some information transfer regulatory loops are removed or disrupted when the agent is passively-coupled to the world. In this case, we observe that  $1/f$  scaling and the width of the multifractal spectrum in the neurodynamic patterns of the agent is reduced and its temporal coherence decays. Similarly, if we directly remove the plasticity of the agent, fractal scaling of sensory patterns and the multifractal spectrum of both sensory and neurodynamic patterns are reduced. The coordinated processes of fractal scaling in the agent's components are precarious in the sense that they are maintained by a continuous sensorimotor interaction. Just feeding the right input is not enough to sustain them but the agent needs to maintain a delicate equilibrium in its interaction with its world. In the absence of the mechanisms that allow the agent an active regulation of its interaction with the environment (an agent without plastic mechanisms) or the normal functioning of the channels that allow the agent to create an asymmetrical agent-environment interaction with the world (a passively-coupled agent) the temporal structures generated in the autonomous sensorimotor regulation of the agent run down.

#### 4.6.2 A SITUATED DYNAMIC CORE

Our model not only provides a generative illustration of the generic intuition outlined by Silberstein and Chemero in which strongly nonlinear and softly assembled interactions allow the emergence of a brain-body-environment system of nested self-organization (*Silberstein and Chemero, 2012*). The results shown in this chapter provide interesting insights into the notion of a dynamic core as a necessary hub for agential autonomous processes. Usually a dy-

dynamic core is circumscribed to distributed clusters of neurons intensely interacting with each other within the brain. We find appealing the notion of a dynamic core composed not only of coupled neural dynamics, but extended to coupled neural and environmental reentrant processes: a *sensorimotor dynamic core*. Edelman and Tononi stress that a dynamic core is a process, not a thing or a place, and is defined in terms of neural interactions, rather than in terms of specific neural locations, connectivity or activity (Edelman and Tononi, 2001). Thus, there is no apparent reason to limit the definition of a dynamic core to brain-bound events. Our model shows instead, that the dynamic core might well be situated, cutting across brain-body-world divisions.

Moreover, our results point in the direction of two critical aspects of a situated dynamic core:

1. The situated nature of the operational closure of the system, which covers both internal dynamics and sensorimotor dynamics of agent-environment interaction.
2. The multiscale causal asymmetry of the agent-environment system, in which the informational complexity in the sensorimotor loop is enhanced respect a passively-coupled situation, generating a structure of enhanced information flows distributed among different spatial scales of the system.
3. A precarious operationally closed network of processes displaying long-range correlations sustained by a circular network of enabling processes, which runs down if this network is disrupted.

This portrays a ‘radically embodied’ (Thompson and Varela, 2001) view of the dynamic core hypothesis of cognition, in which specific neuronal assemblies underlay the operation of cognitive acts, depending crucially on the manner in which brain dynamics are embedded in the somatic and environmental context of the agent’s life, bringing forth the constitution of a sensorimotor cognitive subject which actively constitutes itself as a subject and acts over its world.

## Part III

# Interactive Autonomy in Social Processes



# 5

## Social Agency: interaction-dominant dynamics in social reciprocity

**ABSTRACT:** We argue that interactive sensorimotor constitution of identity can be extended to the social realm. We propose a minimal experimental framework to analyse the dynamics of a shared interactive space between two subjects. Experimental analysis shows that long-range correlations arise in the interactive space only in the case of genuine social interaction.

**CONTRIBUTIONS:**

- We perform an experimental analysis of a modified version of the Perceptual Crossing experiment mixing human players and programmed bots.
- We analyse long time correlations in individual and interactive variables, showing that 1/f patterns only arise in the latter in the case of human-human interaction.

**RELATED PUBLICATIONS:**

- Bedia, M. G., Aguilera, M., Gómez, T., Larralde, D. G., and Seron, F. (2014). Quantifying long-range correlations and 1/f patterns in a minimal experiment of social interaction. *Frontiers in Psychology*, 5:1281.

In Chapter 1 we identified two forms of autonomous organization, one based in the material self-production of an organism, and other based in the maintenance of a domain of interaction of an agent and its surroundings. For the latter, we explored two domains in which interactive forms of autonomy appear: sensorimotor neurocognitive agency and collective agency in the social domain. In this chapter and the next we move into the realm of social cognition with the tools from previous chapters exploring the relation between theoretical and analytical tools used to define autonomy at the level of cells or individual agents and collective entities composed of mutually interacting agents. In Chapter 6 we will analyse in detail a particular case of social autonomy, whereas in this chapter we want to explore the role of so-

cial interaction as a constitutive element of social cognition. What does this claim imply? Can something like a ‘social autonomy’ emerge from the sensorimotor interaction between social subjects? To explore this question we first analyse different psychological perspectives that bring interaction as an essential aspect for social processes in contrast with previous dominant frameworks. Then we propose a minimal experimental setup for studying the constitutive nature of interaction in social cognition in a particular case of social engagement based on the ‘perceptual crossing’ paradigm. We analyse the patterns of individual and shared variables searching for the emergence of multiscale correlations, showing that the space for social cognition emerges in a collective space of interaction instead of a purely individual one. We finish by proposing some ideas to tackle the problem of autonomy in large-scale networks of social interaction.

## 5.1 DIRECT PERCEPTION AND SOCIAL INTERACTION

Traditionally, much work on social cognition research has been guided by the idea that minds are composed of exclusively intracranial phenomena, and therefore perceptually inaccessible to others. Similarly to early computational approaches, in which researchers assumed that we have a poor perceptual access to the world (see section 2.1.2), dominant theories on social cognition assume that we have impoverished access to other people’s minds, and therefore social cognition must be built over some kind of cognitive machinery that supplements perception with inferences about the intentions behind our perception of what others are doing. Today, the two dominant theories in social cognition, ‘theory theory’ and ‘simulation theory’, state that something other than perception is necessary for our ability to understand others. Theory theory (Carruthers, 1996; Ratcliffe, 2006) proposes that our understanding of other persons depends on the practice of *mentalizing*, that is, employing folk psychology (i.e. a basic or ‘naïve’ theory of psychology) to infer the mental states of others. The ability to mentalize may be innate, but it depends on our observation and inference abilities. By observing others, we can gather data allowing us to revise our naïve theories. In contrast, simulation theory (Gallese and Goldman, 1998) posits that we do not need such theories about other’s minds, since we have a model that can be used to simulate other person’s mental states: our own brain. Observing others and their behaviour allows us to model and test their mental states as if we were in their situation.

Shaun Gallagher (Gallagher, 2008) has criticized both theory theory and simulation theory arguing that they assume that, when understanding others, (1) perception is playing a merely observational role, not taking into account the possibility of interaction or the effects of the subject actions, and (2) the perception involved seems to be greatly impoverished or ill-informed about other’s mental states, hence needing to be supplemented by some extra cognitive tools. In contrast, Gallagher proposes the notion of *direct perception* in social engagement to support the idea that we have direct perceptual access to others’ intentions, feelings, etc. For him,

in most of our encounters of our everyday life, direct perception delivers sufficient information for understanding others [...] In the broad range of normal circumstances there is already so much available in the person's movements, gestures, facial expressions, and so on, as well as in the pragmatic or social context, that I can grasp everything I need for understanding in what is perceptually available (Gallagher, 2008, p. 540).

This idea is supported by much experimental evidence showing that cognitive and emotional processes span across body and behaviour, thus making parts of some mental processes visible via their distribution across the visible body (Krueger, 2012). If mental states cut across brain, body and environment, as we saw in previous chapters, this implies that some cognitive and affective states can be accessed just perceiving behaviour. Gallagher also proposes a new interpretation for the role of mirror neurons in social perception. Mirror neurons (Rizzolatti and Craighero, 2004) are neurons in the pre-motor cortex, in Broca's area, and in the parietal cortex of the human brain that are activated both when a subject engages in specific instrumental actions and when she observes someone else engaging in those actions. Typically, mirror neurons are considered as evidence supporting simulation theory and considered to be explicit acts of simulation of others behaviours. In contrast, Gallagher argues that mirror neurons constitute sub-personal mirror resonance processes of which we are not explicitly aware. Thus, mirror neurons' motor resonance processes are important enactive processes that contribute to the constitution of the perceptual access that we have to the intentions of others (Gallagher, 2007).

Hanne De Jaegher et al expand these ideas and argue that perception does not directly deliver what we need for social interaction. Instead, they propose that we can experience directly other's feelings and intentions through skilful interaction with others. For them, interaction is constitutive of social understanding and direct perception (De Jaegher et al., 2010). They claim that social interaction processes produce a 'transparency' that makes perception in the social realm feel direct in the same way that the interaction of a blind person and her cane: when she is interacting with the cane and the environment, she does not perceive the cane, but directly the environment. Nevertheless, the key is that perception is not always direct, since the other's intentions are not always transparent to me. Instead, misunderstandings are a key ingredient of social processes. Loss and recovery of coordination in social interaction is what construct social understanding. Breakdowns and recoveries of transparency are where social perception can be 'enriched' the most and when a skillful social interaction can be developed. Thus, moments of transparency are constructed through a history of interaction and coordination. The faculty to mirror each others' movements, anticipate them and temporally synchronize or desynchronize, can create patterns of coordination that can influence the disposition of the individuals involved in the encounter to sustain, modify, or end it (Jaegher and Paolo, 2007). We can find examples of this in situations when an interaction continues despite the fact that none of the participants want it to continue. A familiar case would be when you encounter someone coming from the opposite direction on a narrow footpath and,

attempting to walk past each other, you repeatedly step towards the same side. This suggests that interpersonal coordination of movements can provide the interaction process a form of autonomy that transcends the individual level. Nevertheless this does not mean that individuals are subsumed passively into the interaction process. On the contrary, there is an interplay between individual autonomy and the autonomy of the interaction process itself. The same as sensorimotor interaction, direct or transparent perception of the world it is only achieved when there is a coordinated interaction within a coupled dynamical system.

Following advances in coordination dynamics, coordination of rhythmically moving limbs between two individuals exhibits the same phenomena predicted by the HKB model in the behaviour of individuals (Haken et al., 1985), such as the existence of two spontaneously stable modes of coordination (in-phase and anti-phase), transitions from anti-phase to in-phase at a critical movement frequency, critical fluctuations (i.e. an increase in coordination variability) preceding the transition, and hysteresis (i.e. a sensitivity to the history of the system) (Schmidt et al., 1990). The same coordination phenomena found in a person's mind and body can be extended to social interactions. Furthermore, synchronization between people has been found to arise spontaneously just as a result of visual information exchange (Oullier and Kelso, 2009b). Are these forms of coordinated interaction a constituent element of social cognition?

Although synchronization between subjects is a product of many forms of social interaction, interpersonal coordination may produce many higher order patterns (Riley et al., 2011). In this chapter we propose that fractal scaling may be one of such patterns indicating that interaction is constitutive of a social cognitive process and propose an experimental framework, in the form of a minimal social setting, to explore this issue. In order to compare genuine and decoupled social interaction, we consider how scaling of behaviour and interactions are distributed for the case of mutual interaction between two humans and disrupted interaction between a human and a software agent.

## 5.2 CONSTITUTIVE INTERACTION IN A MINIMAL SOCIAL SETTING

In Chapter 2 we analysed the role of interaction in cognitive activity using a simulation model. Here we analyse the role on interaction in a social process by analysing a popular minimal social setup: the perceptual crossing paradigm. Furthermore, in this chapter and the next we shift from modelling to experimental approaches in order to explore the possibilities of understanding social processes from the perspective of systemic autonomy. Although an experimental setting gives us less access to the variables of the system than a simulated one, we think that the study of social autonomy still requires experimental research in order to facilitate the design of good models. While in neuroscience there is a vast amount of experimental work from systemic perspectives and much of it addresses aspects related to autonomy (e.g. embodiment, intentionality, consciousness), as well as related models based in experimental evidence (e.g. neural ensembles, the dynamic core hypothesis, integrated information theory), in the social domain there is less consensus about the possibility of a systemic autonomy of groups

of individuals as a collective, and consequently there is less experimental evidence and related models exploring this issue.

Instead of a minimal robotic simulation, in this chapter we use a minimal experimental social setting involving the interaction of two subjects. In previous chapters we compared ‘genuine’ interaction with perturbed configuration as ‘passively-coupled agents’. Comparatively, here we have situations where two humans interact together, and situations where a human interacts with a software robot which does not establish a sensorimotor interaction with its partner.

### 5.2.1 DOUBLE MONITOR EXPERIMENT

Evidence of the constitutive role of interaction in social cognition has been greatly inspired by Lynne Murray and Colwyn Trevarthen’s double-monitor experiment (Murray and Trevarthen, 1985), which showed the role of interaction in social exchange of a mother and her baby through closed circuit TV. In the experiment a 6- to 12-week-olds baby interacted normally with her mother, but at some point the video of the mother was replaced by a replay of the mother’s behaviour during a previous period of live interaction, making the baby upset. The experiment showed that a baby can detect that the mother is not interacting with him when mother’s recorded motions are displayed, and thus does not experience a process of social engagement. The double monitor experiment and its successors (e.g. Nadel et al., 1999; Trevarthen, 1994) suggest that individuals do not become involved in a social process just by receiving an external input from another subject, but this input needs to be coordinated with the individual for it to make sense. In the experiment, it is not enough to perceive the mother’s face and its expressiveness to become socially engaged. Instead, it seems that the ongoingness and contingency of the interaction is a key factor for the infant to recognize a ‘genuine’ interaction. Nevertheless, the experiments should not be necessarily interpreted as indicating the constitutive role of the social process. It may be just an enabling process if ‘recognition’ is performed at a purely individual level. For investigating this, we need a different framework where we can easily analyse processes happening not only at the individual level but also in the collective interactive domain.

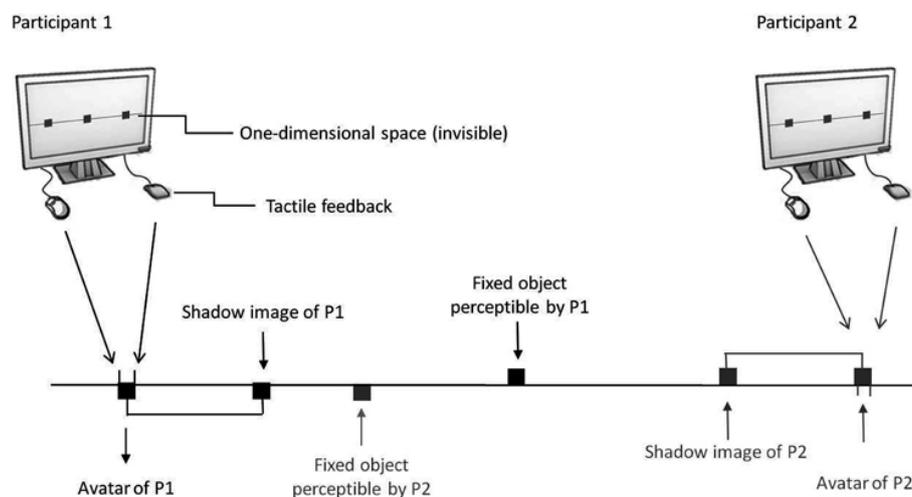
### 5.2.2 THE PERCEPTUAL CROSSING PARADIGM

In recent years, researchers in social cognition have found the ‘perceptual crossing paradigm’ to be both a theoretical and practical advance towards meeting particular challenges. This paradigm has been used to analyse the type of interactive processes that emerge in minimal interactions and it has allowed progress towards understanding of the principles of social cognition processes and the mechanisms that give support to minimal social capabilities (Auvray et al., 2009). This experimental framework is a way to study dyadic interactions and to analyse the perception of someone else’s agency in different situations implemented in a minimal virtual world. Through the self-organized collective patterns that emerge from the interac-

tions (like emergent coordination, turn-taking, etc.), hypotheses about the human capacity for social cognition can be extracted.

Perceptual crossing paradigm constitutes a simple framework for studying social interactions in its simpler form. It consists of a minimal scenario in which two participants, sitting in different rooms, interact with each other by moving a sensor along a shared virtual one-dimensional space using a computer mouse 5.1. The subjects are only allowed to move laterally in a one-dimensional world and perceive the collisions with other human subjects or artificial agents. In the last few years, the perceptual crossing paradigm has become a promising experimental tool for the analysis of dynamic interactions of human social processes. A more detailed analysis leads to two types of experiments: (1) behavioural experimental research and (2) simulated agent modelling. Relating to the former, numerous experiments where real subjects try to identify each other in a virtual world have been carried out, and researchers have analysed the type of behaviours that seem to offer support for social coordination patterns; both in one-dimensional experiments (Auvray et al., 2009) and also in their extensions to two dimensions (Lenay et al., 2011). In some cases, real experiments and phenomena previously tested in simulations were combined, for example in (Iizuka et al., 2009; Iizuka and Ando, 2012), where the authors explored how participants modulated their interaction dynamics to figure out if an interaction was genuine or not. Regarding the latter, i.e. focusing now on computational modelling context, virtual agents have been evolved to locate others in an experimental setup analogous to that used in Auvray's version (Auvray et al., 2009), providing a mathematical analysis that explained how virtual agents managed their own variables, such as size or velocity, to coordinate with others in an extremely robust way (Iizuka and Ando, 2012; Iizuka and Paolo, 2007). Simulation models to build 'social software agents' have demonstrated that this kind of behaviour can emerge from very simple structures without explicit social reasoning capabilities (Froese et al., 2014a). In general, the studies on simulation models complimented the experimental work with humans, sometimes providing proofs of concept, making empirical hypothesis more precise or simply understanding in more detail the possible complex patterns and the dynamical properties of the task.

In this chapter, we analyse whether some critical aspects of these interactions could not have been observed by previous studies. We consider alternative indicators that could complete, or even lead us to rethink, the current interpretation of the results obtained from both experimental and simulated modelling approaches in the fields of social interactions and minimal perceptual crossing. We aim to characterize the individual and interactive domains of interaction for different conditions, trying to depict how a space of social engagement is constituted. In particular, we discuss the possibility that previous experiments have been analytically constrained to a short-term dynamic type of player response. Additionally, we propose the possibility of considering these experiments from a more suitable framework based on the use and analysis of long-range correlations and fractal dynamics. If social interaction is to display autonomous properties it is tempting to explore whether measurement techniques that turned out relevant in the domain of sensorimotor autonomy can also be applied suc-



**Figure 5.1: Perceptual crossing framework.** Schematic illustration of Auvray et al. (2009) experimental set-up. Extracted from Auvray and Rohde (2012).

cessfully to measure social autonomy. We also reveal evidence supporting the idea that social interactions are deployed along many scales of activity. Specifically, we propose that the fractal structure of the interactions could be a more adequate framework to understand the type of social interaction patterns generated in a social engagement.

### 5.3 THEORETICAL FRAMEWORK

Studies of the perceptual crossing experiment have provided insightful evidence about the importance of inter-individual coordination for the emergence of social cognition and agency detection. However, we think that still more advances are needed in order to characterize and better understand how coordinated interactions may give rise to collective social processes. Recently, some authors have emphasized the importance of understanding how distinct time scales and organizational levels are intertwined for the emergence of social cognition (Dumas et al., 2014). At the neural level, there is experimental evidence of the importance of non-linear cross-scale interactions for brain organization (Le Van Quyen, 2011), and in social neuroscience, inter-brain synchronization in multiple frequency bands has been found during imitation of hand movements (Dumas et al., 2010) or synchronization patterns are found during guitar improvisation showing a complex interplay of different frequencies (Müller et al., 2013). It is still missing, to our knowledge, a detailed analysis of this kind of phenomena at a behavioural level. These examples show the potential of a multi-scale account of social cognition and lead us to think that sometimes the analyses developed so far to understand perceptual crossing dynamics may fall short in their ability to characterize the emergent multi-scale nature of social interaction.

A common feature exists in the way in which we deal with the experimental results ob-

tained in previous analyses of behaviour in the perceptual crossing experiment (considering behavioural modelling experiments and simulated agents): the participants' behaviour is analysed only in a short-time scale (this point is explained in detail in the next section). For example, Auvray et al. (2009) analyse the probability of clicking in a two second window after the subject encounters another subject or object, together with standard statistical variables of some system variables (frequency of crossings, correlation between velocity and acceleration, etc.). This type of analysis is implicitly assuming that the emergence of social engagement can be reduced to a scale of short-term activity and that there is no influence of other scales or any inter-scale correlations that are relevant for the subject's behaviour (for example, assuming that there is no interference between the previous collisions of the subject with different kinds of agents and the decision of clicking or not clicking). A similar assumption is also found in the agent modelling field, for example, Di Paolo et al. (2008) focus on finding what kind of short term dynamics in their simulated model (modelled in terms of delays between the perceptual stimulation of the agent and its motor response) is able to create the stable pattern of social interaction that allows the emergence of a dynamic co-regulation. Again, inter-scale correlations in the social interaction process are left out of the analysis and modelling.

In this chapter we propose a quantitative indicator that works as a complementary measure of the analysis addressed in previous perceptual crossing experiments; an indicator that consists of characterizing the cross-scale nature of the interaction through fractal and multifractal analysis (Van Orden et al., 2003) of the collective dynamics. Despite its apparent simplicity, we propose that the perceptual crossing paradigm could comprise several embedded levels of dynamic interaction, resulting in correlations of the signals over different time scales. In section 1.2.3 we explained the concept of self-organized criticality, used to understand the behaviour of a robotic agent in Chapter 4. For a multi-scale approach to social cognition, SOC is appealing because it allows us to imagine systems that are able to self-regulate coordinated behaviour at different scales in a distributed manner and without a central controller. We analyse whether coupling between the players of the perceptual crossing experiment may be operating under a regime of SOC and display  $1/f$  scaling patterns.

### 5.3.1 OUTLINE

In this chapter we try to explore the presence and relevance of multiple scale and inter-scale or long-range correlations in the perceptual crossing experiment. If social interaction is genuinely constituted by the interaction process itself (and not by the inferential capacity of the interacting subjects) social interaction should display long-range correlations and coordinated intermittency in the form of  $1/f$  scaling and a multifractal spectrum. Moreover, multi-scale interactions should be present in collective variables and not only in individual variables, as an indicator of an emergence of a social domain of interaction.

We propose a modified version of the original perceptual crossing experiment, in which the player only faces one opponent, which may be another human player or a programmed agent with two possible kinds of behaviour: a simple oscillatory behaviour or a 'shadow'

behaviour that repeats the movement of the player. More information is given in the next section. The comparison between genuine social human-human interaction and disrupted human-bot interaction is used to test the analysis performed. Thus, in our experimental setup we have different kinds of social interaction: humans recognizing each other as such, humans interacting with programmed agents with artificial behaviour, humans failing to recognize other humans or bots tricking humans. Can we characterize when genuine social interaction emerges? And if so, where does it lie?

Auvray et al. (2009) propose that the sensitivity for recognizing other intentional subjects, instead of being perceived by each of the participants, arises from the dynamics of the interaction itself. In their experiment, the distribution of clicks suggested that social recognition arose from a combination of (1) the ability to discriminate between mobile (human player, shadow) and immobile objects and (2) the stability of mutual interaction patterns between two human partners or between human and an immobile object. This interpretation was inspired by the results in a simulated model which showed the importance of the stability of coordinated behaviour (Di Paolo et al., 2008). However, we think that further evidence supporting the claim that social recognition emerges from interaction dynamics instead of individual sensitivity is necessary. In fact, the model presented by Di Paolo et al. (2008) could be interpreted as showing that relatively simple reactive behaviours could account for a click distribution in which agents appear to ‘recognize’ each other, without a genuine, underlying process of social recognition. We propose that genuine social interaction should arise from the emergence of a complex web of interactions across different timescales between the activity of different agents. For a first approach to support this claim we propose the following schema:

1. Since we consider that inter-scale dynamics might be relevant to characterize perceptual crossing dynamics, we perform similar measures to previous works in perceptual crossing experiments, and explore the existence of a link between ours and previous results, and cross-scale interaction dynamics (section 5.5).
2. We propose that if genuine social interaction is based on cross-scale interactions, a fractal distribution should be present in collective variables of the social process. We propose the difference in the movement of the two players (using the difference between their speeds) as a candidate variable and perform fractal and multifractal analysis of the distribution in the individual rounds of the game, finding a clear  $1/f$  and multifractal spectrum only when two human players interact (section 5.6).
3. Finally, we suggest that as opposed to collective variables, the fractal structure of the individual dynamics of the player or their opponent alone should not be discriminative for the type of ongoing interaction. We analyse this issue repeating fractal and multifractal measures on the movement of the player and the movement of the opponent (using their individual speeds) and conducting linear mixed effects models to assess if the different variables analysed (difference of speeds, speed of the player and speed of the opponent) can discriminate between the type of interaction, finding that only the

collective variable of the relative speeds can discriminate the two types of programmed agents from genuine human interaction (section 5.6.1).

#### 5.4 EXPERIMENTAL SETUP AND METHODS OF ANALYSIS

In this experiment, human participants were assigned to computers for them to interact in pairs, within a shared perceptual space, where some opponents were other human participants and some opponents were computerized agents (bots). Participants were unaware of the nature of their opponents.

Our intention was not to replicate Auvray's experiment, where each participant simultaneously encounters a human partner, a mobile agent (which acts as a shadow of the human partner, repeating the same movements but shifted in space) and a static one (Auvray et al., 2009). In our case, each participant received only a single stimulus in one of the following scenarios: human vs. human, human vs. 'oscillatory agent' and human vs. 'shadow agent'. The 'oscillatory agent' was programmed to deploy an oscillatory behaviour (describing a sinusoidal trajectory of 0.5 Hz and 200 pixels of amplitude), predictable and deterministic. The 'shadow agent' was able to show an irregular pattern because it consists of a 'shadow image' of the participant's movement, i.e, a bot that generates a movement strictly identical to the participant trajectory but delayed 400 ms. in time and 125 pixels in space. Participants were instructed to try to detect whether their opponent was human or not.

When participants arrived at the laboratory they were randomly assigned to a workstation and were provided with headphones. They were informed that the study involved two parts, each independent from the other and that the first one -training stage- would take approximately 3 minutes and the second one -evaluation stage- a further 10 minutes. In order to guarantee confidentiality during the study, identification codes/nicknames were chosen by the participants. Throughout the experiment, participants were provided with verbal instructions regarding the structure of the experiment and their sections.

In the training stage, the participants were informed that it was a simple 'proof of concept' stage and that the purpose was only to learn how the platform worked. Participants were free to move the mouse as they pleased during three sessions of 1 minute each with a short break between them. They played consecutively against three bots of increasing difficulty in the interaction: a static bot, a bot moving at a constant low speed and a bot moving at a constant medium speed.

After that, they were informed of the aim and rules of the evaluation part of the experiment. The experiment consisted of 10 sessions of 40 seconds each. In each session: (1) participants were randomly assigned an opponent (human-human or human-bot) to explore the virtual space; (2) participants were asked to move their mouses in order to detect the movement of their assigned opponents, (3) after each session, participants were asked to make a choice between the two options displayed on the screen in order to guess whether their opponent was a human or a bot, and (4), finally, participants were informed on the screen whether they had guessed successfully or not. After the 10 sessions were completed, the experiment

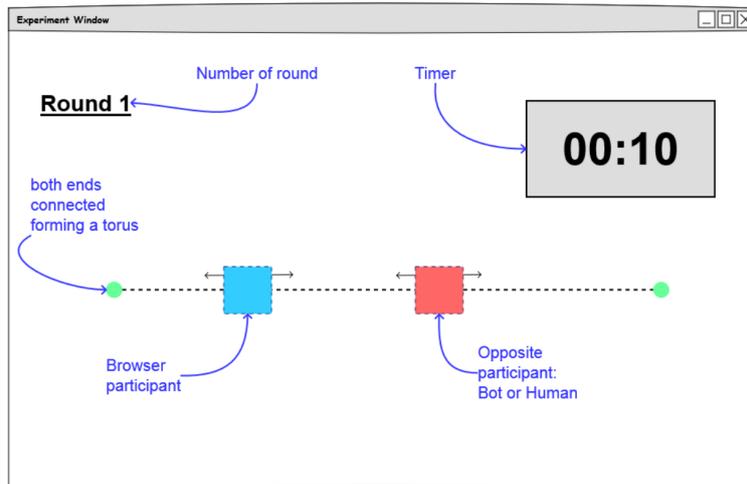


Figure 5.2: Experimental setup. Extracted from Bedia et al. (2014).

was declared finished.

A total of 13 participants (8 females and 5 males) took part in this experiment. Their ages ranged from 16 to 19 years. The final dataset used in the analysis comprises a total of 106 samples of the cursor positions over time of each participant recorded with a sampling period of 1ms.

In order to quantify the fractal-like autocorrelation properties of each of the time series generated, a detrended fluctuation analysis (DFA) algorithm (Peng et al., 2000) was used to estimate the scaling exponent in every fractal series, characterizing long memory dependence phenomena. As we did not assume that the processes were normally distributed, calculations were completed through multifractal detrended fluctuation analysis (MF DFA) (Ihlen and Vereijken, 2010). Both methods are described in section 4.5.1.

DFA bins for parameter  $n$  have been defined logarithmically from 64ms to 10s with an interval of  $2^{0.01}$ s. For the MF DFA we have used the same values for the  $n$  bins and we have taken a value of  $q$  with values from  $-3$  to  $3$  with intervals of  $0.25$ .

## STATISTICAL APPROACH

The design of this experiment involved repeated measures per subject and, in order to account for this characteristic, linear mixed effect models were computed. Analysis of the fractal study data included descriptive statistics and mixed effects models of the relationships between different measurements. These models allowed us to account for the variation between players and the residual variation between the different games played by each respondent. Models were developed using the *nlme* package in the open-source statistical software R<sup>1</sup>. In a nutshell, mixed effect models are regression models that incorporate both fixed and random ef-

<sup>1</sup><http://cran.r-project.org/web/packages/nlme/index.html>

fects. Fixed effects are the independent variables of interest while random effects replicate the structure of the data (i.e. games within player in this case). As a consequence, the unexplained variation can be split into the variation between players and the residual variation between games within players. In this experimental design, the variable ‘type of opponent’ (‘human’, ‘shadow agent’ or ‘oscillatory agent’) acts as the only fixed effect. Each player performs the experiment several times, so we include the variable ‘player’ in order to account for the potential lack of independence of the repeated measures for each participant.

## 5.5 PRELIMINARY ANALYSIS: SHORT-SCALE VS. MULTISCALE DESCRIPTIONS

Above, we proposed that analysis in the perceptual crossing experiments should not be restricted to events with a short temporal span. Can we sustain this claim? Does the distribution of social interaction present a multiscale structure? Before characterizing different forms of engagement in our experimental setup, in this section we perform different tests to explore the possibility of multi-scale interactions shaping the dynamics within the perceptual crossing experiment. We start by analysing our results with similar measures to the ones used in previous analysis and propose the necessity of complementing them with other measures that are not constrained to one particular scale of behaviour.

### FIRST RESULTS

As an example the of use of short span variables of analysis, in (Auvray et al., 2009) the two variables used in order to explain the detection of another human player are the frequency of stimulation (the number of times a player receives an input from its opponent) and the probability of clicking (the probability of the player clicking their mouse in a two second interval after a stimulation). The setup in our task differs in that the players are not asked to click when they recognize a human player. Thus, in order to test a similar measure (note that participants were not asked to click in our experiment) we substitute the probability of clicking with the probability of having a new stimulation in an interval defined as a given window after a previous stimulation. This measure is intended to capture the probability of engagement in an ongoing interaction between the two players.

Unlike Auvray et al. (2009), we do not use a single value for the window length and instead test the values 0.25, 0.5, 1 and 2 seconds (around 95% of stimulations happen within a window of 2 s after the previous stimulation). We refer to the frequency of stimulation as  $F_s$  and the probability of consecutive stimulations in a window of length  $L$  seconds as  $P_s^L$ .

We conduct linear mixed effects modelling of the series corresponding to each measure and we obtain the results shown in Table 5.1. Here we show the p-value resulting from the comparison of the distributions corresponding to players when playing against another human player and when playing against each type of bot .

We can observe in the table how the frequency of stimulation  $F_s$  does not discriminate against different types of players. This result is different from the classical perceptual crossing results, and may be caused by the fact that the participants play individually against each type

Groups	p-value				
	$F_s$	$p_s^{0.25}$	$p_s^{0.5}$	$p_s^1$	$p_s^2$
human-human vs. human-oscillatory	0.2381	0.0000	0.0000	0.0000	0.0496
human-human vs. human-shadow	0.6591	0.6159	0.0000	0.2455	0.0519

**Table 5.1: Statistical analysis of short-scale variables.** Results of the linear mixed-model effects for comparing stimulation frequency  $F_s$  and probability of consecutive stimulations  $P_s^L$  between the rounds where the player was facing other human player and the two cases of programmed agents (oscillatory and shadow agents). Only some particular windows allow us to statistically discriminate human versus bot opponents.

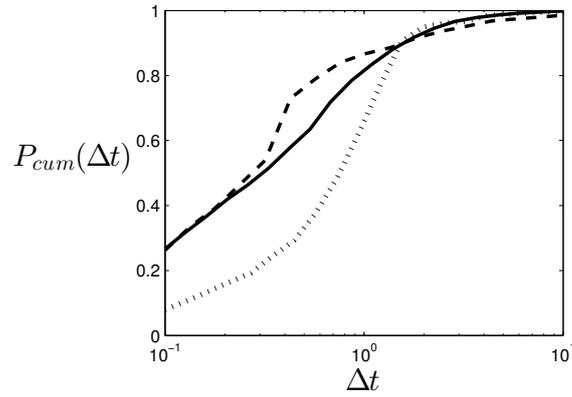
of opponent. For the probability of consecutive stimulations  $P_s^L$  we observe that the result depends largely on the chosen value of  $L$ . For example, for the extreme values of  $0.25\text{ s}$  and  $2\text{ s}$  the p-value is quite high and we cannot discriminate human opponents from either of the two bots (e.g. setting the statistical significance level at  $5\%$ ). Oscillator opponents however can be discriminated for windows of  $0.5\text{ s}$  and  $1\text{ s}$ , and shadow opponents can only be discriminated from humans for windows of  $0.5\text{ s}$ . Then, choosing a value of  $0.5\text{ s}$  would give us a variable that allows us to statistically differentiate the different players, showing that at this particular scale some opponents have more consecutive stimulations with the player than others (in this case, the shadow agent presents a higher probability of consecutive stimulations).

To assess the significance of the statistical results without the bias of choosing particular windows of analysis, we proceed to compute the distribution of inter-stimulation intervals  $\Delta t$ , that is, the time between one stimulation and the next. However, since the data for each player and round are scarce (around 40 mean stimulations by game, depending on the type of agent), we aggregate the data of different players and rounds (although this could entail losing some information about the data structure). The result of the cumulative probability can be observed in Figure 5.3, where we observe that the windows of discrimination in Table 5.1 roughly coincide with the intervals in which the cumulative density functions overlap. This may indicate that what we are doing when we just take the probability of stimulation (or clicking) is integrating the density distribution of a process that unfolds over different scales (in our case ranging from  $0.1\text{ s}$  to  $10\text{ s}$ ).

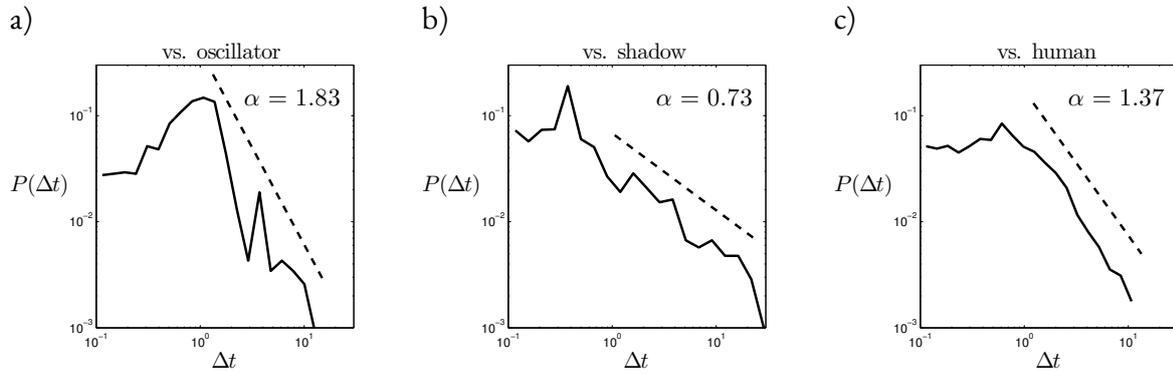
Our hypothesis is that genuine social engagement displays a scale free structure of sensorimotor interactions. Thus, difference between human-human interaction and other modes of disrupted interaction should be reflected in the distribution between scales, and not in any specific scales. So, although we might find that events at a particular scale allow us to statistically differentiate between different behaviours, this would be an epiphenomenon of a deeper change of the inter-scale structure of the interaction process.

### 5.5.1 CHARACTERIZING A MULTISCALE STRUCTURE OF SOCIAL INTERACTION

To test this idea, we should test whether the results for a particular window are discriminative between agents are either the consequence of something relevant happening at that timescale,



**Figure 5.3: Cumulative probability density function of the time between collisions for different types of opponents aggregated among participants and trials.** Values for the regions illustrated are: (dotted line) human vs. oscillatory agent, (dashed line) human vs. shadow agent, (solid line) both participants are human players.



**Figure 5.4: Probability density function of the time between stimulations for different types of opponents aggregated among participants and trials.** Values for the regions illustrated are: (left) human vs. oscillatory agent ('vs. oscillator'), (middle) human vs. shadow agent ('vs. shadow'), (right) both participants are human players ('vs. human').

or are instead caused by the different underlying structures of the temporal density distributions. In order to shed some light on this question we have represented the aggregated density distribution functions of the time between stimulations  $\Delta t$  for the three types of opponents (Figure 5.4). In the figure we can observe the presence of long tails that start around 0.5 s in the case of the shadow and human opponents, and that these long tails have different slopes in a logarithmic plot. This might be indicating that the statistically significant differences in the activity between the different 0.5 s windows are not the result of something happening at that scale, but the product of a deeper change in the temporal structure of the interaction. In that case the statistical difference at windows of 0.5 s may appear because we are integrating along all the smaller timescales.

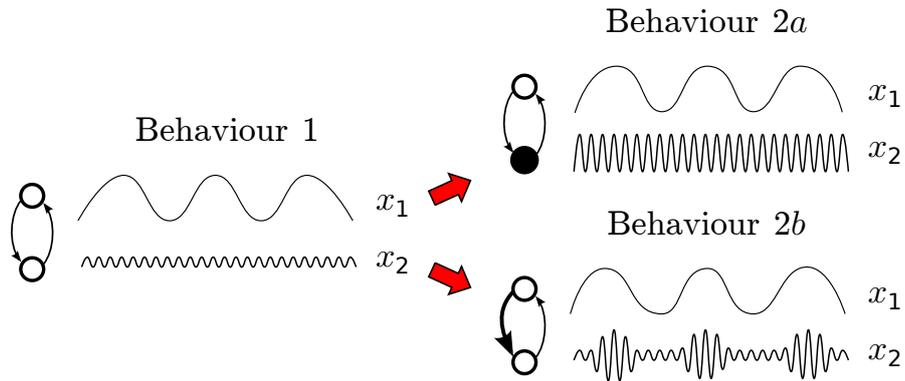
To illustrate this point we offer the following example (Figure 5.5): imagine that we have

a system in which we can access two components  $x_1$  and  $x_2$ , each one being active at a different timescale. The same system may display different behaviours. We can imagine that stimulating  $x_2$  the system switches from behaviour 1 to behaviour 2a. As a result of the behaviour change, we can find statistical differences between the distributions of  $x_2$  in behaviour 1 and 2a respectively. Also, we can consider a different condition in which we enhance the influence of variable  $x_1$  over  $x_2$  (in a process of phase modulation), making the system switch from behaviour 1 to behaviour 2b. Again, we find statistical differences between the distributions of  $x_2$  in behaviour 1 and 2b. The important point is that, while in the first case the statistical distribution of  $x_2$  is provoked by a direct change in the activity of this variable (directly stimulating the component that produces it), in the second case the statistical difference in  $x_2$  can only be explained by a change in the interaction between variables  $x_1$  and  $x_2$ . Similarly, significant statistical changes in a timescale of 0.5 s, might be the result of something relevant happening at that scale, or it may be the result of a reconfiguration of the whole temporal structure relating to different scales of behaviour.

The example in Figure 5.5 indicates that by analysing just one particular scale of the system we may be failing to capture the causes of a change in the system's behaviour even in the case that we were able to find a statistical discrimination of the distribution of a variable. In the case of the perceptual crossing, we propose that previous analysis may be extended with assessment of the activity at different scales and the relation between these scales. We contend that taking into account the changes in the temporal structure of inter-stimulation times allows a fuller explanation of the statistical discrimination offered by simple indices such as the number of clicks or consecutive stimulations within a given window (used by Auvray et al., 2009). Nevertheless, the analysis of the density distribution of aggregated data is too coarse to test this claim. We need to perform a more detailed analysis of the temporal structure within the individual interaction dynamics in each round to provide more conclusive results. We propose that statistical analysis of fractal and multifractal time series may be a better suited tool for this problem.

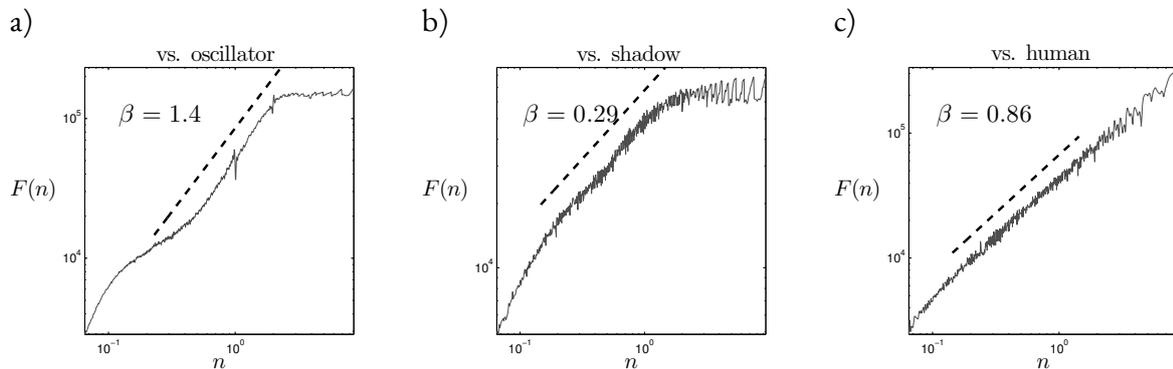
## 5.6 FRACTAL DYNAMICS IN THE STRUCTURE OF SOCIAL INTERACTION

So far, we have observed that the interaction process in our perceptual crossing experiment presents a multiscale structure that appears to change for different types of opponents. Is this multiscale structure just the result of individual sensorimotor behaviour? Or is it the result of the dynamics of a shared interactive space constructed between the two participants? In this section we seek a more detailed analysis of the temporal structure of the interaction between the two players for the three kinds of opponent. In doing so, we need to extract the movements of the two players. In order to analyse the interaction between the subjects, we take the time series of the distance between the two players (or the player and the bot agent): (1) the first derivative of the distance is computed in order to obtain the variations in the distance (whether the players are approaching or distancing themselves at each moment of time given that we are interested in the coordinated movements of the players, not their positions); (2) we



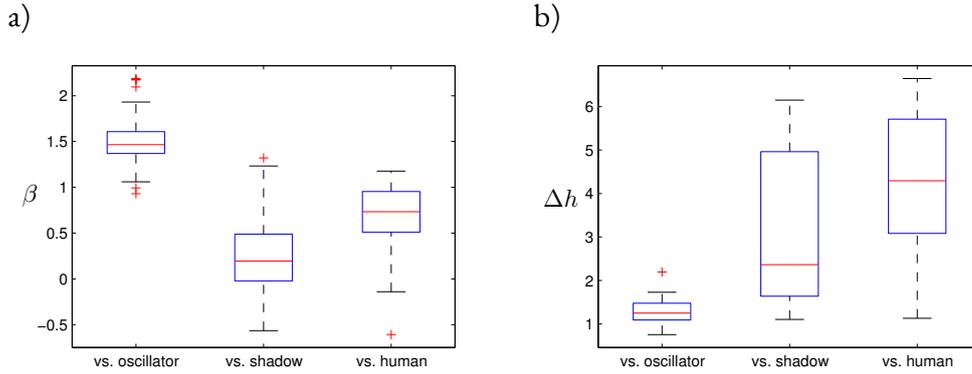
**Figure 5.5: Example of statistical comparison between two multiscale systems.** In case 2a the behaviour is statistically different from 1 at scale  $s_2$  because the intrinsic levels of activity at this scale have been increased. However, in case 2b the statistical differences in respect to 1 at scale  $x_2$  is not due to any intrinsic change in  $x_2$  but instead to a change in the relation between  $x_2$  and  $x_1$ , that now presents a phase modulation from slow to fast frequencies.

use the DFA and MDDFA algorithms to compute the structure of correlations across scales in the data series and (3) we perform a linear mixed-effects modelling in order to observe if the DFA and MDFA exponents are capable of differentiating between the interaction dynamics depending on the type of opponent the player is facing (oscillatory, shadow or human).



**Figure 5.6: Fractal analysis calculated on interactive patterns between two participants.** Values for the regions illustrated are: (left) human vs. oscillatory agent ('vs. oscillator'), (middle) human vs. shadow agent ('vs. shadow'), (right) both participants are human players ('vs. human'). Log-linear slopes are computed in the interval of one decade to the left of the higher scale cut-off point. If there is not cut-off point, the slope is computed in the interval  $n \in [10^{-0.5}, 10^{0.5}]$ . The examples are representative cases of the three kinds of populations in the experiment.

As a first step in the analysis, we observe the results of individual DFAs in different rounds. In Figure 5.6 we show some representative examples of the types of temporal structures displayed by the interactions with each type of agent. Since the slope of the fluctuations in a logarithmic plot is not always linear for all scales, we check if there is any cut-off value in which



**Figure 5.7: Fractal and multifractal exponents of relative velocity between participants time series.** a) Boxplots distribution of  $\beta$  and, b) width of the multifractal spectrum  $\Delta h$  in the time series of the relative velocity between participants. Values illustrated refer to interactions between: a human and an oscillatory agent ('vs. oscillator'), a human and a shadow agent ('vs. shadow') and two human participants ('vs. human').

the linear relationship is truncated. We do this by searching for negative peaks in the second derivative of  $F(n)$ . The search of cut-off values is only performed in the right half of the  $n$  axis, in order to find only the cut-offs at larger scales. Once the cut-off value is found, we analyse the slope  $F(n)$  for the values of  $n$  in the decade just below the cut-off value (e.g. Figure 5.6.a and 5.6.b). In the cases where there is no cut-off value (as in Figure 5.6.c) we analyse the interval  $n \in [10^{-0.5}, 10^{0.5}]$ .

For the oscillatory agent, we can observe in Figure 5.6.a a flatland at higher values of  $n$ , followed by a steep linear slope with a  $\beta$  parameter around 1.5. For lower values of  $n$  the linear slope disappears. This kind of fluctuation is characteristic of oscillatory dynamics, with the transition from flat to slope being equal to the period of the oscillations. In the other case, for the shadow agent, Figure 5.6.b presents something similar to a linear slope in the middle of the fluctuation spectrum, although the slope linearity breaks at the extremes. The slope of the fluctuation gives an exponent somewhat higher than  $\beta = 0$ . This suggests that weak short-range correlations exist (close to a white noise structure), but they do not hold for longer timescales. Finally, in Figure 5.6.c, when a player faces another human player, the fluctuation spectrum displays a linear slope with a  $\beta$  exponent close to a pink noise spectrum ( $\beta = 1$ ). In a large part of the series, the fractal slope reaches the largest timescales, showing that the correlations of the interaction dynamics cover a wide range of the spectrum. In Figure 5.6.c, fractal relations covering the hole spectrum are illustrated, although there are many other cases which present a cut-off point at large scales breaking the linear relation. We propose that the existence of fractal  $1/f$  patterns covering the whole analysed spectrum just in some cases of human-human interaction may be related to the fact that sometimes interaction is successful during a specific round but at other moments it experiences a breakdown in the interaction, leading to disruption in correlation at longer timescales.

Figure 5.6 shows three particularly representative examples of the three kinds of populations in the experiment. Particularly, in Figure 5.7.a we can observe the boxplots of  $\beta$  for the different types of interaction. When the opponent is the oscillatory agent, we find that the values of  $\beta$  in the time series are around 1.5. This means that the interactions are closer to a brown noise structure, signifying that the interaction is more rigid and structured than in the other cases. This makes sense since the movement of the oscillatory agent is constraining the interactions into its cyclic movement structure. On the other hand, when the opponent is the shadow agent, we have the opposite situation in which the interaction dynamics tend to display values of  $\beta$  greater but close to 0. This means that the history of interaction is more uncorrelated. Using a linear mixed-effects model we assess that indeed the three distributions of  $\beta$  are different from each other. We tested this idea appropriately using linear mixed-effects models of the three types of opponents (oscillatory agent, shadow agent and human) to assess the presence of statistically significant differences between the density distributions of  $\beta$ . Using a linear mixed-effects model we can test that beta is a significant parameter for distinguishing the different kinds of interactions depending on the type of opponent [ $F(2, 93) = 258.350$ ,  $p < 0.0001$ ].

As we have mentioned in section 4.5.1, fractal analysis is a mathematical procedure to determine scale invariant structures in a dataset. Monofractal signals have the same scaling properties throughout the entire signal, therefore, it can be indexed by a single global exponent (that is known as the Hurst exponent, see section 3.2). Alternatively, when spatial and temporal variations in a scale invariant structure appear, we get a multifractal structure that can be decomposed into different subsets characterized by different local Hurst exponents (denoted as  $h$ ) which quantify the local scaling of the time series. With this collection of exponents, we characterize their scaling properties: any deviation from the average fractal structure for segments with large and small fluctuations is captured by the multifractal spectrum width, denoted by  $D(h)$ .

In order to verify the nonlinear intermittent structure of the processes behind the patterns analysed above, we also analyse the width of the multifractal spectrum of the derivative of the distance between players. For each case, we calculate the width of the multifractal spectrum using the MFDFA algorithm and plot the distributions of the obtained values depending on the type of opponent (Figure 5.6.b). The probability distribution of the multifractal spectrum width  $\Delta h$  on the oscillatory agent is more concentrated around small widths, indicating little interaction between the time-scale of the oscillation frequency of the agent and the time-scales of the movement of its human opponent. Larger values on the distribution of the shadow agent indicate stronger interaction between its time-scales. Finally, the distribution of the human agent reaches the largest values of the multifractal spectrum width, suggesting a rich time-scale dynamics prompted by the interactivity between the time-scales of the movements of a pair of human opponents. Again, a linear mixed-effects model shows us that the distributions of values of  $\Delta h$  are different depending on the type of opponent [ $F(2, 93) = 258.350$ ,  $p < 0.0001$ ].

The fractal and multifractal spectrum results show that the relative velocity of the player with respect to their opponent in the interaction process displays different distributions depending on whether a genuine social interaction is happening or whether the player is interacting with an artificial agent with trivial (oscillatory) or complex (shadow) patterns of movement. It is interesting that  $1/f$  noise emerges for a collective variable (the derivative of the distance) only in the case of human-human interaction, suggesting that long-range correlations emerge in the shared space of social interactions and genuine social interaction is characterized by the collective evolution of the dyadic exchange. In those cases where the interaction between the players is too rigid or too weak, the emergent multiscale phenomenon disappears. The multifractal analysis seems to support this claim. To further test this proposal and determine if the same results can be obtained from non-collective variables, we compare these results with the behaviour of individual variables of the player and their opponents.

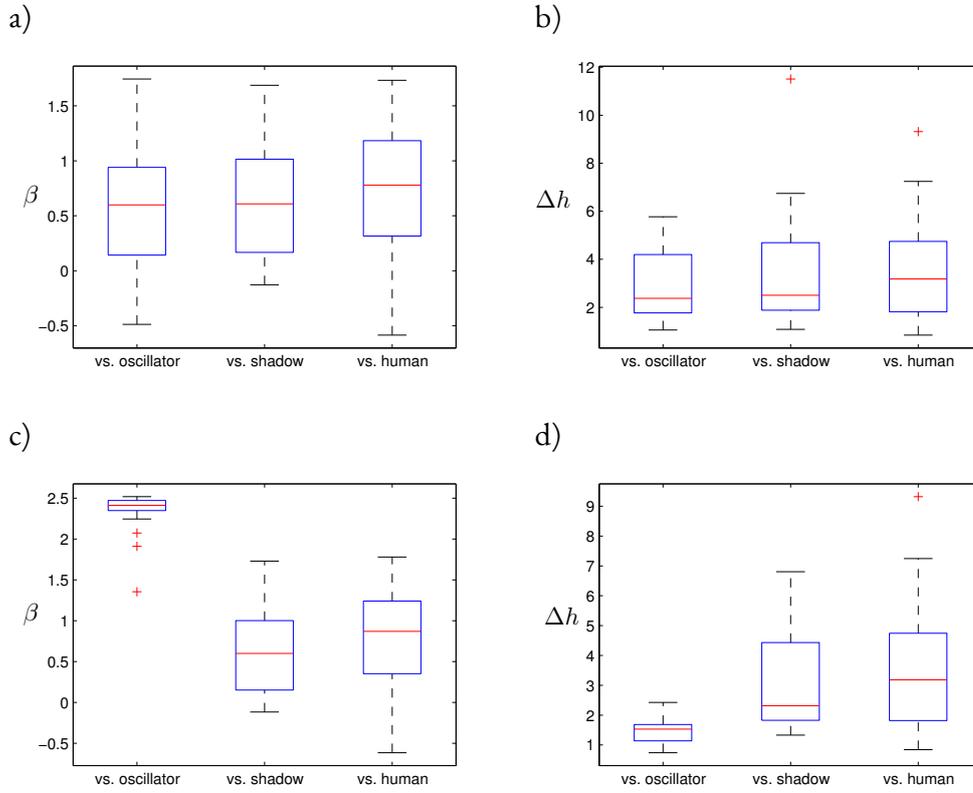
### 5.6.1 COMPARING FRACTAL EXPONENTS OF INDIVIDUAL AND COLLECTIVE VARIABLES

One of the ideas behind much of the work in the perceptual crossing paradigm is that the interaction between subjects is a constitutive element of social cognition (Auvray et al., 2009). If that is true, the characteristics of a genuine social interaction should appear only in dyadic variables such as the relative velocity between subjects and should be absent in individual variables such as the individual movement of the player or their opponent. In order to test this hypothesis, we repeat the fractal and multifractal analysis above using the velocity of the player and the velocity of their opponent, instead of the relative velocity between the two. Thus, we can test if the differences in the fractal emergent structure takes place in the shared space of social interaction or are instead phenomena that may be accounted for by the changes in individual dynamics alone.

In Figure 5.8 we can see how in this case the boxplots of  $\beta$  and the multifractal spectrum width  $\Delta h$  show more overlapping among the distributions corresponding to the different opponents. We tested our hypothesis using linear mixed-effects models of the three types of opponents (oscillatory agent, shadow agent and human) to assess the presence of statistically significant differences between the density distributions of  $\beta$  (Table 5.2) and  $\Delta h$  (Table 5.3) for three different cases: (1) the relative velocity between the player and its opponent (labelled in the tables as the ‘interaction’ case), (2) the individual velocity of the player (labelled as ‘player’) and (3) the individual velocity of the opponent (labelled as ‘opponent’). Both tables include the corresponding p-values resulting from the modelling.

Given the results shown in both tables and setting the significance level at 5%, we can conclude that only in the case of the relative velocity between the agents (‘interaction’ columns) are all three distributions statistically significantly different for both  $\beta$  and  $\Delta h$ .

For the case of the velocity of the player, we cannot confirm a statistically significant difference between the distributions of  $\beta$  and  $\Delta h$ . In the case of the velocity of the opponent, we could only find evidence of statistically significant differences between the oscillatory agent and the other two kinds of opponents, but not between the human opponent and the shadow



**Figure 5.8: Fractal and multifractal exponents of player and opponent's velocity time series.** Boxplot distributions of  $\beta$  (left side) and width of the multifractal spectrum  $\Delta h$  (right side) in the velocity of the players. The top figures (a and b) represent the fractal and multifractal exponents of the velocity of the player. The bottom figures (c and d) represent fractal and multifractal indices of the time series of the velocity of the opponent. Values illustrated refer to interactions between: a human and an oscillatory agent ('vs. oscillator'), a human and a shadow agent ('vs. shadow') and two human participants ('vs. human').

$\beta$	p-value		
	Interaction	Player	Opponent
human-human vs. human-oscillatory	0.0000	0.1106	0.0000
human-human vs. human-shadow	0.0017	0.6831	0.0850

**Table 5.2: Results of the linear mixed-model effects for comparing the fractal  $\beta$  exponent from DFA results between the rounds where the player was facing other human player and the two cases of programmed agents (oscillatory and shadow agents).** The left column (interaction) reflects the results when the relative velocity between the players is analysed, central column (player) shows the results when the velocity of the player is analysed and the right column (opponent) the velocity the opponent.

$\Delta h$ Groups	p-value		
	Interaction	Player	Opponent
human-human vs. human-oscillatory	0.0000	0.1405	0.0000
human-human vs. human-shadow	0.0002	0.8594	0.4601

**Table 5.3: Results of the linear mixed-model effects for comparing the fractal  $\Delta h$  exponents from MF DFA results between the rounds where the player was facing other human player and the two cases of programmed agents (oscillatory and shadow agents).** The left column (interaction) reflects the results when the relative velocity between the players is analysed. Central column (player) shows the results when the velocity of the player is analysed and right column (opponent) the velocity of the opponent.

agent.

Results show that individual variables are not suitable for discriminating between the kind of interaction going on in the case of the shadow agent. This reveals that when the individual behaviours have some kind of complexity, what it is relevant in terms of the emergence of social interaction is what is going on in the interaction between the two subjects and not the complexity of their individual behaviours. As we hypothesized at the beginning of the chapter, the invariances that constitute social engagement (in this case a scale-free dynamic structure) are built over a shared space of mutual interaction, and cannot be explained in individual terms.

## 5.7 DISCUSSION

In previous chapters we developed a notion of sensorimotor autonomy for neurocognitive agents based in the constitutive nature of brain-body-environment interaction and a multi-scale double loop of neurodynamic and sensorimotor regulation. During this chapter we have tried to extend our insights about neurocognitive autonomy to the social realm, analysing the constitutive nature of interaction between engaged agents and correlation at different scales in a minimal social experimental setting. In this setting, the perceptual crossing paradigm, we have questioned the limitations of previous approaches for capturing the complex unfolding of the different levels of cognitive and social interaction. This interpretation offers a new conceptualization of the directions in which we should focus attention: given the results shown in this chapter, it is possible that the emergence of social engagement might not depend solely on either the stability of co-regulative dynamics between two participants as suggested in previous perceptual crossing experiments and simulations (e.g. Di Paolo et al., 2008; Auvray et al., 2009). Instead, the results obtained indicate that genuine social engagement might be better characterized by a structure of cross-scale interactions that we have characterizing analysing fractal  $1/f$  scaling and the multifractal spectrum in relational variables of the process of social engagement. Fractal and multifractal exponents showed no statistically significant differences between different types of opponents when we analysed the velocity of the player or the opponent. In contrast, we did find these differences analysing the relative velocity between them,

leading us to conclude that the emergence of a  $1/f$  structure for genuine social interaction is something that happens only in the shared space between the two subjects, and the process cannot be reduced to their individual dynamics. This suggests that the process of social engagement may have in some cases an autonomy of its own, and it cannot be reduced to autonomy of individuals.

However, this work leaves several questions unanswered. The first concerns what an adequate framework of analysis might be and how previous and new insights can be integrated in a larger framework. The framework presented here still needs to be extended, since  $1/f$  scaling and the multifractal spectrum reduce the complexity of multiscale dynamics to a single exponent that, although detecting the presence of activity at different scales, falls short of characterizing the mechanisms sustaining of cross-scale interactions in social engagement in detail. Multi-scale synchronization analysis employed for measuring inter-brain synchronization in social tasks appears to be a suitable candidate for extending the analysis presented here with multiscale synchronization analysis of behavioural dynamics (Dumas et al., 2010; Müller et al., 2013). Further work should inquire into the mechanisms that make possible the emergence of an interactive space of long-range correlations. More detailed analysis may also offer new points of connection with previous work and alternative explanations for the phenomena observed in the perceptual crossing experiment.

Another way forward may lie in modifications of the perceptual crossing experiment which may prove helpful in better understanding the cross-scale interactions in minimal social interaction. Interesting advances following this approach include the work of (Iizuka et al., 2013), which studies the emergence of a communication system between two participants, using the perceptual crossing set up to collectively categorize different symbols. Also, (Froese et al., 2014b) explore the evolution of interaction of fixed pairs of players during several rounds with the objective of establishing a team for finding each other, observing that at some point the players simultaneously become aware of each other. This kind of extended experiment may allow the study of correlations at larger scales than just instantaneous mutual recognition, allowing us to analyse interesting dynamics such as learning, development of shared patterns and joint development of the player's mutual dynamical entanglement.

We also should take into account that the perceptual crossing experiment is a minimal framework for analysing dyadic social engagement. As in the case of sensorimotor interaction, the emergence of a sensorimotor engagement is not enough for addressing autonomous behaviour. Instead, an autonomous agent constitutes itself as such when it has the ability to engage and disengage different dynamical fields as result of own processes of regulation. Similarly, in social life, meaning and autonomy are constructed when social subjects can get involved sequentially in different social processes at different times. An autonomous social subject should be one that is able to recombine itself to engage and disengage different social actors or processes, regulating some organizational invariants by doing so.

In the next chapter we tackle the question we posed at the beginning of this chapter: is it possible the emergence of forms of autonomy in the social realm? What kind of organi-

zational structure may sustains these forms of autonomy? In our everyday life, dyadic social interactions are pervasive but also transient and intermixed. Social processes are generally not limited to sequential dyadic encounters but composed of one-to-one, one-to-many and many-to-many interactions at several scales. From face to face encounters to more diffuse social exchanges through different communicational and cultural artifacts. We shall move to a particular case of large-scale social autonomous behaviour for illustrating this point: massive collective organizations mediated through digital communication tools.



# 6

## The Collective Mind: large-scale self-organization in social systems

**ABSTRACT:** We assess the possibility of a social autonomy by exploring a particular case of self-organization in a ‘networked social movement’: the 15M movement in Spain. We review the notion of autonomy as applied to social and political systems finding the tools and concepts developed for neurodynamic and behavioural autonomy useful to characterize self-organized forms of construction of collective identity.

**CONTRIBUTIONS:**

- We analyse the impact of ICTs in social organization in modern societies, proposing the notion of ‘multitudinous identity’ for describing a recurrent pattern of social organization during the last years. We propose the 15M movement as a relevant case study of a multitudinous identity.
- We describe the operational closure of a network of geographical nodes the 15M movement based in transfer entropy measures of its Twitter activity.
- We find geographical synchronization patterns at multiple scales in the activity of the 15M movement in Twitter, frequently corresponding to moments of strong embodiment of protests.
- We analyse the relation between embodiment and synchronization of the 15M movement with indicators of multiscale complexity of its organizational structure (TSE complexity and  $1/f$  scaling).

**RELATED PUBLICATIONS<sup>1</sup>:**

- Monterde, A., Calleja-López, A., Aguilera, M., Barandiaran, X. E., and Postill, J. (2015). Multitudinous identities: a qualitative and network analysis of the 15M collective identity. *Information, Communication & Society*, 18(8):930–950.
- Aguilera, M., Morer, I., Barandiaran, X. E., and Bedia, M. G. (2013b). Quantifying political self-organization in social media. fractal patterns in the Spanish 15m movement on twitter. In *Advances in Artificial Life, ECAL*, volume 12:395–402.

---

<sup>1</sup>Although part of the conceptual framework and tools in this chapter were used in related publications, the experimental data, analysis and elaboration of this chapter is original.

In Chapters 3 and 4, we characterized in the cognitive domain a form of interactive autonomy which is constructed over the engagement of regulatory loops of neural self-organized patterns and sensorimotor loops of the behavioural patterns that exist in an interactive space of agent-environment coordination, producing a symmetry breaking where the agent is capable of coping with the complexity of environmental features as it acts as a coherent unit over its environment. We then shifted our attention from sensorimotor autonomy to inter-subjective social interaction. In Chapter 5 we have characterized the interactive nature of certain processes of social engagement, as a requirement for the constitution of a social subject that goes beyond the individual subjects participating in it. In this chapter we analyse the possibility of this form of collective social subject to display a form of autonomy on its own. We propose that the model of interactive cognitive autonomy that we have explored in previous chapters may share some characteristics with modes of social interactive autonomy.

Intuitively, we frequently attribute unitary qualities to large-scale social processes. We talk of ‘social consciousness’, ‘popular indignation’, ‘collective memory’, ‘general will’ and so forth. Often, we find it useful to ascribe psychological states to social entities as if they were autonomous subjects. Is this a mere metaphor and are social states just a statistical average of the individual participant psychological states? Or do groups of people have the ability of acting as a unitary subject with ‘mental’ properties that emerge at scales other than that of the individual? In this chapter we analyse what kind of collective identity would allow a community and its participants to constitute themselves as an autonomous entity with emergent properties of its own. We argue that, in certain cases, social agents can emerge and exhibit a form of collective autonomy. We describe a particular mode of collective autonomy that we name ‘multitudinous identity’, and we explore its internal mechanisms and organization and the particularities and the resemblances with other modes of autonomy.

## 6.1 SOCIAL AGENCY

During the last centuries, two starkly opposed traditions have dominated thinking about group agency (List and Pettit, 2011). On one hand, it is the ‘emergentist’ tradition or the ‘animation theory’ of society, that prospered in the late nineteenth century in history, sociology and political theory. Emergentist approaches appeared in the final decades of the nineteenth century and the early decades of the twentieth, influenced by the work of legal historians like Otto von Gierke and Frederick Maitland (Runciman, 1997). Inspired by medieval and later legal theory, they argued that group agents are emergent entities over and above the individuals who compose them. Using metaphors similar to those of biological vitalists and treating life as a mysterious, organicist force, they proposed that for a collective agent to exist, there must be something present apart from the individuals composing the group. This force takes the form of ‘the pulsation of a common purpose which surges, as it were, from above, into the mind and behaviour of members’ (Barker, 1950, p.61).

This perspective opposed individualistic atomistic views of society, defending the existence of a different level of description which serves as a centre of reference for the action of

the social agent. The emergentist tradition became fashionable in the early twentieth century and was associated with a wide range of political perspectives, often related to socialist perspectives enthusiastically highlighting the independent reality of associations, organizations and states. Nevertheless, as fascism took over in Europe, it also became associated with a totalitarian image of society (List and Pettit, 2011, p.9).

On the other hand, the ‘eliminativist’ tradition, associated with analytical philosophy and economics, arose in the 19th century and became mainstream since the middle of the 20th. This tradition is based on the methodological individualism, which was generalized in much of the economics and social sciences for the second half of the 20th century represented by Karl Popper or Friedrich Hayek (Arrow, 1994). This methodology states that explanations of social phenomena should be based on the agency of individuals and their actions and decisions. Inspired by this methodological approach, eliminativists state that agency only applies to individuals. When such individuals cooperate in groups, they do not bring novel agents into existence. Following Jeremy Bentham claim ‘[t]he community is a fictitious body’ (Bentham, 1789, p.3), eliminativists claim that collective agency is not an appropriate subject of analysis:

Groups are said to have beliefs, emotions, and attitudes and to take decisions and make promises. But these ways of speaking are plainly metaphorical. To ascribe mental predicates to a group is always an indirect way of ascribing such predicates to its members (Quinton, 1975, p.17).

At much, eliminativists defend we can speak of social states as a form of statistical reduction of its participants, but not as an entity that can be governed by a different logic than the individuals that compose it.

As in the cognitive domain, a tension arises between the intuition about a unitary subject that arises from a network of distributed entities, and the absence of mechanistic explanations of how that may happen, leading to the conclusion that it may be a phenomenon of the sum of individual dynamics (see Chapter 2 for a review of reductionist and holistic tensions in cognitive sciences). As in the cognitive domain, new scientific tools developed during the last half century allow the possibility to find a third way for characterizing social agency. Neither a ‘fictitious body’ nor a collective body existing over individuals ‘from above, into the mind and behaviour of members’, an autonomous body constructed through the interaction of different scales of description of the social system may overcome the reductionism-holism dichotomy. We think that a systemic definition of collective embodiment is possible for defining interactive autonomy using the conceptual tools of nonlinear dynamics and complex systems science.

### 6.1.1 THE CYBERNETIC VIEW OF GROUP AGENCY

As in cognitive science, the rise of cybernetics also brought a refreshing perspective about collective agency. In his seminal book 'Cybernetics or Control and Communication in the Animal and the Machine', Norbert Wiener included a last chapter discussing the application of the cybernetic method to social science (Wiener, 1965, p.155-165). For Wiener, communities exist only in terms of the effective transmission of information among its members. For him, agency of a group of animals (or a 'race', using his words) can be measured in terms of the communal information shared among them:

Whatever means of communication the race may have, it is possible to define and to measure the amount of information available to the race and to distinguish it from the amount of information available to the individual. Certainly no information available to the individual is also available to the race unless it modifies the behavior of one individual to another, nor is even that behavior of racial significance unless it is distinguishable by other individuals from other forms of behavior (Wiener, 1965, p.160).

For Wiener, social agency exists as long as there is a new level of description, that of communal information, which cannot be reduced to the information of individual agents alone, but instead is generated in the process of interaction between agents. This new level of description cannot be reduced to the behaviour of individuals but is described in terms of how the behaviour of one individual influences others. Wiener speculates that certain animal species might act as communities at some levels. For example, ant colonies are constituted by chemical communication which allows ants to recognize whether others are members of their own or other colonies. Also, hormones triggering sexual attraction in certain solitary mammals might have a communal role driving the continuation of the species. The emergence of an interactive communicational domain allows a new level of description where a social level of agency can be characterized.

Unfortunately for the objective of this work, with the end of the first cybernetic wave and the rise of computationalist approaches to artificial intelligence (see Chapter 2 for a review), these views were pushed out of the mainstream scientific research program. Instead, the artificial intelligence community shifted its interest to multi-agent systems approaches. In this approach agents are typically designed as having a set of explicitly represented beliefs in terms of some internal logical language, and are able to derive some of the logical consequences of their beliefs and communicate them to others (Wooldridge, 1992; Panzarasa and Jennings, 2001). For behaving collectively, agents must be able to determine which goals they share with other agents through communication protocols (Finin et al., 1994). So, multiagent systems are the natural extension of typical AI knowledge-based systems. Although emergent phenomena are possible in these systems, in general the objective of the designer is to avoid them and reduce the system to the scenarios designed to provide predictable solutions to an engineering

problem in a rigid and controlled structure. Despite multiagent systems approaches are sometimes considered similar to self-organized distributed systems (like artificial neural networks or ant colonies) or game theory models<sup>2</sup> the logic behind the modelling of multiagent systems is quite different and typically dismisses the possibility of emergent phenomena as a fundamental aspect of the system. In a way, by reducing the system to a set of representational goals and interactions between agents mediated by logical communication protocols, the complexity of social interaction is collapsed into a rule-based system, where the behavioural rules of the whole are described using the same language as the description of an individual's behaviour. As in computationalist artificial intelligence, the reduction of the system to only those aspects that can be captured as explicit representations of knowledge forsakes the emergence of different levels of description of the system's activity, falling into an equivalent of an eliminativist perspective of social agency.

However, in the periphery of mainstream computationalist perspectives, the second cybernetic wave and the incipient network and complex systems perspectives brought some interest to different perspectives on social agency. In social science, the term 'social autopoiesis' was coined, impelled by the German sociologist Niklas Luhmann (Luhmann, 1986). For Luhmann, there are not individuals but communications themselves that constitute the (self-reproducing) elements of the autopoietic social system:

Social systems use communications as their particular mode of autopoietic reproduction. Their elements are communication which are recursively produced and reproduced by a network of communications and which cannot exist outside such a network (Luhmann, 1986, p.174).

Thus, social systems would constitute self-referential systems with communications as its elements. Somehow, humans are not the components of the system but the sensors in the environment of the system. Other views suggest that humans are indeed the components of social autopoiesis, and what it is self-produced (by humans) is the permanent creation of the 'unity of human actors and social structures, i.e. human sociality, in society' (Fuchs and Hofkirchner, 2010).

However, Maturana and Varela have argued at different moments that autopoiesis is not directly applicable to the biological realm (Mingers, 1992). Varela showed his scepticism about this arguing that:

---

<sup>2</sup> In parallel with the development of cybernetics and artificial intelligence, research in economics popularized game theory and the idea of Nash equilibrium, where emergent phenomena are considered but typically behaviour is modelled in purely individualistic terms, where agents only seek to maximize its own utility function (Mirowski, 2001). Nevertheless, modern research has discovered a number of important complex properties of the Nash Equilibrium (Fellman, 2011) which may be used to address complex forms of organization.

These ideas are based, in my opinion, on an abuse of language. In autopoiesis, the notion of boundary has a more or less precise meaning. When, however, the net of processes is transformed into one 'interaction among people', and the cellular membrane is transformed into the limit of a human group, one falls into an abuse (Varela, 2000).

Moreover, the application of the concept of autopoiesis to the social realm, often not in a very rigorous way, has negatively affected the concept, which has sometimes been seen unscientific and 'new-agey' (Luisi, 2003, p.50). Although these first attempts to apply autopoietic ideas to society encountered some problems, it is convenient to remember that autopoiesis is a specific materialization of autonomy in the biological realm. As in the case of cognitive autonomy, we can think that a different materialization of autonomous organizations might be taking place in the social realm, one that demands a different conception of closure and self-nonsel distinction while encompassing the multiscale and modulatory aspects we have contributed in previous chapters.

Regarding this materialization, we might ask whether large human societies have the mechanisms that may allow the emergence of forms of social autonomy. For Wiener, communities (i.e. social agency) emerge in closely knit groups of humans, in which different homeostatic mechanisms play a role in creating respectable standards of care for members of the community and maintaining communal information (Wiener, 1965, p.157). However, he claimed that this cannot happen at a larger scale, in societies that remain too large to ensure direct contact between all its members. According to Wiener, vertical structures in large societies introduce a series of anti-homeostatic factors, the most important of which are mass communication media, since they operate by rules of accumulation of power and wealth rather than creating communal information (Wiener, 1965, p.161). It is by means of those anti-homeostatic mechanisms that 'the Lords of Things as They Are protect themselves from hunger by wealth, from public opinion by privacy and anonymity, from private criticism by the laws of libel and the possession of the means of communication, that ruthlessness can reach its most sublime levels' (Wiener, 1965, p.160).

An interesting perspective about the possibilities of social autonomy was soon to be materialized by the ground-breaking management experiment performed on the earliest seventies in Allende's Chile (Medina, 2006). Project Cybersyn was a decentralized decision support system to aid in the management of the national economy, a kind of proto-Internet built from telex machine and a computer operating at a national scale. The director of the project was Stafford Beer, christened by Wiener as the 'father of management cybernetics', which designed the architecture of the project based on his viable system model theory (Beer, 1995b) and a neural network approach to organizational design. The system was based on a network of telex machines in state-run enterprises transmitting and receiving information with the Chilean's government, in what Beer conceived as a 'decentralizing, worker-participative, and anti-bureaucratic manner' (Medina, 2006, p.572). Although the project was dismantled after the coup that overthrew Allende, the parts that became functional provided a coordi-

nating network structure of increasing importance for the Chilean Government. (Medina, 2006, p.593-594). Moreover, for Beer, factory management was an initial step, and envisioned numerous cybernetic applications to the Chilean social system, including the installation of 'algedonic' meters in a representative sample of Chilean homes that would allow citizens to transmit their pleasure or displeasure with political decisions enabling the government to respond rapidly to public demands, rather than repress opposing views (Medina, 2006, p.599). Interestingly, Project Cybersyn conceived social autonomy as something to be built through communication technologies, and the application of new networked technological tools was a key aspect for overtaking some of the communicational limits observed by Wiener in our present societies in the form of centralized monopolies over media.

### 6.1.2 MASS SELF-COMMUNICATION

At the time when Wiener wrote his book (and still today in most cases), communication media had a hierarchical structure in which information and control fluxes were unidirectional (e.g. television, radio). There is one centralized sender and a mass of receivers with a very limited capacity of interaction at large scales. More complex and distributed channels of communication could be created at small levels, but it was difficult for these forms of communication to scale-up to large social segments. Nevertheless, there may be some interesting cases at some particular moments and timescales, as postal networks during the Republic of Letters in the 17th and 18th centuries (Chang et al., 2009), or the widespread print-shop networks of radical reforming movements as a fundamental part of the 16th German Reformation with the generalization of the printing press (Hill, 2015). Interestingly, the rise of new digital communication tools and network technologies is pushing forward forms of fast bidirectional communication, generating what Manuel Castells has defined as mass self-communication (Castells, 2007). According to Castells, these new forms of communication increase the autonomy and influence of the social groups that use them.

Others have stressed the role of information and communication technologies (ICTs) in the emergence of collective intelligence based in distributed, enhanced, real time coordination resulting in effective mobilization of skills (Levy, 1999), and by creating a collective pool of social knowledge by expanding the domain of human interactions (Flew, 2011). Similarly, Howard Rheingold has coined the term *smart mobs*, to characterize large groups that cooperate in novel ways taking advantage of the communication and computing capabilities of widespread technological devices (Rheingold, 2007).

In general, new digital tools seem to make the idea of large-scale autonomous social organization more plausible (although this claim is not without criticism, e.g. Morozov, 2012). Moreover, Castells argues that a product of the creation of horizontal and interactive communication networks is creating a wave of self-organized 'networked social movements' aiming to redefine the rules of our social and political life and exploring new forms of being and acting together (Castells, 2012). For him, the Icelandic 'Kitchenware Revolution', Wikileaks 'Cablegate' and Anonymous' network defence, the 2011 Arab Spring, the Spanish 15M movement

and the Occupy Movement are but a few among the many examples of the increasing role played by social media in grassroots political organizing. While these and similar social movements differ in many important ways and they are grounded in specific social contexts, there is one thing they share in common: they are all interwoven through ‘autonomous’ communication networks supported by the Internet and wireless communication (Castells, 2012).

Social media has provided the tools for creating horizontal and interactive communication networks, boosting enormously the possibilities for self-organized political processes. Particular forms of collective action operating at different scales reached the public arena with an intensive use of digital networks, amplifying their events around the world, engaging thousands of people within shifting political scenarios, singular forms of political subjectivity and collective agency, and generating emergent forms of identity (e.g. Anonymous, 15M or Occupy). The presence of Internet and other digital media and the increasing use of multidirectional and interactive mass communication networks is starting to change radically the way societies organize themselves to constitute counter power or change power relationships with dominant institutions (Castells, 2007).

Moreover, intensive use of digital communication tools also provides an opportunity for new research methods in social sciences. In his book, Wiener was sceptical about the possibility of applying the cybernetic method to social sciences. For him, the difficulty of minimizing the coupling between observer and the observed phenomena was a prohibitive obstacle (Wiener, 1965, p.163). Fortunately, the rise of social media and digital data-mining opens a new door for a genuine analysis and synthesis of human social life (Lazer et al., 2009).

What kind of organizational structures can arise in these social media-based mobilizations? Are we witnessing the emergence of new modes of social agency? Do new forms of communication facilitate the constitution of a communal body that sustains modes of interactive autonomy? What is the degree of collective ‘political consciousness’ in social media-based mobilizations? We provide some answers to these questions analysing one particular case of social agency: the collective identity generated in the Spanish 15M movement as a particular example of networked social movements.

### 6.1.3 MULTITUDINOUS IDENTITIES IN RECENT NETWORKED MOVEMENTS

During the last few years, recent movements as the Arab Springs, the Occupy movement or the 15M movement have triggered an active debate in communication and political sciences about the nature of collective identity in these movements, although sometimes they are faced with the limits of previous dichotomies between holistic and individualist approaches. For example, Lance Bennett and Alexandra Segerberg have proposed that new forms of communication promote a ‘logic of connective action’ based on sharing personalized contents through social media, which is different to earlier forms of collective action linked to ‘high levels of organizational resources and the formation of collective identities’ (Bennett and Segerberg, 2012). Along the same line Kevin McDonald proposes to shift the focus ‘from collective identity’ to ‘the public experience of self’ (McDonald, 2002, p.109) and the concept of ‘fluidarity’

for understanding new forms of political organization. However, Paolo Gerbaudo (2014) is critical of these views as they downplay collective and cultural dimensions of protest through its methodological individualism. Instead, Gerbaudo suggests that an emerging culture of digitally mediated activism brings about the primacy of processes of aggregation, ‘a process of *reductio ad unum* – a reduction of the complexity of the social’, as he characterizes it, quoting Ernesto Laclau (Gerbaudo, 2013). Furthermore, he points out that aggregation implies ‘an emphasis on unity, collectivity and uniformity’ distant from what he sees as the familiar, individualistic discourse on networks. Along similar lines, Juris (2012) argues that social media contributed ‘a “logic of aggregation”, which entails the assembling of masses of individuals from diverse backgrounds within physical space’ (Juris, 2012, p.260). This logic ‘generates particular patterns of social and political interaction that involve the viral flow of information and subsequent aggregations of large numbers of individuals in concrete physical spaces’ (Juris, 2012, p.266).

Although these analyses bring into the picture different interesting aspects of the new forms of organization, a tension arises between approaches prioritizing informational and organization dimensions and those calling for a culture-centred analysis (Gerbaudo and Treré, 2015), which reminds us of the old dispute between emergentism and eliminativism opposing a microscale personal level and macroscopic notions of aggregation. In order to go beyond this tension (Monterde et al., 2015), we have proposed a different approach based on the notion of a *multitudinous identity*, a collective identity based on a statistically irreducible internal complexity based on Hardt and Negri’s notion of ‘multitude’:

The components of the masses, the mob, and the crowd are not singularities - and this is obvious from the fact that their differences so easily collapse into the indifference of the whole [...] The multitude, designates an active social subject [...] an internally different, multiple social subject (Hardt and Negri, 2005, p.99-100).

Rather than by direct, upwards identification with a symbol, person or cause, multitudinous identities emerge from large-scale processes of self-organized, continuous, interaction, where relationships of various kinds (including relations some-to-many, and many-to-many, distributed and transient), are the norm rather than the exception. We do not deny the importance of the personal dimension and singular actors, neither the role of mechanisms of aggregation of large masses of individuals. Nevertheless, the notion of multitudinous identity emphasizes the role of mesoscale levels of activity, actors and interactions, emphasizing the multiplicity, changing nature and diversity of the interactions between singular actors, groups and collective initiatives.

We contend that the idea of the ‘multitudinous identity’ helps to understand some features of the internally complex macroscopic identity of the forms of organization of new networked movements, describing it as the interrelation between irreducible scales – from micro-personal exchanges to transiently coordinated activity of large groups of synchronized actors,

up to the evolution of the organization of the macroscopic network as a whole – through distributed interaction and transient leadership among the different participants in the network (Monterde et al., 2015, p.17). Challenging more traditional forms of centralized and hierarchical political organization, the idea of multitude is irreducible to particular individuals in the network or to a fixed organizational structure.

This approach, on top of getting us beyond previous tensions between reductionist and holistic approaches, provides us with an entry point to apply notions we used to describe interactive autonomy into the social realm. Particularly, the notion of multitude resonates with one of the characteristics that we have highlighted for neurocognitive autonomy: the ability of presenting at the same time an integrated nature and a high level of differentiation of its components.

The objective of this chapter is to deepen into the notion of multitudinous identity and to characterize whether a particular example of one of the most prominent networked movements – the Spanish 15M movement – presents an autonomous organization in the terms we presented in previous chapters. As we will see below, the 15M movement is composed of a rich network structure of heterogeneous, internally multiple actors, composed of a myriad of collective initiatives, with collective goals and messages (even if frequently transient and prioritizing participation over delegation or representation), and recursivity. Our objective is to analyse what kind of mechanisms may be sustaining the multitudinous identity of the 15M movement.

## 6.2 THE 15M MOVEMENT AND ITS EVOLUTION

We take the Spanish 15M movement as a case study to depict some fundamental characteristics of the forms of collective identity arising around the wave of network-movements since 2011 (Castells, 2012). The 15M movement displays a rich, ongoing history of evolution and maturation (Monterde et al., 2015). Nevertheless, in this chapter we focus only on the first wave of massive mobilization that started the movement around May 15, 2011. In March 2011 a citizen platform was created around a manifesto calling for mass demonstrations on May 15 asking for a ‘real democracy’ in Spain. The name adopted by the platform, ‘Democracia Real Ya’ (Real Democracy Now, DRY hereafter), was soon taken-up by citizens around Spain who were encouraged to create their local nodes, without requiring any central authorization or supervision (Toret et al., 2015, p.75-77). This platform was coordinated mostly through a web forum and Facebook groups, plus a myriad of small local groups that emerged in over 60 cities during its first days of existence.

On May 15th 2011, simultaneous demonstrations took place in 60 cities up and down the country. Despite the silence of mainstream media, about 130,000 people joined the demonstrations (Serrano, 2012). Between May 16th and 22nd, a small sit-in in Madrid’s Puerta del Sol (Acampada Sol), evolved into a camp that became the first node of a network of camps that soon reached over 130 cities across Spain and another 60 more around the world. Camps spread throughout the Spanish territory through socio-technologically structured processes of

replication (Toret et al., 2015, p.88). This replication process (characteristic of different phases of the 15M movement, the Occupy movement and others) was based on the proliferation of ‘prototypes’ that served in the construction of new networked nodes (Gutiérrez, 2013b,a); those prototypes were easily replicable and highly adaptive to different local contexts. Thus, a myriad of mesoscale initiatives (contained within the 15M macro-identity) replicated quickly, creating the backbone of the camp network. A vast network of connected camps and squares emerged for the duration of the protests allowing information, calls and actions to circulate and reverberate engaging an estimated 6 to 8 million people in the protests (RTVE, 2011).

When compared with forerunners such as the alter-globalization movement (Juris, 2008), the 15M movement exhibits an increase in the range and variety of the networked practices of its participants, e.g. viral campaigns on Facebook, activity coordination on Twitter or protest live-streaming (Pérez and Gil, 2014). Unsurprisingly, Internet traffic in Spain increased by 17% from April to May 2011, indicating an intense use of Internet-related technologies (Monterde and Postill, 2014).

During the following years, the movement evolved through a continuous renewal of its repertoire of practices, combining moments of latency with periodical outbreaks of massive mobilizations, displacing the centre of gravity from the initial claims and the camps to new initiatives. These new initiatives included global joint mobilizations with other movements as the Occupy movement and anti-austerity protests in Greece and Portugal (such as the October 15, 2011, global demonstration), legal actions such as the crowdsourced campaign *15MpaRato* against a prominent banker accused of fraud and corruption, massive acts of civil disobedience such as the surrounding of the Spanish Congress by 60.000 people (*Rodea el Congreso*) in September 2012, or popular legislative initiatives for example, the one impelled by the Mortgage Victims Platform (PAH) to stem the flood of foreclosures and evictions, which garnered 1.5 million signatures in favor of the initiative. Some of these initiatives acquired special importance within the ecosystem of the 15M movement, and became regular spaces for action (e.g. the teachers and public health workers’ mobilizations under the label *mareas* (tides), or the fight against housing evictions). These practices entailed intensive use of social media and digital tools, open prototypes that can be easily appropriated, and transversal calls to actions that did not appeal to a shared identity or ideology (Candón Mena, 2013; Monterde and Postill, 2014; Toret et al., 2015). As these lines are being written, 15M-derived political initiatives are irrupting in the electoral arena. In 2014, a political party labelled ‘Podemos’ (We can) obtained 1.2 million votes in May 2014 European elections. In May 2015, Podemos was the third force in Spanish regional elections. Furthermore, several 15M-inspired independent citizen platforms concurred to May 2015 municipal elections in many Spanish cities, winning the city councils of three of the five largest Spanish cities (Madrid, Barcelona and Zaragoza). In all these elections the two main political parties in Spain both obtained the worse result in their history, indicating a renewal of political organization and real electoral effects of the claims and communication structures of the 15M movement.

### 6.3 THE 15M AS A COORDINATIVE STRUCTURE

Undoubtedly, the 15M movement has put into practice new forms of organization, knitting a powerful network of distributed mobilization. But, what kind of processes allow the construction and maintenance of this network?

It has been found that the activity of the 15M movement in social media displays structural scale-free topologies at their mesoscale, and that its dynamics exhibits typical features of critical systems such as robustness and power-law distributions (Borge-Holthoefer et al., 2011). Also, it has been shown that participants are recruited into the network in a distributed manner, through the combination of local network and global bridges percolating a coherent whole (González-Bailón et al., 2011). Furthermore, Borge-Holthoefer et al. (2015) have analysed the emergence of the 15M movement, showing that after the 15M demonstration the system undergoes a transition from a structure of geographically centralized information flows to a distributed scenario of symmetric information flows coupling together the different parts of the network. Many of these properties – criticality, compromise between local and global dynamics, emergent coordination – suggest that the system is organized through principles of distributed coordination. As we saw, the extrapolation of biological autonomy as material self-production of the system to social systems is problematic in various senses. Instead, we consider that it is more plausible that, as in neurocognitive autonomous systems, coordinative activity may be playing the role of building an operationally closed structure supporting the emergence of mesoscopic levels of regulation sustaining the autonomy at a macroscopic scale.

Coordinated activity is a powerful force in creating and sustaining human communities. William McNeill (2008) has reviewed how coordinated rhythmic movement has been a mechanism for holding human groups together for diverse historical phenomena such as warfare, plague or the pursuit of power. For McNeill, visceral and emotional sensations of shared movement, what he calls ‘muscular bonding’ is what creates the sense of community, independently of it taking the form of dances expressing revolutionary self-consciousness, the goose-stepping Nazi formations, or the morning exercises of factory workers in Japan. For example, ‘French revolutionaries experimented with Liberty tree dances and civic festivals that gave muscular public expression to the revolutionary principles of liberty, equality, and fraternity’ (McNeill, 2008, p.136). No wonder that a few years after the storming of the Bastille, when revolutionary ardour was at its peak, commemorative speeches devoted to July 14 presented it, in the words of a modern scholar, as ‘a dance, a masterpiece of pure activity in which the movements of a unanimous people were miraculously ordered in the figure of a gracious, gratuitous ballet’ (Ozouf and Sheridan, 1991, p.172).

In the case of the 15M movement, activists have described the activity of the movement as the cyclic dynamic of a collective heart, in which the activity of the system is governed by organismic cycles (Malo and Pérez, 2012) or as a ‘self-organizing climate’ which envelopes society like waves of wind (Fernández-Savater, 2012). Do networked movements present new forms of coordinated activity boosted by digital technology? Monterde and Postill (2014) argue that the recent proliferation of mobile Internet lines in 2011 allowed the use of smartphones as ar-

tulators of coordinated online spaces and occupied physical spaces. Using social media like Twitter or live streaming, demonstrators were able to create ‘mobile ensembles’ generating specific collective actions (e.g. occupying a square, preventing an eviction, or holding a general assembly). This hypothesis seems supported by the fact that, from the 16th to the 18th of May 2011, there was a 20% increase in mobile data traffic in Spain, with the number of requests being even higher, which at one point resulted in collapsed services at *Puerta del Sol* in Madrid (Monterde and Postill, 2014, p.9). In the case of other networked movements like the Occupy movement, Bennett et al. (2014) argue that an essential aspect of its network organization is the ‘dynamic integration’ which allow the network to partially integrate its participating sub-networks, depending on the ability of its users to create and recreate multiple connections between them.

Nonetheless, the mechanisms creating ‘mobile ensembles’ or ‘dynamical integration’ in mass participation in networked movements are far from clear. We consider that methodological advances are required in order to unveil the underlying dynamic and organizational structures of networked movements. In a recent book chapter, we proposed as a hypothesis that the organization of the 15M movement might be sustained by something similar to the dynamic core hypothesis in neuroscience (Barandiaran and Aguilera, 2015, p.204-207), and that it may allow an integrated and coordinated interactive assemblage, which was also distributed and reconfigurable in its structure and composition. This is supported by the fact that, during the evolution of the movement, different collective initiatives have transiently assumed roles of leadership reconfiguring the organization of the rest of the network, in what has been described as a model of ‘distributed transient leadership’ (Toret et al., 2015, p.72). In the following sections we analyse if a dynamic core organizational model is suitable to explain and characterize the behaviour of the 15M movement and what processes may be sustaining it.

Specifically, our hypothesis is that self-organized mass synchronization might be operating for the maintenance of a coherent identity, in a somewhat analogous manner to the way in which neurodynamic identities are formed. We agree with authors that see autopoietic autonomy as unfit for social processes. Instead, we consider that the model of interactive autonomy depicted in previous chapters, based on regulatory loops of coordinated activity bringing forth the operational closure of the system, might be more adequate to describe modes of social agency than those based on the material constitution of an agent. In order to explore this hypothesis, during the rest of the chapter we analyse the delimitation, underlying coordination principles and the multiscale organizational structure of the 15M identity.

#### 6.4 DELIMITATION AND CHARACTERIZATION OF THE 15M COLLECTIVE IDENTITY

In what follows we depict the identity of the 15M movement as a specific form of collective identity that we call ‘multitudinous identity’. This is a form of systemic identity that emerges from the networked interactions of heterogeneous actors (bloggers, activist reporters, alternative and independent media, etc.) including collective ones. Our interest lies in 15M’s sys-

temic, macro-identity, its emergence and its evolution over time. Interestingly, the 15M identity arises at a national level as a bottom-up, grassroots process. How does this identity emerge encompassing large groups of geographically distant people? In the following sections, we analyse the relation between the main geographical nodes of activity in social-media.

In this section, and during the rest of this chapter, we analyse a dataset of 1,444,051 tweets from 181,146 users, collected between 13 May 2011 and 31 May 2011. This dataset (Peña-López et al., 2014) was extracted from the Twitter streaming API, which provides information on the time and content of the tweet, as well as information on the sender, including location. Messages were captured when they contained one of the following hashtags or keywords (which were selected as some of the most relevant during the emergence of the 15M movement): #15M, 15-M, #democraciarealya, #tomalacalle, #Nolesvotes, #spanishrevolution, #acampadasol, #acampadabcn, #indignados, #notenemosmiedo, #nonosvamos, #yeswecamp. Since the location is a field of the description of the user, it does not necessarily correspond to the real location. We thus ran a test on geolocalized Twitter data from Spain, observing that for a set of 20.000 random tweets in a 80.25% the profile location corresponds with the actual location. Since the 15M was (at least during the first days) mainly an urban phenomena, we analyse geographical interactions between the 15 cities with more activity in Twitter during 17 days of the protests.

As stated above, we focus on the systemic dimension of collective identity formation and evolution. This aspect of ‘operational identity’ does not exclude, but should not be confused with, mutual identification or solidarity at a personal or collective level, or as a social or psychological state of symbolic subsumption ‘I/we am/feel part of X’, being X the collective identity. In this sense, we leave aside the cognitive and cultural aspects of how a collective identity is formed. We focus instead on the communicative *interaction processes* that make a diffuse social entity such as 15M emerge and autonomously maintain itself over time, that is, independent from an external agent (e.g. the State or the media) that identifies or circumvents that collective entity.

#### 6.4.1 INTERACTIVE DELIMITATION OF THE 15M IDENTITY

We defined an operationally closed system as one in which all the processes that form part of the system condition and are conditioned by other of the processes that form part of the system. In the case of a directed graph (a network whose links have a defined direction) of the influences between the nodes of the network, we may identify an operationally closed unity as a strongly connected component structure in the network (Virgo et al., 2011). A strongly connected component consists of a set of nodes of a directed graph in which, for any pair of nodes of the set, there is a path linking them. That is, information can circulate within a strongly connected component, potentially departing from and reaching any node of the component. If a node can either send information to or receive information from a set of nodes, but not the opposite, it is not part of the strongly connected component, and therefore

cannot be part of its systemic identity<sup>3</sup>. Yet, depending on what kind of data we are looking at (the time-span, the level of detail of an interaction network, or the thresholds used to define the network graph), almost any social system can be pictured as a strongly connected network. For this reason, we should think carefully what we define as processes in the system and the condition for assuming that one process ‘conditions’ another.

Using geographical information of the tweets, we compute the time-series of geographical activity at 15 cities<sup>4</sup>. Time series are defined summing the number of tweets for a step of 60 seconds. We make use of transfer entropy measures (see Section 3.6) to test whether activity at one location influences another. Transfer entropy is a good candidate to statistically analyse the causal relationships between the nodes of the network. Transfer entropy between two cities  $x$  and  $y$   $TE(\tau)_{xy}$  is computed for a sliding window of 86.400 s (i.e. 24 hours). The window is shifted for intervals of 10.800 s (i.e. 3 hours). For each window, we simplify the signal into discrete symbols by computing the presence of bursts of activity. In order to reduce the computational cost of calculations (especially for the generation of surrogates and calculating TSE complexity, which we will describe in a subsequent section), we reduce each signal into a binary time series which represents if the signal is bursting higher than its logarithmic mean. For each city  $c$ , we define a variable  $B_c$ , which represents when the activity at city  $a_c(t)$  burst to levels higher than the logarithmic mean of the activity of the window in which the signal is defined. For computing  $B_c$ , we compute the mean of the activity in a particular window, defining the activation of bursts as:

$$B_c(t) = \begin{cases} 1 & \text{if } \log(a_c(t)) > \langle \log(a_c(t)) \rangle \\ 0 & \text{otherwise} \end{cases} \quad (6.1)$$

where  $a_c(t)$  is the activity of city  $c$  in a particular window. Finally, the values of  $\tau$  represent the length of the temporal correlations analysed for computing transfer entropy. We analyse values between 60 s and 10 800 s (3 hours), logarithmically distributed using 20 representative values.

In order to statistically validate that one location’s activity is influencing other, we used data random surrogates using amplitude adjusted Fourier transform using the TISEAN software<sup>5</sup>. Amplitude adjusted Fourier transform surrogates (Schreiber and Schmitz, 1996) are time series that preserve the power spectrum of a distribution and a distribution of values, but remove the temporal correlations present in the original signal. For each city and window

---

<sup>3</sup>Although the semantical content of the messages may provide valuable information about the identity and closure of the network, here we only consider information from the point of view of the dynamical influence of one node to another

<sup>4</sup>The list of cities analysed is (sorted by volume of activity): Madrid, Barcelona, Seville, Valencia, Málaga, Zaragoza, Granada, Murcia, Bilbao, Alicante, Valladolid, London, Salamanca, Las Palmas de Gran Canaria, A Coruña. London is not a Spanish city, but it presented a high volume of activity due to the large number of Spanish young people who have emigrated there.

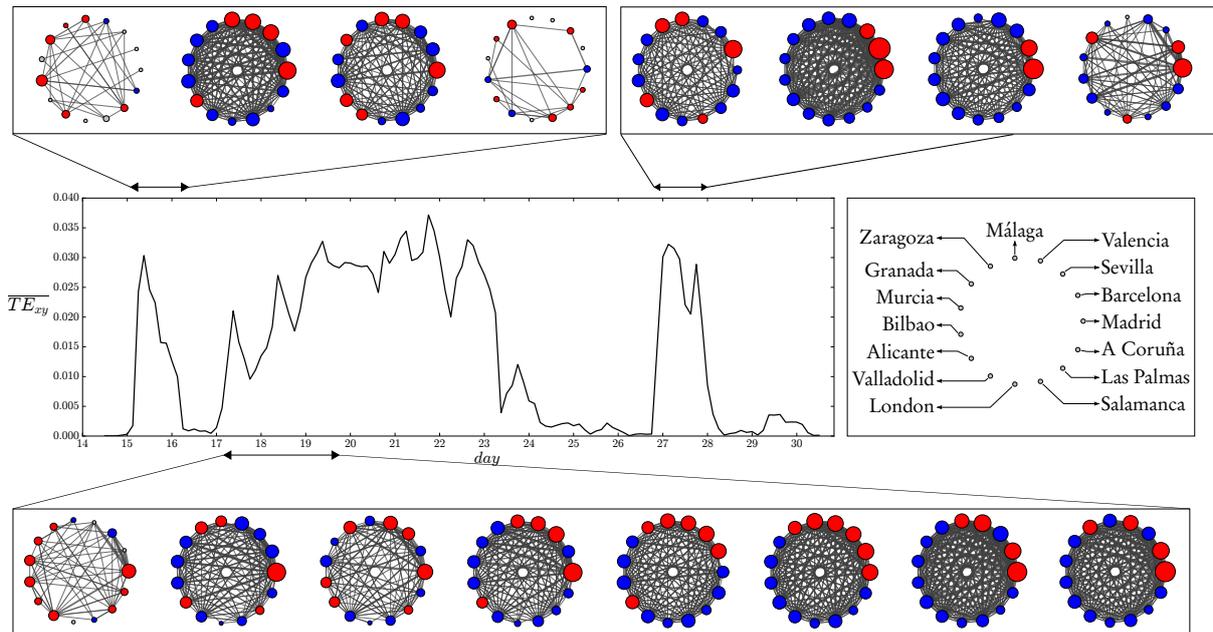
<sup>5</sup>TISEAN 3.0.1. Nonlinear Time Series Routines [http://www.mpipks-dresden.mpg.de/~tisean/Tisean\\_3.0.1/](http://www.mpipks-dresden.mpg.de/~tisean/Tisean_3.0.1/)

location, we generate 200 surrogate time series of the original data and compute the values of transfer entropy between each pair of cities for comparing them to the original data.

For testing if one process influences another, we compute the value of  $\tau$  in which transfer entropy is maximum, and check whether this value is higher than a threshold equal to 99% of the surrogate data series (we tested other percentages with similar results). This means that the values of transfer entropy presented in the original data are unlikely to be obtained by chance. Whenever a value of transfer entropy is higher than the threshold established, we consider that there is a directed causal link between two geographical nodes of the network. The intensity of the link depends of the peak value of transfer entropy between the two nodes.

In Figure 6.1 (centre) we can see the mean value of Transfer Entropy of the links in the network at different moments. As we can see, the threshold used allows us to determine moments of statistically significant causal links between nodes and moments in which this causal influence is not significant. We can observe how the periods presenting significant causal links are May 15 (the day of the demonstration that triggered the movement) the period between May 17 and May 22 (the period of most intensity of protests in the form of camps) and day 27 (a violent eviction attempt of the camp in Barcelona provokes a massive peaceful protest). As well, we can observe for some windows the shape of the resulting networks of interactions and the nodes that are within a strongly connected component. We represented each city as a node of the network and the links represent values of transfer entropy that are higher than the threshold. Madrid corresponds to the node at the right of the network (0° in polar coordinates). The rest of the cities follow counter-clockwise with steps of 23° sorted by decreasing volume of activity (Madrid, Barcelona, Seville, Valencia, Zaragoza, etc.). The size of each node is proportional to the average transfer entropy from that node to others. The width of the links is proportional to the value of information transfer. Moreover, for each network we have computed the largest strongly connected component. This is represented by the colour of the nodes. Nodes that are outside the strongly connected component are coloured in gray, whereas nodes that belong to the strongly connected components are either red or blue. Nodes are red when they have more transfer entropy in their outgoing links than in their ingoing links, and blue otherwise.

In 6.1 (top left), we can observe the emergence of the operationally closed network during May 15, in which the structure of information flows keeps changing while maintaining its closure. Then, the network slowly fades out. In 6.1 (bottom), we can observe how an operationally closed network arises again when the network of camps emerges. After the eviction of around 200 people camped in Madrid, every major city witnesses the creation of protest camps. We can see that during the first moments Madrid is the greatest information source, and then the network grows to be almost fully connected for several days, shifting which nodes act as information sources. Finally, the operationally closed network vanishes after May 22 (in which the country held its local and regional elections) and arises again on May 27 when police tried to evict the camp in Barcelona. In 6.1 (top right) we can see how the Barcelona node is the one that provides more information to the rest of the network for the first moments.



**Figure 6.1: Operational closure of the network.** In the middle, we display the mean transfer entropy links of the links higher than a threshold generated from surrogate data for windows of 24 hours. In the top and bottom figures we observe the network of links of causal influences between nodes and their operational closure. The distribution of cities in the network is shown in the right plot. The size of each node is proportional to the average transfer entropy from that node to others. The width of the links is proportional to the value of information transfer. Red and blue nodes belong to the largest strongly connected component of the network. Nodes are red when they have more transfer entropy in their outgoing links than in their ingoing links, and blue otherwise.

Towards the end of the day, the eviction is stopped and solidarity rallies are held in every city, changing the distribution of the network and the sources of information flows.

Recapitulating, we observe that the network of the 15M identity emerges as an operationally closed network with a shifting structure. The network continuously changes its configuration and the direction of information flows but it is able to maintain its operationally closed structure for several days. In the next section we analyse the mechanisms that may be maintaining this operationally closed network. Specifically, we test the hypothesis that synchronization mechanisms are behind the coherence of the network.

## 6.5 SELF-SYNCHRONIZING AGENCY

One of the most prominent features of the 15M movement was its fast territorial development. Without any coordination centre or any formal organization, the movement was able to reproduce a network of camps across Spanish in a period of a few days. In this section we analyse the coordination activity between the population of the main Spanish cities. Our hypothesis is that, in parallel to what we observed in neural activity, ‘coalitions’ or transiently

synchronized masses of users may be one of the basic building blocks of emergent distributed autonomous organizations. Moreover, we hypothesize that coordination phenomena is not a merely internal process, but if it is sustaining the organization of a collective agency it should be grounded in the embodied activity of the social body of the multitude that conforms the movement. We predict that if phase-locking mechanisms are playing a role in constituting a social agent in the 15M movement, metastable geographical coalitions should arise as it emerges in different modes of interaction with its environment.

We analyse moments of synchronization between different cities using phase locking statistics (Lachaux et al., 2000) and wavelet filtering. Wavelet filtering is used to quantify phase-locking at different moments and timescales. The span of the analysis is the 17 day period and a frequency span between  $[1.67 \cdot 10^{-3}\text{Hz}, 9.26 \cdot 10^{-5}\text{Hz}]$  (from 10 minutes to 3 hours) mapped into a logarithmic sequence with intervals of  $10^{0.01}$ . Following Lachaux et al. (2000), we define phase locking values PLV between two cities  $x$  and  $y$  as:

$$\text{PLV}_{xy}(f, t) = \left| \frac{1}{\delta} \sum_{\tau=-\delta/2}^{\delta/2} \exp(i(\theta_{y,d}(f, t + \tau) - \theta_{x,d}(f, t + \tau))) \right| \cdot A_{xy}(t) \quad (6.2)$$

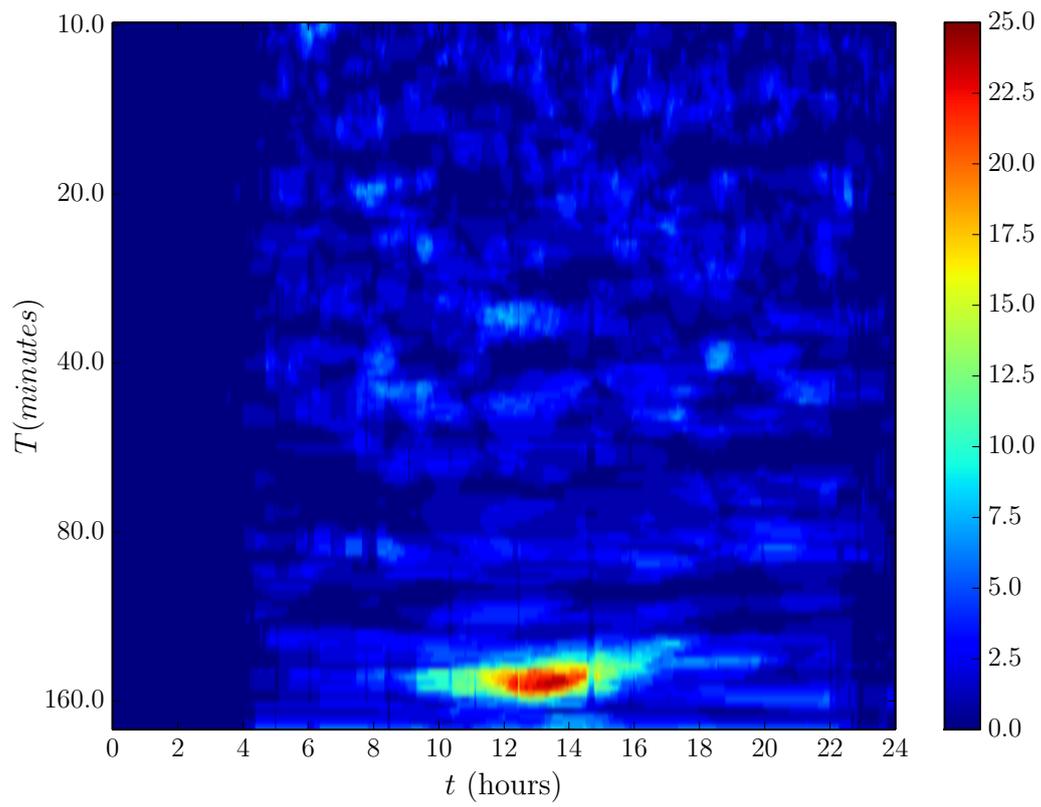
where  $\delta$  is the size of the window of temporal integration:  $\delta = \frac{n_{co}}{f}$ , being  $n_{co}$  the number of cycles in which we analyse phase-locking. Throughout the chapter, all calculations are performed taking a value of  $n_{co} = 8$ . A window of integration of 8 cycles is similar to the values typically used in neuroscience, and ensures that we are detecting sustained synchronization.  $A_{xy}(t)$  is a corrector factor which is zero when the mean activity of nodes  $x$  or  $y$  for a moving window of 30 minutes is below a threshold of 0.25 times its mean activation. The function of this corrector factor is to remove spurious synchronization when the network is inactive during sleeping hours.

The degree of statistical significance of each phase locking value is determined by comparing it to the phase locking values of surrogate time series obtained using amplitude adjusted Fourier transform as in the previous section.

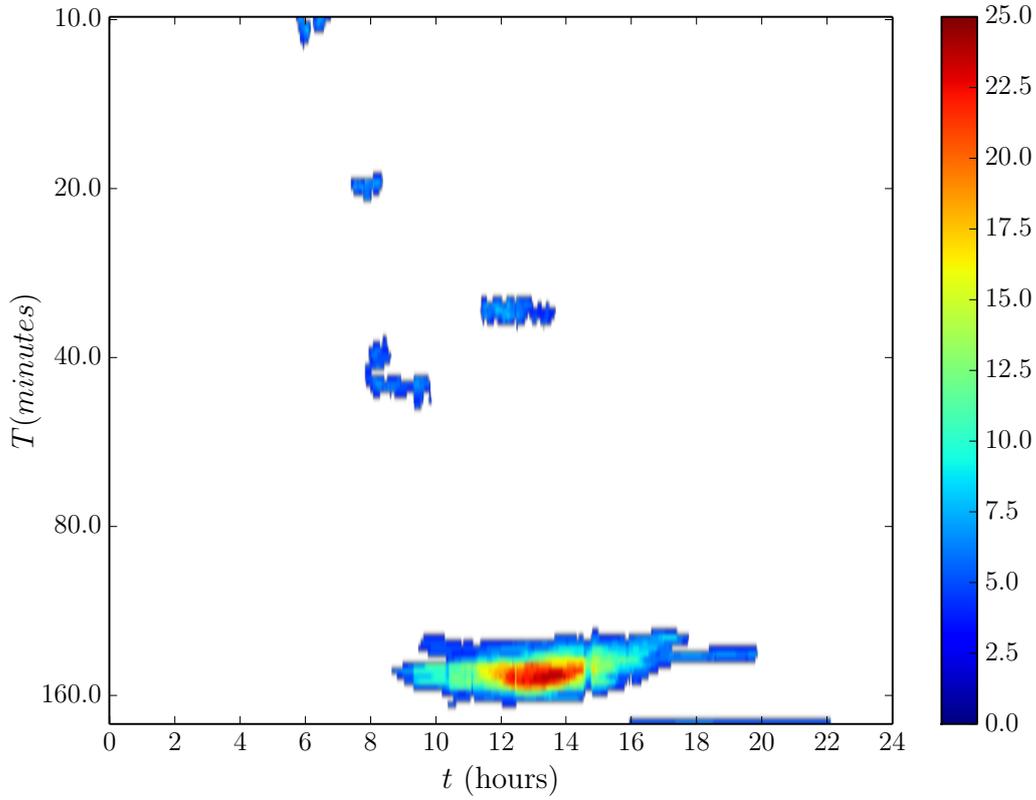
We created 200 surrogate time series. For each value of  $t$  and  $f$  these time series were used to estimate a significance threshold for the values of PLV. For each frequency  $f$ , we compute the mean average phase locking value, and we use them to compute a threshold  $\text{PLV}_{th}(f)$ , indicating a value higher than 99% of surrogate data. Thus, we define a variable defining the presence of statistically relevant phase locking links as:

$$\text{PLL}_{xy}(f, t) = \begin{cases} 1 & \text{if } \text{PLV}_{xy}(f, t) \geq \text{PLV}_{th}(f) \\ 0 & \text{otherwise} \end{cases}$$

Moreover, we compute the total number of phase locking links and values of  $f$  and  $t$ , as  $\text{PLL}(f, t) = \sum_{x,y} \text{PLL}_{xy}(f, t)$ . In Figure 6.2 we observe the resulting number of statistically significant phase locking links for a particular day (May 27). We clearly observe some moments of significant phase locking, especially at larger timescales, indicating that a particular coalition is active.



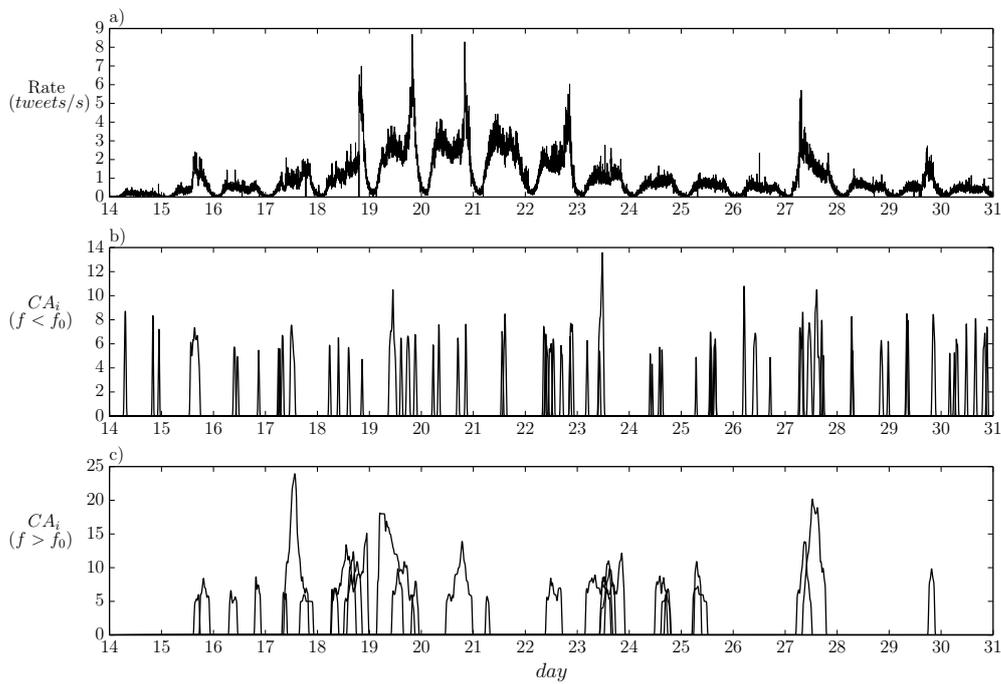
**Figure 6.2: Phase locking statistics.** Sum of the number of phase locking links between all cities for day May 27. The x axis represents the different moments during the day, and the y axis represents the period  $T$  of the different wavelets using in the filtering. The colour represents the intensity of phase locking values.



**Figure 6.3: Phase-locked coalitions.** Detected phase-locked coalitions for day May 27 applying a threshold over the phase locking statistics being higher than 99% of those of surrogate data. The x axis represents the different moments during the day, and the y axis represents the period  $T$  of the different wavelets using in the filtering. White regions represent values under established threshold determined from surrogate data  $PLV_{th}(f)$ , and coloured regions values over the threshold.

From the function  $PLL(f, t)$ , we extract the activation of coalitions when many nodes are phase-locked around a particular frequency for a prolonged span of time. For doing so, we use a Gaussian bidimensional filter with variance 0.25 for reducing noise in the function, and extract connected regions in which the value of  $PLL(f, t) \geq 4$  that spans for a time interval of at least 2 time cycles ( $\frac{2}{f}$ ). We will separate the function  $PLL(f, t)$  into different functions  $PLL_i(f, t)$ , which will be equal to  $PLL(f, t)$  in the region of the detected coalition and 0 otherwise. We observe the result in Figure 6.3.

Once the activation of coalitions is detected, we define the activation level of each coalition  $i$  as  $CA_i(t) = \max_f(PLL_i(f, t))$ . Also, the frequency carrier of each coalition is defined as  $f_i(t) = \operatorname{argmax}_f(\sum_t PLL_i(f, t))$ . In Figure 6.4 we represent the activation of the different coalitions  $CA_i(t)$ . To smooth the variations in  $CA_i(t)$  (which are abrupt since it changes in discrete steps) we filter each function  $CA_i(t)$  with a Gaussian kernel with a standard deviation



**Figure 6.4: Phase-locked coalitions.** Number of phase-locked links for the different coalitions activated above the selected duration and activation threshold. For facilitating visualization, we have separated assemblies in the lower and upper half of the logarithmic frequency spectrum, where  $f_0$  is the middle frequency. We observe that there is a correspondence between days with more intensity in coalition activation and days in which relevant events take place (e.g. days 15, 17-20, 27).

of 600S (10 minutes). Since mixing coalitions at different timescales in the same graph makes it hard to read, for facilitating their visualization we have separated the coalitions with frequency carriers in the lower and upper part of the analysed logarithmic frequency spectrum. As we show, most of the coalitions arise at relevant moments of time. Roughly, we see that concentrations of coalitions take place around significant moments of the protests (e.g. days 15, 17-20, 27).

Now, we proceed to a more detailed analysis of the relation between emergent phase-locked coalitions and moments of embodiment of the protests.

#### THE EMERGENCE OF A COLLECTIVE BODY: FIRST DAYS OF THE 15M MOVEMENT AND THE CAMPS

As we described, the 15M movement started by a simultaneous demonstration in several Spanish cities, followed by camps in the main squares of each city, where thousands of people demonstrated every day. What is the role of phase locking and coalitions during the emer-

gence of the 15M movement? Phase locking analysis segmented into different days (Figure 6.5) allows us to observe which coalitions were active at each moment:

- Day 14: day previous to the initial demonstration. Only small coalitions take place at short frequencies.
- Day 15: initial demonstration in more than 50 cities. We observe medium size coalitions at different timescales. Intermediate scale cycle arises around the beginning demonstration (which started at 18:00). A slow-scale coalition arises around 20:00 when the demonstration is about to end, which is the moment of the highest rate of tweets per second. Around 01:00, a group of 40 people camp in Madrid (*acampada Sol*).
- Day 16: groups of people start camps in cities such as Barcelona, Valencia or Granada and call for the creation of camps in other cities in the following days. Short moments of fast-scale and medium-scale coalitions appear distributed along the day. Participation peaks around 1000 people in Madrid and 200 in Barcelona.
- Day 17: Police evicts *acampada Sol* (Madrid) before dawn. The 300 people present there resist peacefully and record the incident with their mobile phones, triggering a huge reaction of indignation boosting participation. New camps are created in most major Spanish cities since the beginning of the day (Zaragoza, Málaga, Seville, etc.)<sup>6</sup>, and a demonstration of 20.000 people peacefully retake Sol square and rebuilds the camp (Toret et al., 2015, p.51). We observe how fast and medium scale coalitions start arising during the morning, triggering a large slow-scale coalition which is active for most of the day.
- Days 18, 19, 20: the highest levels of participation in the camps take place during these days. Every day, every major city in Spain witnesses major mobilizations with rallies in their main squares<sup>7</sup>. In most places, synchronized protest actions take place in many cities at the same time like ‘caceroladas’ (pots and pans demonstrations) which were repeated every day in many cities or a ‘silent scream’ which takes place on day 20 at midnight, during the last minutes before the beginning of the electoral silence day. We observe that during these days there is high presence of phase-locked coalitions, both at

---

<sup>6</sup>Source:

[https://es.wikipedia.org/wiki/Anexo:Protestas\\_en\\_Espa%C3%B1a\\_de\\_2011-2014\\_%28mayo\\_de\\_2011%29#Martes\\_17\\_-\\_Comienzo\\_de\\_las\\_acampadas](https://es.wikipedia.org/wiki/Anexo:Protestas_en_Espa%C3%B1a_de_2011-2014_%28mayo_de_2011%29#Martes_17_-_Comienzo_de_las_acampadas)

<sup>7</sup>Sources:

[https://es.wikipedia.org/wiki/Anexo:Protestas\\_en\\_Espa%C3%B1a\\_de\\_2011-2014\\_%28mayo\\_de\\_2011%29#Mi.C3.A9rcoles\\_18](https://es.wikipedia.org/wiki/Anexo:Protestas_en_Espa%C3%B1a_de_2011-2014_%28mayo_de_2011%29#Mi.C3.A9rcoles_18)

[https://es.wikipedia.org/wiki/Anexo:Protestas\\_en\\_Espa%C3%B1a\\_de\\_2011-2014\\_%28mayo\\_de\\_2011%29#Jueves\\_19](https://es.wikipedia.org/wiki/Anexo:Protestas_en_Espa%C3%B1a_de_2011-2014_%28mayo_de_2011%29#Jueves_19)

[https://es.wikipedia.org/wiki/Anexo:Protestas\\_en\\_Espa%C3%B1a\\_de\\_2011-2014\\_%28mayo\\_de\\_2011%29#Viernes\\_20\\_-\\_Formulaci.C3.B3n\\_de\\_propuestas](https://es.wikipedia.org/wiki/Anexo:Protestas_en_Espa%C3%B1a_de_2011-2014_%28mayo_de_2011%29#Viernes_20_-_Formulaci.C3.B3n_de_propuestas)

slow and fast timescales. Also, the rate of tweets per second and the level of activation of the coalitions (the number of cities synchronized) is higher than any other day (except for day 27 in which, as we analyse below, another mayor mobilization takes place).

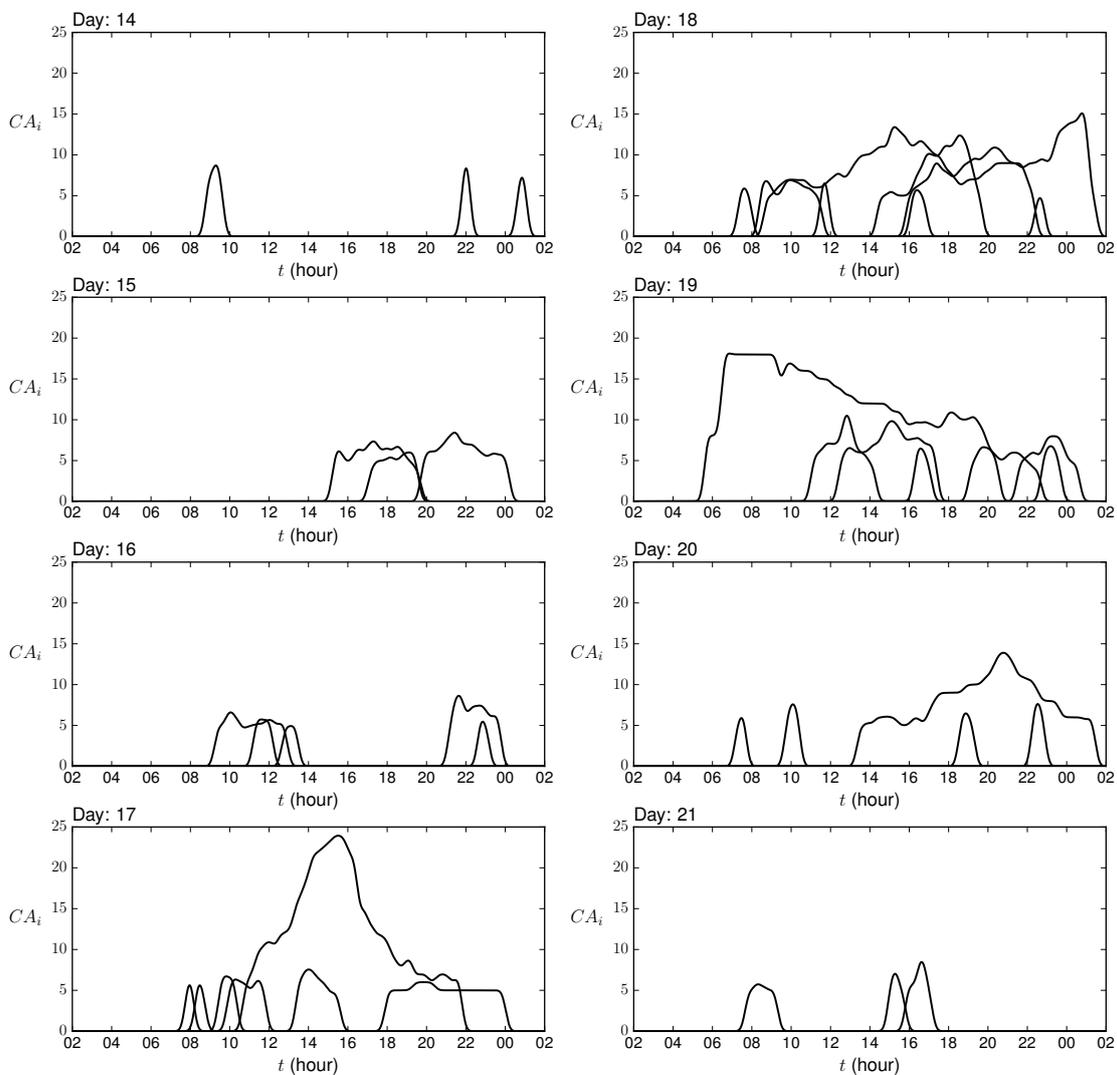
We observe that for some days massive mobilizations are accompanied with episodes of strong synchronization in social media. In Figure 6.4 we observe that during days 21 and 22 both activity and phase-locked coalitions decrease. This is probably due to the fact that day 21 was ‘reflection day’ in Spain, in which 24 hours of election silence are mandated, and day 22 was election day. After the election, on day 23, a widespread debate emerged about the need to create a minimum consensus about the goals of the movement, but in general, the level of activity and synchrony decreased until day 27 (see Figure 6.4).

In general, we observe that the emergence of the movement is connected with moments of synchronization. Moreover, it is interesting to observe the correspondence between moments in which the protest is embodied in the streets (camps, rallies, etc.) and moments of intense phase locking in social media, suggesting that the communication circuitry in social media is strongly coupled with physical events taking place in the streets. Similarly, when the activity in the camps and the streets decreases, phase-locked coalitions are reduced as well. As we hypothesized at the beginning of the section, processes of coordination between geographical parts of the network seem to be grounded in the embodiment of the participants of the network.

#### ADAPTIVE AGENTIAL RECONFIGURATION: RESISTING EVICTION

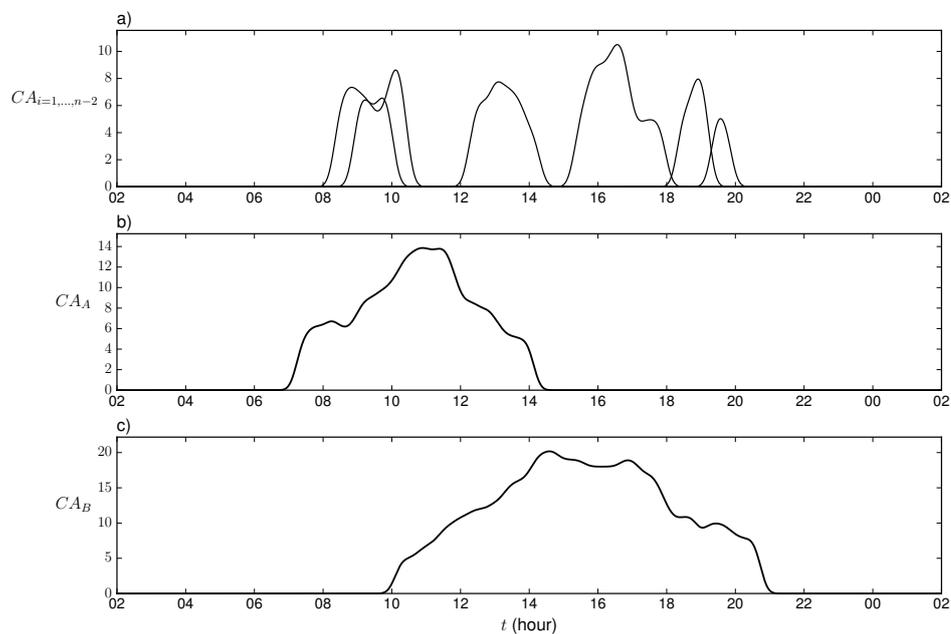
Given only the analysis above, we might think that what we are witnessing are just waves of synchronized activity as the system responds to protest events. Does the synchronization of the network have a role in the organization and reconfiguration of the movement’s activity? That is, is transient synchronization playing a constitutive role in the organization of a collective agent or is it just an epiphenomenon of other processes? We analyse this using a particular case. We have shown that after some days of lower levels of activity, on May 27 we observe the emergence of a new episode of high rates of activity and large phase-locked coalitions. This day, the camp at *Plaza Catalunya* in Barcelona (one of the most representative) is evicted by the police. If we analyse the sequence of events during this day (Toret et al., 2015, p.139-142) and our phase locking analysis (Figure 6.6), we observe that there is a strong relation between the reconfiguration of phase-locked coalitions between different cities and the course of embodied events in the protest camps. As we saw in Figure 6.3, during this day there are several fast-scale coalitions (Figure 6.6.a) and two main slow-scale coalitions  $CA_A$  and  $CA_B$  that last for several hours (Figures 6.6.b and 6.6.c).

Events start around 6:30, when the police arrive to the camp with the intention of starting an eviction and demonstrators start peacefully resisting and recording everything with their mobile phones. By 7:20 the hashtag *#BarcelonaSinMiedo* (Barcelona without fear) becomes a ‘trending topic’ in Spain. At the same time, coalition  $CA_A$  emerges (Figure 6.6.b).



**Figure 6.5: Phase-locked coalitions activated for particular days.** Number of phase-locked links for the different coalitions activated above the selected duration and activation threshold for windows compressing specific days. As we review in the tests, moments with the highest synchronization correspond with moment of strong embodiment of the protests in particular events taking place in the street (during days, 15, 17, 18 and 19)

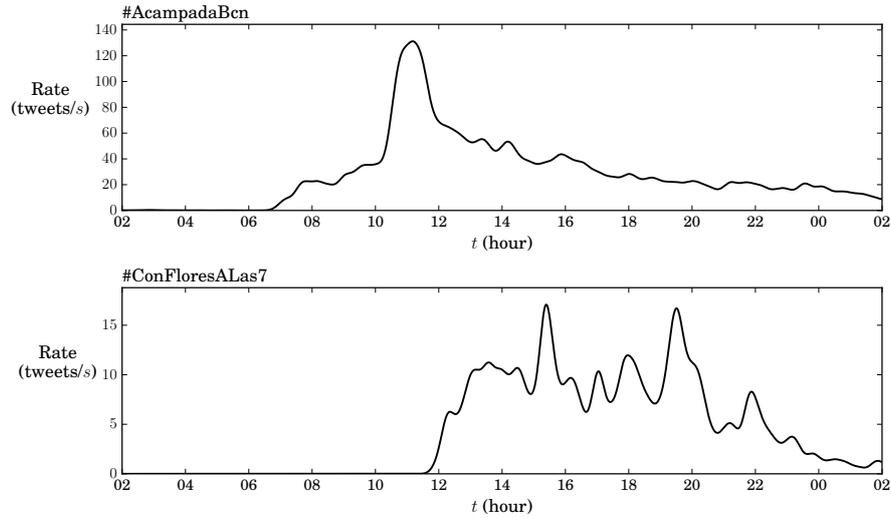
This activity starts raising attention and people start to flock to Plaza Catalunya, while fast-scale coalitions emerge (Figure 6.6.a). Around 10:00 a thousand people are present at Plaza Catalunya, blocking the police from evicting the camp. While more and more people arrive, police start beating people with batons, generating images depicting violence from the police towards peaceful demonstrators, which became viral in social media, multiplying activity in Twitter where sympathizers, press and other media covered the events. At 11:30, in a moment of most intense police violence, a national TV channel airs live coverage of the event showing live police brutality. At 12:00, thousands of people follow the images of what is happening in Plaza Catalunya. At this moment, every trending topic in Spain is related to the eviction of the square. Simultaneously, a new medium-scale coalition arises between 12:00 and 14:00.



**Figure 6.6: Phase-locked coalitions activated on May 27.** Number of phase-locked links for the different coalitions activated above the selected duration and activation threshold for day May 27. a) Medium and fast-scale coalitions activated during day May 27. b) First slow-scale coalition  $CA_A$ . c) Second slow-scale coalition  $CA_B$ .

While more and more people arrive, peaceful resistance to police violence gets stronger and demonstrators become more determined, finally making police desist from their attempts of evicting the square and around at around 14:00 a wave of people reconquer the square and start rebuilding the camp in a state of shock from police violence (Toret et al., 2015, p.140). The slow-scale coalition  $CA_A$  starts decreasing when the square is reconquered and fades out a couple of hours later.

In parallel with the events in Barcelona, a call for a mobilization in solidarity with the eviction attempt of Plaza Catalunya was emerging. This call was soon materialized in the



**Figure 6.7: Hashtag evolution.** Evolution of tweet rate for tweets containing the hashtags a) #AcampadaBcn and b) #ConFloresALas7.

hashtag ‘#ConFloresALas7’, proposing a rally in every Spanish city at 19:00 bringing flowers to protest against police brutality (Figure 6.7). Interestingly, a slow-scale coalition  $CA_B$  arises around a couple of hours earlier than the hashtag (Figure 6.6.b). This coalition lasts the whole day and has its peak in the afternoon.

The results so far suggest that two macroscopic processes are taking place. First, the reaction to the eviction, which is correlated with the emergence of a long-lasting coalition  $CA_A$  (Figure 6.6.b), and at least a related hashtag ‘#AcampadaBcn’<sup>8</sup> (Figure 6.7.a). Another process is the organization of solidarity demonstrations, which is correlated with the emergence of another long-lasting coalition  $CA_B$  (Figure 6.6.c) and the hashtag ‘#ConFloresALas7’ (6.7.b).

We analyse some aspects of their structure by analysing the information flows in each of them. In order to filter the activity related to each one, we compute the carrier frequency for each coalition (the frequency in which significant phase-locking last longer) and we filter the activity at each city using wavelets scaled to the carrier frequencies of coalitions  $CA_A$  and  $CA_B$ . Convolution of the wavelets with the series of each city  $c$ , we obtain the complex signals  $w_{A,c}(t)$  and  $w_{B,c}(t)$ . To extract the information contained in the phase of each signal, we extract the relative phase of each signal in respect to the mean phase of the coalition  $CA_A$  for a particular city  $c_i$  as:

$$\phi_{A,c_i}(t) = \angle(w_{A,c_i}(t)) - \angle\left(\frac{1}{C} \sum_{j=1}^C w_{A,j}(t)\right) \quad (6.3)$$

<sup>8</sup>Although the hashtag #BarcelonaSinMiedo was the first one about the eviction to become Trending Topic, in our dataset it was not gathered until a few hours after it was first used. Thus, we chose to use the hashtag #AcampadaBcn which was also used to talk about what was happening in the camp.

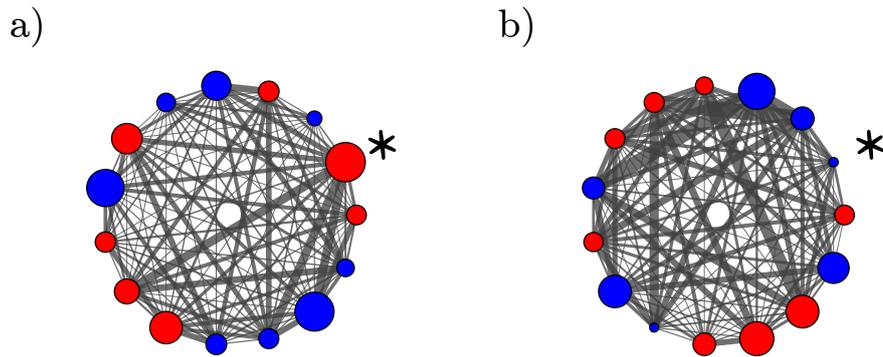
where  $C$  is the number of cities. Similarly, we can obtain the relative phase of signals in coalition  $CA_B$ .

Having computed the phase information, we use transfer entropy to compute flows of information in the network of each coalition. For doing so, we discretize the values of  $\phi_{A,c}$  simply by defining a signal  $\phi_{A,c}^*$  which is equal to one when  $\phi_{A,c} > 0$  and zero otherwise. For each pair of nodes  $c_i$  and  $c_j$ , we compute the transfer entropy  $TE(\tau)_{\phi_{A,c_i}^* \rightarrow \phi_{A,c_j}^*}$  using values of  $\tau$  logarithmically distributed between 60 s and 10 800 s (3 hours), using 20 representative values. Then we take the peak value of  $TE(\tau)_{\phi_{A,c_i}^* \rightarrow \phi_{A,c_j}^*}$ , and assign this value for the link between nodes  $TE_{A,ij} = \max_{\tau}(TE(\tau)_{\phi_{A,c_i}^* \rightarrow \phi_{A,c_j}^*})$ . Since we are interested in analysing which nodes act as information sources or sinks, for every pair of nodes we compute the differential information  $\Delta TE_{A,ij} = TE_{A,ij} - TE_{A,ji}$ . In Figure 6.8 we can observe the differential information flows in the network of coalitions  $CA_A$  and  $CA_B$ . The width of the edges represent the differential transfer entropy between each pair of nodes  $\Delta TE_{A,ij}$ . The size of the node represents the absolute mean differential transfer entropy between one node and the others  $\frac{1}{C-1} \sum_{j=1}^C \Delta TE_{A,ij}$ , and the colour of the node represents the sign of the mean differential transfer entropy. Red nodes characterize information sources, and blue nodes represent information sinks.

We can observe in Figure 6.8.a that in coalition  $CA_A$ , the node of Barcelona (marked with an asterisk) is the main information source in the network. In contrast, we observe in Figure 6.8.b that in coalition  $CA_B$ , the node of Barcelona is no longer the main source of information and information flows are distributed among other nodes of the network. This suggests that coalitions not only serve to bind together the activity of different geographical nodes, but they also serve to allow the distribution of information flows between them. We see here how at the phase-locking frequency of  $CA_A$  information flows from the node of Barcelona to the rest, distributing information about the eviction and leading a synchronized wave of activity to counteract the eviction attempt in the city of Barcelona. At a slower frequency carrier, coalition  $CA_B$  distributes information about the organization of a network of solidarity demonstrations that take place in the afternoon. Here, the lead of the coalition is not assigned to the node of Barcelona but is distributed among different cities. Interestingly, as we can see in Figures 6.6.b and 6.6.c, both coalitions overlap for some time. Although it is not shown in the figure, some nodes are synchronized to both coalitions at the same time. We hypothesize that the different frequencies of the coalition may serve to filter the content related to different events: the local response to the eviction attempt, and the global response of the whole system as a solidarity demonstration protesting against the eviction.

## 6.6 MULTISCALE ORGANIZATION AND EMBODIED ACTION

Our hypothesis in this chapter is that the collective agent behind the 15M is organized in the form of a multitudinous identity, generated by the operational closure of the system through transient synchronizing links at different scales. One of the properties of multitudinous identities is its multiplicity of scales of description. Multitudinous identities conform unitary sub-



**Figure 6.8: Integration flows in phase-locked coalitions for May 27.** The networks represent the transfer entropy information flows between the phase of the activity at each node for coalitions  $CA_A$  (a) and  $CA_B$  (b). The width of the edges represent the differential transfer entropy between each pair of nodes  $\Delta TE_{A,i,j}$ . The size of the node represents the absolute mean differential transfer entropy between one node and the others  $\frac{1}{C-1} \sum_{j=1}^C \Delta TE_{A,i,j}$ , and the colour represents the sign of the mean differential transfer entropy. Red nodes characterize information sources, and blue nodes represent information sinks. While coalition  $CA_A$  distributes information preferentially from Barcelona to other nodes, coalition  $CA_B$  information flows are distributed more homogeneously.

jects which at the same time are internally heterogeneous and singular actors. We have studied the inter-city heterogeneity and coordinative activity but other investigations have highlighted and studied in detail the non-localized diversity of the 15M networks (Monterde et al., 2015). In Chapter 4, we described how the forms of closure emerging in an autonomous neurocognitive agent displayed temporal and spatial multiscale structures. These organizational structures allow a neurocognitive agent to flexibly adapt to the complexity of their environment by generating internal complex patterns and, at the same time, they are able to integrate those patterns at particular moments in order to act as a coherent entity over their environment, sustaining internal organizational invariances. Thus, sensorimotor neurodynamic structures of an autonomous agent are internally multiple and differentiated amongst a rich repertoire of behavioural modes, and at the same times it exhibits a unitary and coherent structure. We suggest that a similar mode of organization may be taking place in multitudinous identities.

Above we have analysed how moments of statistically significant phase-locking emerge at different moments and timescales. Moreover, we have analysed how, in many cases, the emergence of phase-locked coalitions coincides temporally with significant events in which the activity of the network is physically embodied in the activity of crowds occupying camps and squares. We may ask ourselves what kind of organizative structure produces these phase-locking moments. Are they the product of social media activity just reactively activating waves of synchronization in response to protest events? Or are they the product of a multiscale internal regulatory structure organizing a heterogeneous and internally multiple social agency?

In the previous section we observed how phase-locked coalitions arise at multiple scales, and how the system switches from one to the other as it engages in different modes of interaction with its environment. Moreover, information flows seem to go back and forth between the embodied activity and the self-organized emergent coalitions, suggesting that complex mechanisms are involved in the interaction between of phase-locked coalitions and embodied protests events. In order to explore what the organizational structure of the 15M collective agent is, we analyse the appearance of significant indices of TSE complexity and  $1/f$  scaling, and their correlation with the emergence of waves of synchronization and embodied protest events.

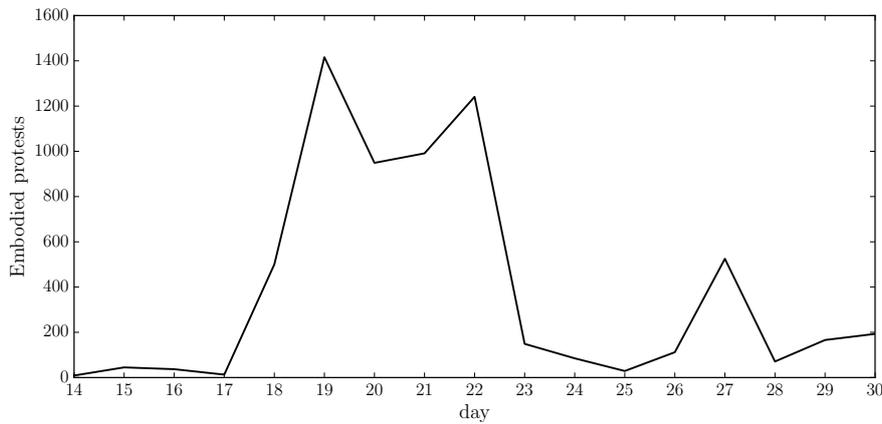
#### EMBODIMENT OF PROTESTS AND SYNCHRONIZED ACTIVITY

In previous sections, we have described specific moments in which we can observe an interaction between the emergence of phase-locked coalitions and events of embodied protests. We argued that both of them are intertwined in the organization of the 15M multitudinous identity. So far, we have described the embodiment of protests in physical demonstrations using only qualitative descriptions extracted from the literature. Here, we want to compare a quantitative description of protest events with indices of synchronization and complexity. For doing so, we made use of the GDELT Event Database, which records over 300 categories of physical activities around the world each day extracted from news reports from a variety of international sources <sup>9</sup>. We extracted events that happened in Spain between May 14 and May 30 2011 labelled with the codes 140 to 149 (representing protest events). For each event, we extracted the number of mentions which we added to the number of events on each day. In Figure 6.9 we can observe the number of mentions from news sources about protest events in Spain, which we will use as a proxy for the intensity of protest events during each day. We observe how protest intensity rises for day 18 until day 22, presenting a peak for day 27. However, the absence of protests during days 15 to 17 has to be taken carefully, since it has been indicated that during those days mainstream media blacked out the protest from their news bulletins (Toret et al., 2015, p.95).

In order to compare embodied protest activity with synchronization between the network's nodes, we computed the mean activation of the sum of coalitions for a sliding window of 86.400 s (i.e. 24 hours) as  $S = \langle \sum_i CA_i(t) \rangle$ . The window is shifted for intervals of 10.800 s (i.e. 3 hours). As we can observe in Figure 6.10, peaks in synchronization arise around days 17-19, 23 and 27. As we saw above, some of these peaks coincide with particular events like the mobilization of camps (days 17-19) and the eviction attempt of Plaza Catalunya (day 27). The only peak that does not correspond with a major event of physical mobilization is days 23, which correspond to the day after electoral result, in which the movement was intensively debating the results and the strategy to be followed around the hashtag #ConsensoDeMínimos. Although some of these debates took place physically in assemblies in the squares, much of the debate took part in social media. We can say that there is a rough correlation between

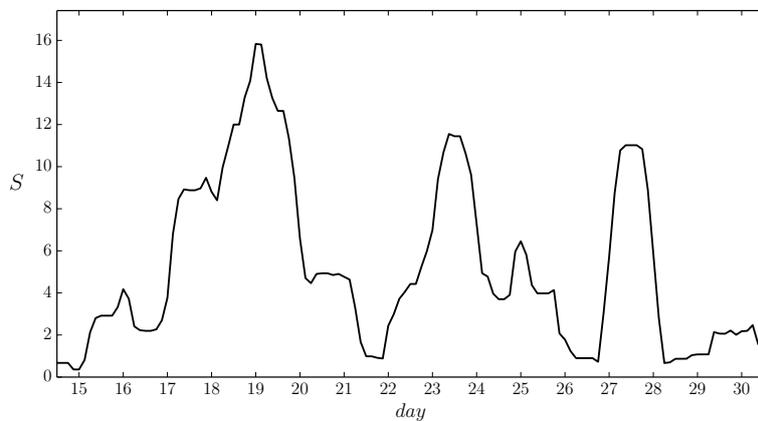
---

<sup>9</sup><http://gdeltproject.org/>



**Figure 6.9: Embodied protest activity.** Sum of mentions to protest events in Spain for different days of May extracted from the GDELT Event Database.

synchronization indices and protest events, although some synchronized coalitions emerged without the presence of protest events (day 23) and some protest events did not show significant synchronized structures (the continuation of camp protests during days 20 to 22).



**Figure 6.10: Sum of active coalitions.** The mean of the sum of active coalitions computed for a sliding window with a length of 24h.

### TSE COMPLEXITY

As we saw in chapter 4, TSE complexity is a measure of information complexity in the spatial domain, allowing us to identify the simultaneous presence of high global integration between the components of a system and high differentiation of their local activity. We analyse TSE complexity of the system by measuring the levels of integration and segregation amongst the

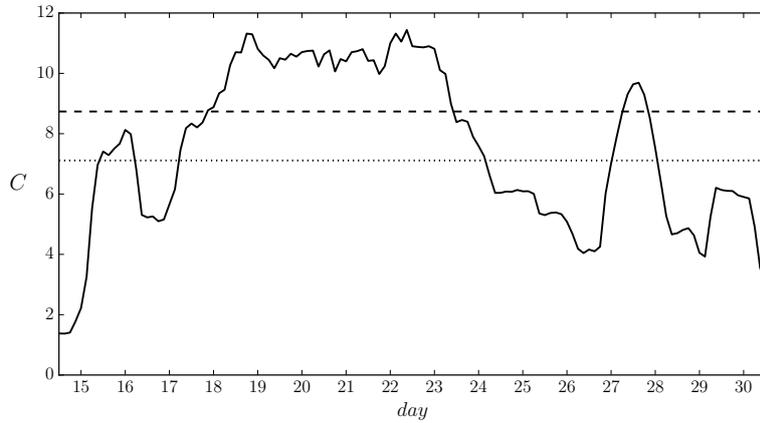
different analysed cities. High levels of TSE would suggest that while the activity at each city is differentiated and adapted to its local context, the different cities are at the same time integrated into a coherent pattern of activity. This is precisely what the notion of multitudinous identity tries to capture; an identity that is not built on the homogeneity but on the diversity of its participants coordinated in a shared common domain. Again we have focused on inter-city diversity and multiplicity but a more distributed analysis is possible (where diversity of nodes are depicted in terms of types of users, types of political organization, etc.). In order to avoid the high computational costs of computing joint density functions in our 15-dimensional system, we compute TSE complexity over the binary variables  $B_c(t)$  computed from Equation 6.1.

We proceed to compute TSE complexity according to Equation 4.4. In Figure 6.11 we can observe the resulting values of TSE complexity  $C$  for a sliding window of 86.400 s (i.e. 24 hours). The window is shifted for intervals of 10.800 s (i.e. 3 hours). Furthermore, we compare the results to that of surrogate data using the amplitude adjusted Fourier transform as in previous sections removing the correlations in the signal. We computed TSE complexity for 200 series of surrogate data, and extracted a threshold of the value  $C$  being higher than 99% of surrogate data series (dashed line in Figure 6.11). As shown, we find values of high complexity (above threshold) during days 17-23, and day 27. Also, on day 15 we find values higher than the surrounding days, being close to the threshold. A first difference we observe between the values of TSE complexity and the activation of coalitions is that the former changes more gradually. A possible interpretation is that apart from phase-locking dynamics, there is a complex structure of integrated and differentiated activity that is self-sustained in a longer timescale. For example, during days 21 and 22 the strong embodiment of the protests disappears (probably because it was election day on day 22 and mandated election silence on day 21, so massive protests did not take place), but the levels of integration and segregation of the network are maintained.

## FRACTAL ANALYSIS

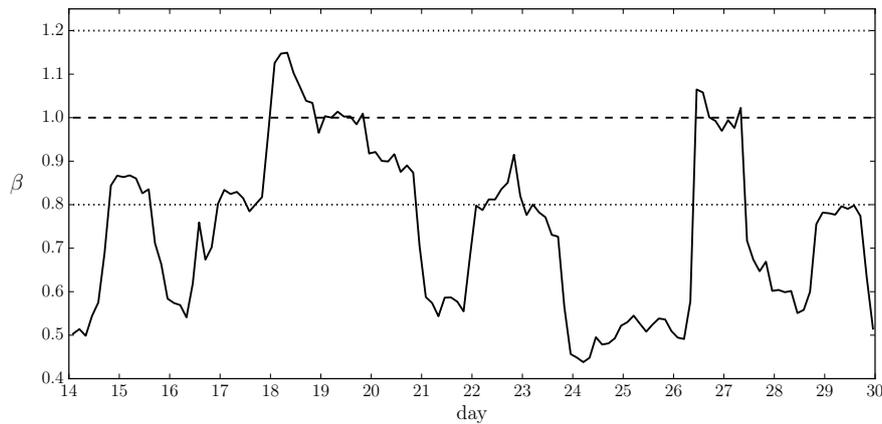
Finally, we analyse the presence of long-range correlations in the activity of the system. For this, we use the DFA algorithm defined in Equation 4.8 to analyse scale-free dynamics in windows of 12 hours of activity (we did not use longer windows because daily cycles can introduce periodicities that change the spectral distribution of the series) for time series of the activity rate at every moment (measured as the total number of tweets per second computed for every value of the time series). In most cases, the shape of the function  $F(n)$  had a good log-linear fit, thus we only extracted the exponent of the slope  $\alpha$  and transform it into  $\beta$  ( $\beta = 2\alpha - 1$ , see section 4) to directly compare the fractal scaling of the system to that of a  $1/f^\beta$  spectrum. In Figure 6.12 we observe how long-range correlations are present at different moments. For example, if we set an arbitrary threshold of  $\beta \in [0.8, 1.2]$ , we find values around scale-free distributions ( $\beta = 1$ ) for days 15, 17-21, 23 and 27.

These results show that scale-free dynamics emerge when there are many coalitions present



**Figure 6.11: TSE Complexity.** Values TSE complexity for the activity each day. The dashed line represents a threshold value for the value of TSE complexity being higher than 99% of surrogate data series. The dotted line represents the mean value of TSE complexity for surrogate data.

in the system. This makes sense since coalitions arise at many temporal scales. Also, it suggests nested dynamical structures with several bottom-up and top-down regulatory loops.



**Figure 6.12: Fractal scaling.**  $\beta$  coefficient from the DFA algorithm for different windows with a span of 12 hours.

## RECAPITULATION

In general, we observe that the moments with the most intense embodied activity (days 17-20 and day 27) present many coalitions arising at different scales, as well as both high TSE complexity and long-range correlations in the form of  $1/f$  patterns. In Figure 6.13 we compare different indices to observe the correlations between the different indices we have measured. First, we have the indices of embodied protest activity and the mean active synchronization

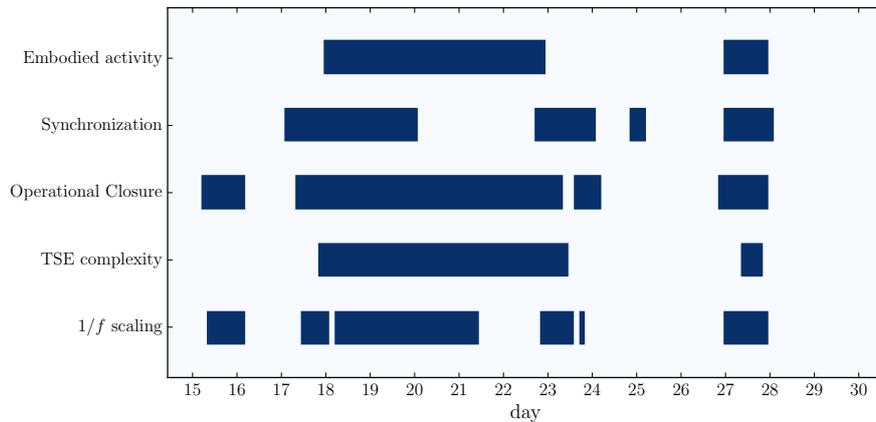
coalitions. Embodied activity and synchronization are computed as the number for mentions to protest (Figure 6.9) and the sum of active coalitions (Figure 6.10) being higher than their mean value. We also show the emergence of the operational closure of the system (Figure 6.1), measured as the moments in which the strongly connected component of the networks in Figure 6.1 covers the whole network of 15 cities. These three indices are indicators of how the network emerges as a coherent, agencial, entity, i.e. as a network that is operationally closed, coordinated at different scales and strongly embodied in the physical world. Finally, we show the indices of temporal and spatial multiscale complexity. The index for TSE complexity is computed as the moments in which values of TSE complexity (Figure 6.11) is higher than 99% of the values of surrogate data. Finally, the index for  $1/f$  scaling is computed as the moments in which the  $\beta$  exponent (Figure 6.12) is between 0.8 and 1.2. These indices account for the temporal and spatial structure of the network's organization, representing the internal multiplicity in their activity.

We find a correlation between the indices of embodiment, synchronization and operational closure and the indices of spatial and temporal multiscale complexity. This suggests that agency in mass mobilizations is facilitated by coordination between different spatial and temporal scales of the system, based on distributed phase-locking mechanisms allowing an increase in both informational complexity and long-range correlations. The system emerges as a unitary subject acting in space observing the increase of protest activity and synchronized activity in the system. We can confirm that the constitution of the system as a collective agent (with strong internal and external cohesion) is accompanied with the configuration of a multiscale structure at different moments in both the temporal and spatial domain. This suggest that the form of agency that is being constituted is indeed a form of multitudinous identity, where the creation of a unitary subject is based in the increase of the internal multiplicity in an interactive network structure both at the level of distributed locations (analyzed here) and at the level of different initiatives, users and types of collectives (as we analyzed in *Monterde et al., 2015*).

## 6.7 DISCUSSION

In this chapter we have applied analytical tools used for characterizing autonomous behaviour in neurocognitive agents to massive collective agents. We first reviewed the ontological tension in sociological literature between microscale personal networking and macroscopic notions of aggregation. We have proposed that some forms of social coordination might go beyond this dichotomy, allowing the emergence of statistically irreducible networks of communicative interaction at a mesoscopic level (between multiple singularities, groups, collective initiatives and spatial locations). Moreover, we have proposed that the intensive use of digital technologies might be boosting these forms of collective organization.

Our hypothesis is that mechanisms allowing massive distributed communication between human groups makes possible the emergence of collective forms of autonomous agency, that we labelled multitudinous identities, that differ from traditional views of collective groups



**Figure 6.13: Comparison between indicators of agency and indices of multiscale complexity.** Comparison from indices from Figures 6.9, 6.10, 6.1, 6.11 and 6.12. Embodied activity and synchronization are computed as the number for mentions to protest (Figure 6.9) and the sum of active coalitions (Figure 6.10) being higher than their mean value. Operational closure is measured as the moments in which the strongly connected component of the networks in Figure 6.1 covers the whole network of 15 cities. The index for TSE complexity is computed as the moments in which values of TSE complexity (Figure 6.11) is higher than 99% of the values of surrogate data. Finally, the index for  $1/f$  scaling is computed as the moments in which the  $\beta$  exponent (Figure 6.12) is between 0.8 and 1.2.

based on hierarchical entities or representative delegation of decision-making. Rather than by direct, upwards identification with a symbol, person or cause, multitudinous identities emerge from large-scale processes of self-organized continuous interaction. This interaction compress relationships of various kinds (including relations some-to-many, and many-to-many, distributed and transient) at different scales. In other words, the intensive use of social media and related sociotechnical practices boost a specific form of collective identity from the micro to macro-level perspective: multitudinous identities.

We propose that the Spanish 15M movement is a prominent example of multitudinous identities found in the myriad of ‘networked social movements’ that have emerged during the last few years (also including others as the Arab Spring or the Occupy movement), and we have focused our analysis on this exemplar. Previous qualitative and quantitative analyses have described how the most innovative 15M ‘patterns’ take the form of a complex organization of interactions which are internally multiple, exhibiting multi-scale recursivity –and still generate an emerging 15M systemic, albeit diffuse and fluid, identity–(Aguilera et al., 2013b; Monterde et al., 2015; Toret et al., 2015).

We propose that multitudinous identities are irreducible to individuals’ aggregated activity, viral flows, or simple online-offline ‘sub-sequence’ schemes. This irreducibility stems from the interrelation of the singularities that compose it. The system cannot be reduced to one-to-one relations among individuals. Instead, interactions should be present and coordinated at different scales: from micro-personal exchanges to transiently coordinated activity of large groups of synchronized actors, up to the evolution of the organization of the macro-

scopic network as a whole. We aim to contribute to explanations about mechanisms that are behind the constitution of multitudinous identities. To do so, we have analysed a large dataset of Twitter activity during the first weeks of the 15M movement in Spain.

First, we have characterized the operational closure of the network of geographically distributed activity using transfer entropy measures, finding that the network arises as a web of distributed dependencies, which continuously changes its structure of information flows while maintaining its operational closure for long periods. This suggests that the operational closure is not being maintained by the conservation of a particular structural configuration, but rather by the maintenance of higher order invariances on top of a highly plastic underlying network.

Second, we analysed statistically salient phase locking values in the coordination between activity at different cities during the emergence and consolidation of the 15M movement, finding that episodes of transient phase locking emerge at different scales and moments, apparently coinciding with important episodes of massive mobilization of activists in the camps and the squares. As we observed, particular phenomena like settling camps, organizing massive spontaneous rallies, or creating large scale peaceful response to police attacks, were triggered by complex sequences of phase-locked coalitions at different timescales. Moreover, we find that phase-locked coalitions are transient, and frequently one coalition is dissolved so others can emerge at a new carrier frequency. This suggests that the activity of the different nodes is coordinated through metastable synchronizing tendencies in a manner that resembles the formation of conscious activity in the brain (Rodriguez et al., 1999; Lutz et al., 2002). Also, we find that the formation of metastable phase-locked coalitions allows transient modes of leadership (which we also analysed for a different case in Monterde et al., 2015) in which different nodes alternate the lead of the coalition at different moments depending on the process of interaction.

These results suggest that synchronization through social media cannot be seen just as a homogeneous reaction to, or amplification, of external events. Instead, global synchronization is triggered by different parts of the network and is tightly related to specific actions and action contexts, while synchronization links change through time displaying a great plasticity. This suggests that synchronization is not the product of a homogeneous resonant media or a simple and sudden aggregation activity, but a manner of choreography resulting from a network of distributed online interactions, where complex resonances emerge and vary depending on the specific action context and articulation of the network. Phase locking episodes at different scales take place interleaved and in coordination with external events creating complex self-regulation loops through the divisions between communal digital communication tools, collective bodies and the environment. We suggest that regulatory mechanisms triggering coalitions at different scales may operate as a ‘metastable circuit breaker’ (see section 3.2.2), capable of disrupting a particular mode of engagement of the work to trigger a new one. For example, this could take the form of organizing particular mobilization, or shifting the focus of collective attention to a new event of interest. This ‘metastable circuit breaker’ is being not

the mechanism that controls the behaviour of the collective body (which is constrained by many external factors) but a mechanism which is able to engage and disengage the collective body in different dynamical fields of interaction.

The analysis of values of TSE complexity (the simultaneous presence of high levels of integration and segregation in the system) for the bursting activity of the different cities shows that the days with more activity and more presence of phase-locked coalitions also present high values of TSE complexity, although the latter seems to change at a slower pace, suggesting a slower transformation of the interactive architecture of the system. Also, analysing temporal long-range correlations revealed how  $1/f$  scale-free dynamics emerge in the moments with more presence of coalitions at all timescales, suggesting that coalitions are organized in a metastable mode. In previous work, analysing the activity of the 15M movement a year from its inception, we found that the most innovative and participatory initiatives also displayed  $1/f$  patterns in their dynamics (Toret et al., 2015; Monterde et al., 2015; Aguilera et al., 2013b), whereas  $1/f$  scaling was lost for similar initiatives when they were organized as more traditional hierarchical organizations as labour unions.

More importantly, we found clear correlations between episodes of strong embodied activity, high synchronization and internal operational closure and the presence of indicators of complex multiscale activity in the spatial and temporal domain in the form of TSE complexity and  $1/f$  scaling. This supports our hypothesis about multitudinous identity, since the constitution of the 15M as an agential autonomous entity seems to be constructed over the organization of its multiplicity emerging from systemic interactions within and between irreducible scales.

Due to the nature of the system analysed – its huge number of dimensions and the fact that we do not have access to variables pertaining to the network structure (e.g. which users are connected or ‘follow’ others) – our analysis cannot characterize the constitution of the system as an autonomous system with the same level of detail we did in Chapter 3 for our homeostatic agent. Nevertheless, the analysis performed suggest that the model of neurodynamic autonomy we proposed there might be extended to cases of large-scale collective autonomy, at least in some cases with intensive use of decentralized communication tools and strong physical embodiment.

In a late interview, *Autopoïese et émergence*, Varela argues that an extension of autopoietic concepts to the social domain might imply negative social and political consequences:

I refuse to apply autopoiesis to the social plane. That might surprise you, but I do so for political reasons. History has shown that biological holism is very interesting and has produced great things, but it has always had its dark side, a black side, each time it's allowed itself to be applied to a social model. There's always slippages toward fascism, toward authoritarian impositions, eugenics, and so on (English translation extracted from Protevi, 2009).

Protevi (2009) connects this authoritarian logic with Varela's experience of the Chilean Civil War, in which during the 3 years of government of Allende a 'polarity created a continual exaggeration of the sense of boundary and territoriality' [min.24-25](Varela, 1978) between the supporters and detractors of Chileans popular front. Varela uses the example of lack of consensus of print media such as newspapers – that portrayed two incompatible realities – as an example of how everyday life had been polarized into two enclosed communities imbued with the effects of politics [min.26](Varela, 1978). From what we have seen here, autonomy in large social systems facilitated by digital communication tools might be more analogous to autonomy in the brain than to autopoietic cellular autonomy, meaning that the emergence of an autonomous system does not have that much to do with the enclosure of material components of the system, i.e. the creation of a physical border, but with the coherence of information flows through the system using mechanisms of transient synchronization. Thus, social autonomy is not constrained by clear boundaries encapsulating the members of a community, but is built through flexible synchronizing and desynchronizing tendencies. The community is not based in an enclosed ideological structure or a particular social segment, but it is built on continuous interactions, allowing different social groups to fluidly engage and disengage in the movement at different moments of time. This form of metastable aggregation can potentially sustain more democratic and participatory forms of coordination as opposed to the authoritarian connotations that Varela connects with an autopoietic organization of society. Different studies and polls show that the 15M movement exhibited wide consensus between the Spanish population (around 75% about its objectives and 50% about its strategy). (Sampedro, 2014) shows that this consensus is transversal among different segments of the population (age, sex, economic status, location) and this consensus has been maintained for various years.

We may argue that it is precisely the use of networked digital communication tool that allows bypassing local material limitations of short range human communication and large-scale unidirectional mass media and break dynamics of polarization and atomization, and instead allow fluid modes of autonomous composition at the larger scale. This hypothesis needs to be further tested. Comparative analysis is to be developed over the different movements that have arisen from starkly different cultural and social substrates as the Occupy Movement in the US, the 'Yo soy 132' movement in Mexico, 2014 protests in Brazil and Turkey, or the Umbrella Movement in Hong Kong. Since most instances of networked movements have just arisen in the last few years, it is too soon to tell if new forms of organization boosted by social media are the patterns of social mobilization to come in the 21st century. Nevertheless, the possibilities of social and political change seem enormous. In short, distributed networked architectures of communication make possible the emergence of forms of social autonomy that need not fall under the uniformity and mass control structured that Varela feared but make possible a multitudinous, complex and diverse identity in continuous collective negotiation. While these lines are being written, in Spain the 15M evolution has displayed a long and dynamic trajectory, facing many challenges and structural mutations and maintaining an immense popular support. The latest of these has promoted the emergence of several citizen

platforms running to and winning the city hall elections in many of the largest Spanish cities in May 2015, all while maintaining a distributed organizational structure closer to the 15M movement than to traditional political parties (Aragón et al., 2015). A systemic understanding and operational treatment of the complex structure and dynamics displayed by networked movements is slowly emerging, potentially enriching our understanding of collective identities in the social media era. We hope to have contributed to this endeavour.

Part IV  
Conclusion



# 7

## Summary and Conclusion

### 7.1 RECAPITULATION

During this thesis, we have explored autonomous agency as a form of organization built over a network of sensorimotor interactions between a self-constituted agent and its surroundings. Furthermore, we have explored the extension of this form of autonomy into collective social agency. This has been accomplished through the combination of conceptual, modelling and experimental contributions to characterize and demarcate the constitution of the autonomous agent and the sensorimotor interactive domain that delimits it in front of its environment.

To conclude the work presented in previous chapters, we would like to present a schematic summary of the main conclusions drawn from each chapter. The models and analytical methods presented along the thesis require deepening in complex and technical details, and it is not always easy to follow the thread from the conclusions of one chapter to the next. In this section we want to recapitulate the obtained results and take a step back to gain a unified view and articulate the relationships between models, results and relevant conclusions. This synthesis will allow us to grasp the significance of our contribution in the current scientific context and in the study of autonomy.

### PART I. CONCEPTUAL AND METHODOLOGICAL REVIEW

CHAPTER I. INTERACTIVE AUTONOMY. Chapter 1 articulated the need to ground an operational theory of the autonomy of living beings as a point of reference for further chapters.

1. We reviewed the thread of historical conceptual notions that have led to modern views of living autonomy, highlighting three main ideas:

- (a) Persistence: a property of some systems that can maintain their behaviour despite the effect of external perturbations.
  - (b) Adaptive self-regulation: systems with this property can maintain certain internal invariants within some viability region, triggering adaptive responses for avoiding the loss of viability.
  - (c) Autonomy: autonomous systems are adaptive systems which self-regulate their own organization as the invariant they maintain constant (i.e. certain relations between its components).
2. We reviewed models proposed during the last few decades to address the theoretical notions summarized above, exploring a parallel historical thread of mechanistic explorations from cybernetics to complex systems science;
    - (a) We analysed the potential of models of distributed self-synchronization as a more suited mode of persistence than mere homeostatic self-regulation of individual components.
    - (b) We described metastable systems as systems able to maintain homeodynamic states, in which certain invariances are kept within the system while allowing them to flexibly switch from one state to another.
    - (c) We reviewed the notion of self-organized criticality as a state of generalized metastability at all scales of the system, allowing regulation between different levels of description of the system without collapsing the dynamics of the system into any of them.
  3. We analysed the application of autonomy at different scales of living: cellular autonomy, organismic autonomy, neurocognitive autonomy and social autonomy, focusing on the last two of them.
  4. We provided a first guiding definition of autonomy expressed in three conditions that a system must meet: individuality through operational closure; interactive asymmetry; and precariousness.

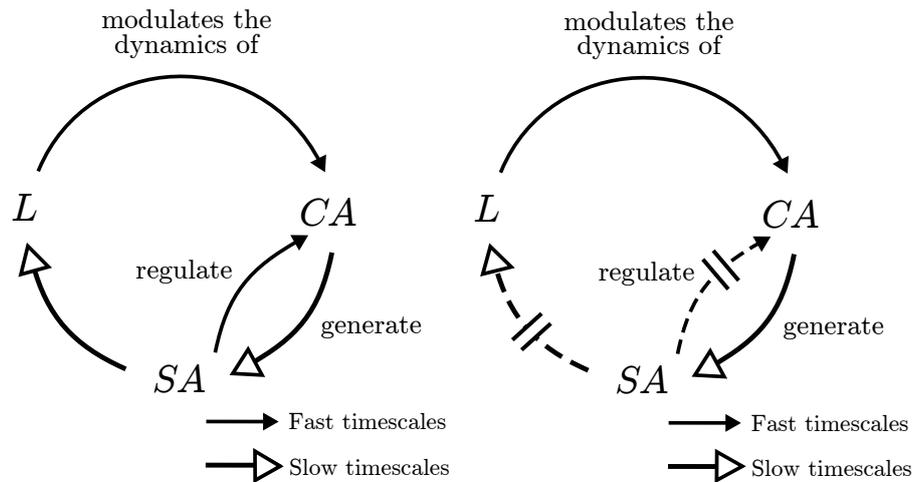
## PART II. SENSORIMOTOR CONSTITUTION OF NEUROCOGNITIVE AUTONOMY

CHAPTER 2. CONSTITUTIVE INTERACTION: THE SITUATED-HKB MODEL. Here, we explored sensorimotor interaction as a constitutive element of neurocognitive autonomy. We argued that much modern neuroscience is focused in the study of the brain in isolation from its environment, while we hypothesized that it is the shared space of sensorimotor interaction where neurocognitive autonomy is constituted.

1. We took a popular minimal model of coordination dynamics, the HKB model, embedding it into a simple robotic model in a gradient-climbing task. In order to address the importance of sensorimotor coupling we analysed two different models: the ‘situated HKB’, which normally performs a gradient-climbing task, and the ‘passively-coupled HKB’, an identical replica of the ‘situated HKB’ controller (with different initial conditions), which instead of being placed into a gradient environment, it is simply fed with an input recorded from the ‘situated HKB’. The comparison of these two conditions allowed us to analyse how real-time coupling affects the case of two similar agents receiving an input with the exact same structure.
2. Due to the minimality of the system, it could be reduced to three differential equations. We solved the system of differential equations and found that the system is controlled by one attractor and one repeller. We discovered that modifying a sensitivity parameter of the agents revealed phase transitions between different behaviours which could only be explained by taking into account the coupling between the neural controller and its environment.
3. We found that when adding perturbations to the system, only the situated agent was able to compensate for these perturbations, while in the passively-coupled agent perturbations were amplified provoking strong fluctuations in the system.
4. Analysing the response of the two agents to a series of different stimulus, we computed a ‘dynamic signature’ of the neural controller of the agent, showing a simplified representation of its phase space. We found that the type of coupling (passive or situated) severely transformed the phase space of the HKB system, and the specificity of functional neural signatures is lost when the agent is passively-coupled.

CHAPTER 3. SITUATED AUTONOMY: ORGANIZATIONAL HOMEOSTASIS IN THE SENSORIMOTOR LOOP. In this chapter, we explored how the constitution of the agent as an adaptive entity that regulates its sensorimotor flows to maintain its own organization. We proposed a specific model of an agent in a behavioural preference task, controlled by a network of oscillators with relational homeostasis. The analysis of the interaction between the macro-variables of the system shows how agency is built from the interaction between emergent patterns at different scales:

1. The system presented a circular operational closure of its processes at three different levels: patterns of sensorimotor interaction (L), oscillatory patterns (CA) and patterns of synaptic connectivity generated by synaptic plasticity (SA). As we observed in Figure 3.17, the autonomous organization of the agent is reflected on a circular closure of information flows between different patterns that constitutes the sensorimotor agent as a coherent unity.



**Figure 3.17: Organizational Closure.** Simplified information flows for the situated (left) and passively-coupled (right) agents.

2. Moreover, the operational closure of the agent created an interactional asymmetry between the agent and environment, being the agent an entity which is affected by bottom-up causal flows from the variability of sensorimotor interactions, and at the same time is able to develop a downward top-down causation over its sensorimotor interaction.
3. The operationally closed loop is broken when the agent is passively-coupled to the environment, showing that the processes of agency are fundamentally interactive and constructed in a sensorimotor space.

**CHAPTER 4. SITUATED AUTONOMY: COMPLEXITY, METASTABILITY AND SELF-ORGANIZED CRITICALITY.** After characterizing the macroscopic regulatory loops that constitute the operational closure of the agent in the previous chapter, we continued exploring the mesoscale patterns that arise in the agent. Specifically, we analysed the presence of spatial and temporal multiscale patterns in the processes that compose the agent:

1. We analysed the oscillatory patterns that emerge under certain behaviours of the agent. We observed that these patterns only arise when the agent is situated in the environment and it is not passively-coupled. We found that situated agents were able to generate coherent internal oscillatory patterns for different behavioural modes, and concurrently they present a high degree of metastability, allowing flexible switching between patterns.
2. We analysed the interaction at different scales of subsets of variables of the system making use of TSE-complexity. We observed that high levels of complexity arise when the agent is both coupled to its environment and plastic mechanisms are present. The

combination of plasticity and situated interaction (in contrast with passively-coupled agents) is what creates a circular interactive asymmetry in the agent.

3. We analysed the presence of long-range interactions in different variables in the form of  $1/f$  scaling in their time series. We found  $1/f$  scaling in variables corresponding to both sensorimotor interaction and oscillatory patterns. We hypothesize that plastic mechanisms are restricted to mesoscale dynamics that allow the connection between sensorimotor microscale fluctuations and oscillatory macroscopic patterns. Moreover,  $1/f$  scaling is only present in both variables when sensorimotor interaction and synaptic plasticity are simultaneously present, being sensorimotor scaling dependent on internal synaptic plasticity and scaling in oscillatory patterns dependent on situated coupling. This cross-dependency suggests that the system as a whole is sustained as a precarious organization by continuous interaction between itself and the environment.

### PART III. INTERACTIVE AUTONOMY IN SOCIAL PROCESSES

CHAPTER 5. SOCIAL AGENCY: INTERACTION-DOMINANT DYNAMICS IN SOCIAL RECIPROCITY. In this chapter, we explored the situation where two autonomous agents interact in a social task. We used the perceptual crossing experimental paradigm as a minimal model for analysing social engagement in a simple interactive task. We found that the social interactive space is constitutive of the ongoing social process:

1. We proposed a novel modification of the perceptual crossing task, in which the player is faced with individual opponents that can be either other human player or programmed bots. This setup allows us to characterize genuine social interaction versus other types of interaction.
2. We showed how previous approaches focused only in behaviour at a specific scale might be missing part of the dynamical richness of social engagement processes. Instead, we defend using fractal scaling analysis to discriminate different modes of engagement.
3. We found that genuine interaction can be discriminated by fractal and multifractal measures only when the interactive space between the two players is analysed. In contrast, social engagement can not be characterized by the individual behaviour of the interacting agents studied in isolation.
4. We concluded that some social processes are constituted in the interactive space and, therefore, the possibility of a social autonomy which takes place at a different plane from individual neurocognitive autonomy is possible.

CHAPTER 6. THE COLLECTIVE MIND: LARGE-SCALE SELF-ORGANIZATION IN MASSIVE SOCIAL NETWORKS. Here we explored the possibility of large-scale social autonomy. We hypothesize that the intensive use of digital communication technologies is opening the door for

new forms of autonomous social organization at scales precluded by previous communication forms. We focused on the analysis of the new wave of ‘networked movements’, constitutively shaped by the use of ICT tools. We proposed the notion of ‘multitudinous identity’ to characterize the forms of autonomous agency of these movements. To explore our hypothesis, we analysed a case study based on the 15M movement in Spain, one of the most representative cases of these network movements, using information theory and dynamical systems tools over a dataset of Twitter activity:

1. We characterized the operational closure of the geographically distant nodes using information transfer measures and comparing them with surrogate data. We showed how the system emerges as a strongly connected component of circular interactions among the geographical parts of the network. Also, we were able to show how the sources of information and the links of the network are continuously changing and reconfiguring the network’s effective connectivity while maintaining the operational closure of the system.
2. We hypothesize that, like in neurodynamic organization, phase-locking phenomena might be playing a role facilitating the emergence of coherent patterns while maintaining a flexible configuration allowing it to rapidly switch between different metastable states. We hypothesized that these types of mechanisms might be behind the constitution of collective agency in multitudinous identities. We used wavelet filtering and phase-locking statistics to characterize the coordination between geographical units at different temporal scales:
  - (a) We characterized the emergence of multiscale phase-locked coalitions during the first days of the 15M movement. Comparing phase-locking statistics of analysed series with surrogate data we found that the emergence of phase-locked coalitions is statistically significant. We found that moments of strong embodiment of the protests are correlated with the emergence of coalitions coordinating the activity at many cities.
  - (b) Analysing a particular case, we exemplify rapid switching between two modes of synchronized coalitions at different carrier frequencies. We analysed the reaction and reconfiguration of the movement to a violent eviction attempt of one of the main protests camps in Barcelona, showing how the system shifts from one coalition synchronized at one specific frequency in which the node of Barcelona is the source of information to all the other nodes mobilizing a mass of peaceful protesters that blocked the eviction, to a new coalition synchronized at a different frequency in which the network organized a country-wide response in the form of protest rallies, where information sources are distributed among several cities.
3. We analysed the presence of spatial and temporal multiscale complexity at different moments of the protests, showing how high values of TSE complexity and scale free

1/f patterns are correlated with both a high presence of synchronized coalitions and a strong embodiment of protests in the streets. We hypothesized how different modes of synchronization might be coordinating the dynamics in social media with the dynamical field of embodied protests, which is what constitutes the emergence of a collective social agent.

## 7.2 CONCLUSION

We have presented different instances of how autonomous selves (be it in neurocognitive or social domains) emerge and maintain fluid, recombinatory modes of organization that sustain a myriad of modes of engagement with the world which are flexibly switched through a metastable transient synchronization. The modes of autonomy that emerge here are different of those of autopoietic organization, in which autonomy is based in the emergence of a material frontier between the self and the external world and operational closure identified as networked self-production. Previous approaches to autonomy inspired in autopoiesis have frequently focused on internal processes leaving agent-environment interactions or coupling as a secondary, surrogate property to be accounted for after the autonomy of the system is determined. Instead, when looking at the autonomy from the point of view of interactions (sensorimotor, social or otherwise), autonomy appears as the maintenance of a network of relational invariants in the coupled activity of the agent-environment system (e.g. long-range dependencies in different parts of the system and its interaction with the environment, or maximization of the informational complexity of interactions between different scales of the system). Needless to say, these interactive forms of autonomy might well be argued to demand a more fundamental or basic form of autonomy, where the internal aspects and material closure need to be granted. Our research does not lessen previous research on autonomy but expands it, bringing new dimensions of autonomy to the foreground providing tools and theoretical progress to study the ways in which autonomy emerges in and through interaction.

The different analyses presented here pose a challenge to one of the most entrenched assumptions of our scientific heritage: that the world and the knower are independent entities. In Varela's words:

If we are forced to admit that cognition cannot be properly understood without common sense, and that common sense is none other than our bodily and social history, then the inevitable conclusion is that the knower and known, mind and world, stand in relation to each other through mutual specification or dependent coordination (Varela et al., 1993, p.150).

We could advance that it is not only history (evolutionary and developmental) but an extended present constituted by coordinations at different scales, that shapes the 'mutual specification'.

In the work presented here, we observe how we cannot distinguish an agent capturing the structure of a pre-given world, but it is the very agent (together with its correspondent world

of meaning) that arises through embodied interaction with its environment. The coupling between the components of the agent and the environment gives rise to transient moments of coherence that structure the agent as an embodied entity. The history of the sequence of these moments of coherence with its correspondent coordinated shaping of internal and external dynamics is what generates an identity that constitutes the autonomous self of the agent.

A robotic agent in a world of two lights does not independently choose a route of behaviour or a light preference. Instead, it is the online regulation between the agent's neurodynamic and behavioural patterns that stabilizes certain precarious patterns of behaviour (e.g. going to light A), which will tend to reinforce themselves until the agent-environment interaction faces a breakdown, in which new patterns of behaviour can arise. The agent as a 'light-switching' entity does not exist *a priori* from its world, but emerges from a specific sequence of interactions with a particular environmental setup. As we observed, if the sensorimotor loop of the agent is broken (e.g. when the agent is fed an input recorded from other agent) or if the agent gets stuck in a particular attractor of behaviour (e.g. when plasticity is stopped), the interactive ecosystems of habits that sustain the mental life of the agents is dissolved, and the autonomous self disappears.

Similarly, a collective subject demanding a series of democratic reforms does not exist prior to its particular embodiment. A process of mass self-communication through online social media allows the coordination of hundreds of thousands of people creating a collective body that enacts moments of collective coherence in a history of demonstrations, camps, campaigns, and so forth. But it is the very process of coordinating this emerging social body that shapes its form: the ideas it sustains, how it communicates with other agents or how it responds to attacks, and so forth. The history of moments of coherence that emerge from distributed coordinations and how the system jumps from one to the other is what constitutes the autonomy of the social subject. A form of social life emerges as long as this autonomy is preserved. If the social substrate that generates the moments of coherence that constitute the system was to be weakened, the collective subject would die as such. Equally, if the emergent configuration of the subject was to be captured in a way in which the movement gets crystallized into a particular configuration or a bureaucratic organization in which individuals cannot influence the moments of coherence that emerge from it, the autonomous self would also cease to exist as such.

Although the project developed here demands further work to provide a robust account of interactive autonomy in the neurocognitive and social domains, we believe that we have offered two contributions to the complex and multilayered research endeavour of unveiling the nature of cognitive autonomy:

1. We have contributed to the development of a modelling route for testing, comparing and validating hypothesis about neurocognitive autonomy. Through two models of neurodynamic controllers embedded into a robotic agent in behavioural tasks, we have provided a proof of concept for some properties described by different hypothe-

ses about cognitive autonomy (e.g. embodiment and situatedness, sensorimotor operational closure or the emergence of a dynamic core). Furthermore, our models allow us to tackle conceptual obstacles precluding the understanding of autonomy in the sensorimotor domain (how operational closure emerges, what its role in generating an agent-environment interactive domain is) and to generate new predictions about the nature of autonomy (multiscale-symmetry breaking, the emergence a double metastable regulatory loop or the coordination of spatial and temporal multiscale complexity).

2. We have explored the extension of our model of cognitive autonomy to the social realm. We analyse two cases from an experimental perspective: the constitution of a collective subject in the sensorimotor domain in a social interactive task, and the emergence of an autonomous social identity in a large-scale technology-mediated social system. We gathered experimental evidence indicating that in some cases social autonomy might emerge based on mechanisms also present in neurocognitive autonomy (transient coalitions of embodied action, the coordination of spatial and temporal multiscale complexity) which might open-up the perspectives for thinking about social and political autonomy with new conceptual and analytical tools.

Moreover, from a methodological point of view, we have shown the potential of indicators of complexity (information theory measures, metastable synchronization or critical scaling) to act as proxies for quantitative and qualitative aspects of autonomy. The study of interaction (between elements, across temporal or spatial scales) highlights the importance of the influence and modulation between micro, meso and macro levels of activity. Autonomy is neither merely a process of closure or isolation from the world nor simply a structural property. Instead, autonomous agency is a process that is deployed at different scales, which are coordinated through circular and asymmetrical causal influences, bringing forth the agent, the world it acts upon, and the interaction domain that connects both. Over and above of these indicators, we have shown how conceptual aspects about how we apply these measures to our models and experimental analyses – for example the operational closure of the network and its precarious cross-dependences or the situated versus passively-coupled comparison – provide a rich perspective for testing specific properties of an autonomous organization. Although indicators of complexity *per se* are not sufficient for detecting if an organism, agent or domain of interaction displays an autonomous organization, they can enrich qualitative intuitions indicating so. The quantitative measures that we have presented bring us closer to be able to measure and characterize the nature of autonomy with greater precision.

Future development of this work should include experimental validation of the results of our model, together with further theoretical and experimental support of how models of neurocognitive autonomy might be useful for understanding large-scale social autonomy, as well as an analysis of which aspects of social autonomy are particular to this domain and cannot be captured with notions from the neurocognitive domain (e.g. some aspects of technologically mediated social dynamics may be drastically different from neurodynamic patterns as digital communication tools overcome some of the limitations of biological networks). Ironically,

this thesis may be complemented with a ‘reverse thesis’, analysing real-world experimental data from sensorimotor processes and exploring simulation models of social agency.

#### TOWARDS A SCIENTIFIC UNDERSTANDING OF AUTONOMY

This thesis is focused on autonomy as a pervasive property of – biological, mental or social – life. Why autonomy? In a world of increasing large-scale connectivity, the technological, political and economic complexity of our societies confronts us with both the promise of a self-organizing collective intelligence freeing and empowering individuals, and the threat of those systemic power dynamics enslaving individual freedom. Cybernetic utopian views tend to present self-organizing mechanisms as ‘neutral’ processes increasing the balance and equilibrium in systems, as if there was some hidden ‘natural order’ waiting to emerge when the components of the system are connected. Nonetheless, far from being neutral, self-organized networks and systems can emerge with little respect for the self-determination of their components. If we think of dramatic financial fluctuations, ecological catastrophes, or examples of collectively sustained forms of networked authoritarianism (e.g. Treré and Barassi, 2015), we can easily imagine a situation in which forms of complex organization threaten the individual’s freedom rather than empowering it.

In the midst of this tension, autonomy appears as a conceptual tool to break the neutrality of systemic perspectives. Delimiting an agential self and its interaction with its surroundings, the components of a system are no longer interchangeable mechanical units but play a different role depending on their position in respect to the self-environment interactive domain. Individual or social action can be evaluated as the action of an active agent modulating its interaction with its environment and other agents, or as being dominated by endogenous or exogenous dynamics. Moreover, an interactive and multiscale view of autonomy allows the existence of agential entities at different levels of description in such a way that being part of a collective with an autonomous identity does not necessarily undermine individual autonomy. Depending on the causal relationships between scales, we can distinguish between processes of empowerment or oppression from the asymmetry between the interactions between agent and its environment. In an increasingly complex and interconnected world, a scientific understanding of autonomy helps us promote the self-determination of individuals and collectives while avoiding their subjugation to pernicious or authoritarian emergent dynamics.

The emergence of a science of adaptive complexity, together with the development of digital technologies (from advances in neuroimaging or computational capacity to digital communication tools) has opened the door to the possibility of an unprecedented understanding of our minds and our world: from nervous systems, organisms, to ecosystems and societies. Indeed, we have at our disposal a methodological and conceptual opportunity to surf the wave of autonomous self-organization in complex open environments. Our notion of autonomy proposes a framework to unveil the loops of interaction that allow living entities to emerge as self-referential, self-modifying, and self-sustaining dynamical processes in front of their worlds. The patterns that they exhibit are neither imposed from outside, nor arise in-

ternally, but are the consequence of an interface of embodied action. This ‘knotty dialectic’ is where living systems are created (from cells, to individuals, to communities), and where they create a world of meaning, will, intention and consciousness. This is where the (neuro-techno) dynamical machinery sustaining the system is infused with the vitality of the ghostly shadow of autonomous agency, which is progressively becoming closer to scientific modelling and quantification.



# References

- Aaronson, S. (2014). Why I Am Not An Integrated Information Theorist (or, The Unconscious Expander).
- Abbott, L. F. and Regehr, W. G. (2004). Synaptic computation. *Nature*, 431(7010):796–803.
- Abeles, M., Bergman, H., Margalit, E., and Vaadia, E. (1993). Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. *Journal of Neurophysiology*, 70(4):1629–1638.
- Acebrón, J. A., Bonilla, L. L., Pérez Vicente, C. J., Ritort, F., and Spigler, R. (2005). The Kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of Modern Physics*, 77(1):137–185.
- Adam, H. and Galinsky, A. D. (2012). Encloded cognition. *Journal of Experimental Social Psychology*, 48(4):918–925.
- Adams, F. R. and Aizawa, K. (2008). *The Bounds of Cognition*, volume 14. Blackwell Pub.
- Aguilera, M., Barandiaran, X. E., Bedia, M. G., and Seron, F. (2015). Self-Organized Criticality, Plasticity and Sensorimotor Coupling. Explorations with a Neurobotic Model in a Behavioural Preference Task. *PLoS ONE*, 10(2):e0117465.
- Aguilera, M., Bedia, M. G., Santos, B. A., and Barandiaran, X. E. (2013a). The situated HKB model: how sensorimotor spatial coupling can alter oscillatory brain dynamics. *Frontiers in Computational Neuroscience*, 7:117.
- Aguilera, M., Morer, I., Barandiaran, X. E., and Bedia, M. G. (2013b). Quantifying political self-organization in social media. fractal patterns in the Spanish 15m movement on twitter. In *Advances in Artificial Life, ECAL*, volume 12, pages 395–402.
- Alimov, S. A. (2001). Green function. In Hazewinkel, M., editor, *Encyclopedia of Mathematics*. Springer.
- Anderson, M. L. (2010). Neural reuse: a fundamental organizational principle of the brain. *The Behavioral and Brain Sciences*, 33(4):245–266; discussion 266–313.

- Anderson, M. L. and Chemero, T. (2013). The problem with brain GUTs: conflation of different senses of "prediction" threatens metaphysical disaster. *The Behavioral and Brain Sciences*, 36(3):204–205.
- Anderson, M. L., Richardson, M. J., and Chemero, A. (2012). Eroding the boundaries of cognition: implications of embodiment(1). *Topics in Cognitive Science*, 4(4):717–730.
- Aragón, P., Volkovich, Y., Laniado, D., and Kaltenbrunner, A. (2015). When a Movement Becomes a Party: The 2015 Barcelona City Council Election. *arXiv:1507.08599 [physics]*. arXiv: 1507.08599.
- Arnellos, A. and Moreno, A. (2015). Multicellular agency: an organizational view. *Biology & Philosophy*, 30(3):333–357.
- Arnellos, A., Moreno, A., and Ruiz-Mirazo, K. (2013). Organizational requirements for multicellular autonomy: insights from a comparative case study. *Biology & Philosophy*, 29(6):851–884.
- Arrow, K. J. (1994). Methodological Individualism and Social Knowledge. *The American Economic Review*, 84(2):1–9.
- Ashby, R. W. (1946). 2072 2073 of Ross Ashby's Journal.
- Ashby, W. R. (1954). *Design for a brain*. New York : Wiley. 00292.
- Ashby, W. R. (1962). Principles of the self-organizing system. *Principles of Self-organization*, pages 255–278.
- Aspray, W. (1990). *Computing Before Computers*. Iowa State University Press.
- Auvray, M., Lenay, C., and Stewart, J. (2009). Perceptual interactions in a minimalist virtual environment. *New Ideas in Psychology*, 27(1):32–47.
- Auvray, M. and Rohde, M. (2012). Perceptual crossing: the simplest online paradigm. *Frontiers in Human Neuroscience*, 6.
- Bak, P., Tang, C., and Wiesenfeld, K. (1987). Self-organized criticality: An explanation of the 1/f noise. *Physical Review Letters*, 59(4):381–384.

- Barandiaran, X. (2008). *Mental life: A naturalized approach to the autonomy of cognitive agents*. Unpublished PhD Thesis, University of the Basque Country, Spain.
- Barandiaran, X. E. (2015). Autonomy and Enactivism; some historical pitfalls and a contemporary re-appraisal (in press). *Topoi*.
- Barandiaran, X. E. and Aguilera, M. (2015). Neurociencia y tecnopolítica: hacia un marco analógico para comprender la mente colectiva del 15m. In Toret, J., editor, *Tecnopolítica y 15M. La potencia de las multitudes conectadas*. Editorial UOC, Barcelona.
- Barandiaran, X. E. and Chemero, A. (2009). Animats in the Modeling Ecosystem. *Adaptive Behavior*, 17(4):287–292.
- Barandiaran, X. E., Di Paolo, E., and Rohde, M. (2009). Defining agency: Individuality, normativity, asymmetry, and spatio-temporality in action. *Adaptive Behavior*, 17(5):367–386.
- Barandiaran, X. E. and Egbert, M. D. (2014). Norm-establishing and norm-following in autonomous agency. *Artificial Life*, 20(1):5–28.
- Barandiaran, X. E. and Moreno, A. (2006). ALife Models as Epistemic Artefacts. In *Artificial Life X: Proceedings of the Tenth International Conference on the Simulation and Synthesis of Living Systems*, pages 513–519. The MIT Press (Bradford Books).
- Barker, E. (1950). Introduction. Gierke, Natural Law and the Theory of Society. E. Barker.
- Bedia, M. G., Aguilera, M., Gómez, T., Larrode, D. G., and Seron, F. (2014). Quantifying long-range correlations and 1/f patterns in a minimal experiment of social interaction. *Frontiers in Psychology*, 5:1281.
- Beer, R. D. (1990). *Intelligence as adaptive behavior: an experiment in computational neuroethology*. Academic Press, Boston.
- Beer, R. D. (1995a). A dynamical systems perspective on agent-environment interaction. *Artificial intelligence*, 72(1-2):173–215.
- Beer, R. D. (1997). The dynamics of adaptive behavior: A research program. *Robotics and Autonomous Systems*, 20:257–289.

- Beer, R. D. and Williams, P. L. (2015). Information Processing and Dynamics in Minimally Cognitive Agents. *Cognitive Science*, 39(1):1–38.
- Beer, S. (1995b). *Brain of the Firm*. Wiley, 2 edition edition.
- Bennett, W. L. and Segerberg, A. (2012). The logic of connective action: Digital media and the personalization of contentious politics. *Information, Communication & Society*, 15(5):739–768.
- Bennett, W. L., Segerberg, A., and Walker, S. (2014). Organization in the crowd: peer production in large-scale networked protests. *Information, Communication & Society*, 17(2):232–260.
- Bentham, J. (1789). An introduction to the principles of morals and legislation. T. Payne, London.
- Bernard, C. (1974). *Lectures on the phenomena of life common to animals and plants*. Thomas.
- Bertalanffy, L. v. (1950). An Outline of General System Theory. *British Journal for the Philosophy of Science*, 1(2):134–165.
- Bertalanffy, L. V. (1969). *General System Theory: Foundations, Development, Applications*. George Braziller, Inc., revised edition edition.
- Bertschinger, N., Olbrich, E., Ay, N., and Jost, J. (2008). Autonomy: An information theoretic perspective. *Biosystems*, 91(2):331–345.
- Bich, L. and Damiano, L. (2012). Life, Autonomy and Cognition: An Organizational Approach to the Definition of the Universal Properties of Life. *Origins of Life and Evolution of Biospheres*, 42(5):389–397.
- Boltzmann, L. (1995). *Lectures on Gas Theory*. Courier Corporation.
- Borge-Holthoefer, J., Perra, N., Gonçalves, B., González-Bailón, S., Arenas, A., Moreno, Y., and Vespignani, A. (2015). The dynamic of information-driven coordination phenomena: a transfer entropy analysis. *arXiv:1507.06106 [nlin, physics:physics]*. arXiv: 1507.06106.

- Borge-Holthoefer, J., Rivero, A., García, I., Cauhé, E., Ferrer, A., Ferrer, D., Francos, D., Iñiguez, D., Pérez, M. P., Ruiz, G., Sanz, F., Serrano, F., Viñas, C., Tarancón, A., and Moreno, Y. (2011). Structural and Dynamical Patterns on Online Social Networks: The Spanish May 15th Movement as a Case Study. *PLoS ONE*, 6(8):e23883.
- Braitenberg, V. (1986). *Vehicles: Experiments in Synthetic Psychology*. MIT Press.
- Brecht, M., Singer, W., and Engel, A. K. (1999). Patterns of synchronization in the superior colliculus of anesthetized cats. *The Journal of Neuroscience* 19:3567–79. *Journal of Neuroscience*, 19:3567–3579.
- Broad, C. D. (1925). *The Mind and Its Place in Nature*. Routledge.
- Brooks, R. A. (1985). A Robust Layered Control System For a Mobile Robot. Technical report, Massachusetts Institute of Technology, Cambridge, MA, USA.
- Brooks, R. A. (1991). Intelligence without representation. 47(1-3):139–159.
- Brosch, M., Bauer, R., and Eckhorn, R. (1997). Stimulus-dependent modulations of correlated high-frequency oscillations in cat visual cortex. *Cerebral Cortex*, 7(1):70–76.
- Buzsáki, G. (2006). *Rhythms of the Brain*. Oxford University Press.
- Buzsáki, G. (2010). Neural syntax: cell assemblies, synapsembles and readers. *Neuron*, 68(3):362–385.
- Calvo, P., Symons, J., and Martín, E. (2012). Beyond “Error-Correction”. *Frontiers in Psychology*, 3.
- Campbell, D. T. (1974). II. ‘Downward Causation’ in Hierarchically Organised Biological Systems. In Ayala, F. J. and Dobzhansky, T. G., editors, *Studies in the Philosophy of Biology: Reduction and Related Problems*, page 179. University of California Press.
- Candón Mena, J. (2013). *Toma la calle, toma las redes: El movimiento 15M en Internet*. Editorial Atrapasueños, Sevilla.
- Cannon, W. B. (1929). Organization for Physiological Homeostasis. *Physiological Reviews*, 9(3):399–431.

- Carruthers, P. (1996). Simulation and self-knowledge: A defence of theory-theory. *Theories of theories of mind*, pages 22–38.
- Carruthers, P. (2006). *The Architecture of the Mind*. Oxford University Press, USA.
- Castells, M. (2007). Communication, Power and Counter-power in the Network Society. *International Journal of Communication*, 1(1):29.
- Castells, M. (2012). *Networks of Outrage and Hope: Social Movements in the Internet Age*. Polity, Cambridge, UK ; Malden, MA, 1 edition edition.
- Castelo-Branco, M., Goebel, R., Neuenschwander, S., and Singer, W. (2000). Neural synchrony correlates with surface segregation rules. *Nature*, 405(6787):685–689.
- Chang, D., Ge, Y., Song, S., Coleman, N., Christensen, J., and Heer, J. (2009). Visualizing the Republic of Letters. *Stanford: Stanford University*. Retrieved April, 21:2014.
- Chemero, A. (2009). *Radical Embodied Cognitive Science*. The MIT Press.
- Chemero, A. and Richardson, M. (2014). Complex Dynamical Systems and Embodiment. In *Routledge Handbook of Embodied Cognition*. Routledge.
- Chialvo, D. R. (2006). The brain near the edge.
- Chiel, H. and Beer, R. (1997). The brain has a body: Adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in neurosciences*, 20(12):553–557.
- Chomsky, N. (1959). A review of BF Skinner's Verbal Behavior. *Language*, 35(1):26–58.
- Christensen, D. W. (2006). The evolutionary origins of volition. In Spurrett, D., Kincaid, H., Ross, D., and Stephens, L., editors, *Distributed Cognition and the Will: Individual Volition and Social Context*. MIT Press.
- Christensen, K. and Moloney, N. R. (2005). Ising Model. In *Complexity and Criticality*, volume 1, pages 115–247. Imperial College Press.
- Christensen, W. D. and Bickhard, M. H. (2002). The process dynamics of normative function. *The Monist*, 85(1):3–28.
- Clarac, F. (2008). Some historical reflections on the neural control of locomotion. *Brain Research Reviews*, 57(1):13–21.

- Clark, A. (1989). Microfunctionalism: Connectionism and the Scientific Explanation of Mental States. In Clark, A., editor, *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*. MIT Press.
- Clark, A. (1997). The dynamical challenge. *Cognitive Science*, 21(4):461–481.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and Brain Sciences*, 36(3):181–204.
- Clark, A. and Chalmers, D. (1998). The extended mind. *analysis*, pages 7–19.
- Cliff, D., Husbands, P., and Harvey, I. (1992). *Evolving Visually Guided Robots*.
- Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*. Wiley-Interscience, New York, 99 edition edition.
- Crick, F. and Koch, C. (1990). Towards a neurobiological theory of consciousness. In *Seminars in the Neurosciences*, volume 2, page 203.
- Damasio, A. (2000). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Mariner Books, New York, 1 edition edition.
- Damasio, A. R. (1989). The Brain Binds Entities and Events by Multiregional Activation from Convergence Zones. *Neural Comput.*, 1(1):123–132.
- De Jaegher, H., Di Paolo, E., and Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in cognitive sciences*, 14(10):441–447.
- Deco, G. and Thiele, A. (2009). Attention - oscillations and neuropharmacology. *European Journal of Neuroscience*, 30(3):347–354.
- Descartes, R. (1984). *Principles of Philosophy*. Springer Science & Business Media.
- Dewey, J. (1896). The reflex arc concept in psychology. *Psychological Review*, 3(4):357–370.
- Dewey, J. (1914). Psychological Doctrine and Philosophical Teaching. *The Journal of Philosophy, Psychology and Scientific Methods*, 11(19):505–511. ArticleType: research-article / Full publication date: Sep. 10, 1914 /.
- Dewey, J. (1922). *Human nature and conduct*. Courier Dover Publications.

- Di Paolo, E. A. (2000). *Homeostatic Adaptation to Inversion of the Visual Field and Other Sensorimotor Disruptions*.
- Di Paolo, E. A. and Iizuka, H. (2008). How (not) to model autonomous behaviour. *Biosystems*, 91(2):409–423.
- Di Paolo, E. A., Noble, J., and Bullock, S. (2000). Simulation models as opaque thought experiments.
- Di Paolo, E. A., Rohde, M., and Iizuka, H. (2008). Sensitivity to social contingency or stability of interaction? Modelling the dynamics of perceptual crossing. *New Ideas in Psychology*, 26(2):278–294.
- Di Paolo, E. A. and Thompson, L. (2014). The Enactive Approach. In Shapiro, L., editor, *The Routledge Handbook of Embodied Cognition*. Routledge.
- Dietrich, E. and Markman, A. B. (2003). Discrete Thoughts: Why Cognition Must Use Discrete Representations. *Mind & Language*, 18(1):95–119.
- Dixon, J. A., Holden, J. G., Mirman, D., and Stephen, D. G. (2012). Multifractal Dynamics in the Emergence of Cognitive Structure. *Topics in Cognitive Science*, 4(1):51–62.
- Dombeck, D. A., Harvey, C. D., Tian, L., Looger, L. L., and Tank, D. W. (2010). Functional imaging of hippocampal place cells at cellular resolution during virtual navigation. *Nature Neuroscience*, 13(11):1433–1440.
- Dombeck, D. A., Khabbaz, A. N., Collman, F., Adelman, T. L., and Tank, D. W. (2007). Imaging large scale neural activity with cellular resolution in awake mobile mice. *Neuron*, 56(1):43–57.
- Dotov, D. G. (2014). Putting reins on the brain. How the body and environment use it. *Frontiers in Human Neuroscience*, 8:795.
- Dotov, D. G., Nie, L., and Chemero, A. (2010). A Demonstration of the Transition from Ready-to-Hand to Unready-to-Hand. *PLoS ONE*, 5(3):e9433.
- Dumas, G., Kelso, J. A. S., and Nadel, J. (2014). Tackling the social cognition paradox through multi-scale approaches. *Cognitive Science*, 5:882.

- Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., and Garnero, L. (2010). Inter-Brain Synchronization during Social Interaction. *PLoS ONE*, 5(8):e12166.
- Dupuy, J.-P. (2009). *On the origins of cognitive science*. A Bradford Book.
- Edelman, G. and Tononi, G. (2001). *A Universe Of Consciousness How Matter Becomes Imagination*. Basic Books.
- Edelman, G. M. (2011). Biology of consciousness. *Frontiers in Consciousness Research*, 2:4.
- Eerland, A., Guadalupe, T. M., and Zwaan, R. A. (2011). Leaning to the left makes the Eiffel Tower seem smaller: posture-modulated estimation. *Psychological Science*, 22(12):1511–1514.
- Elman, J. L., Bates, E. A., and Johnson, M. H. (1998). *Rethinking Innateness: A Connectionist Perspective on Development*. Mit Pr, Cambridge, Mass., edición: revised. edition.
- Engel, A. (2012). Rules Got Rhythm. *Neuron*, 76(4):673–676.
- Engel, A. K., Fries, P., König, P., Brecht, M., and Singer, W. (1999). Temporal Binding, Binocular Rivalry, and Consciousness. *Consciousness and Cognition*, 8(2):128–151.
- Engel, A. K., König, P., and Singer, W. (1991). Direct physiological evidence for scene segmentation by temporal coding. *Proceedings of the National Academy of Sciences*, 88(20):9136–9140.
- Engel, A. K. and Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Sciences*, 5(1):16–25.
- Ernst, M. O. and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433.
- Fan, D., Rich, D., Holtzman, T., Ruther, P., Dalley, J. W., Lopez, A., Rossi, M. A., Barter, J. W., Salas-Meza, D., Herwik, S., Holzhammer, T., Morizio, J., and Yin, H. H. (2011). A Wireless Multi-Channel Recording System for Freely Behaving Mice and Rats. *PLoS ONE*, 6(7):e22033.
- Fang, F. C. and Casadevall, A. (2011). Reductionistic and Holistic Science. *Infection and Immunity*, 79(4):1401–1404.

- Fellman, P. V. (2011). The Nash Equilibrium Revisited: Chaos and Complexity Hidden in Simplicity. In Minai, A. A., Braha, D., and Bar-Yam, Y., editors, *Unifying Themes in Complex Systems*, pages 105–112. Springer Berlin Heidelberg.
- Fernández-Savater, A. (2012). ¿Cómo se organiza un clima? <http://blogs.publico.es/fueradelugar/1438/¿como-se-organiza-un-clima>.
- Finin, T., Fritzson, R., McKay, D., and McEntire, R. (1994). KQML As an Agent Communication Language. In *Proceedings of the Third International Conference on Information and Knowledge Management, CIKM '94*, pages 456–463, New York, NY, USA. ACM.
- Flew, Richard Keith, T. S. (2011). *New Media : An Introduction*. Oxford University Press, Don Mills, Ont., first printing edition edition.
- Floreano, D. (1993). Robogen: A software package for evolutionary control systems. Technical report, Release 1.1. Technical report LabTeCo.
- Fodor, J. A. (1983). *The modularity of mind: an essay on faculty psychology*. MIT Press.
- Freeman, W. J. (2001). *How Brains Make Up Their Minds*. Columbia University Press, 1 edition.
- Freeman, W. J. (2005). A field-theoretic approach to understanding scale-free neocortical dynamics. *Biological cybernetics*, 92(6):350–359.
- Freiwald, W. A., Kreiter, A. K., and Singer, W. (1995). Stimulus dependent intercolumnar synchronization of single unit responses in cat area 17. *Neuroreport*, 6(17):2348–2352.
- Friston, K. J. (2000a). The labile brain. I. Neuronal transients and nonlinear coupling. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 355(1394):215–236.
- Friston, K. J. (2000b). The labile brain. II. Transients, complexity and selection. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 355(1394):237–252.
- Friston, K. J., Daunizeau, J., Kilner, J., and Kiebel, S. J. (2010). Action and behavior: a free-energy formulation. *Biological Cybernetics*, 102(3):227–260.
- Froese, T. (2010). From Cybernetics to Second-Order Cybernetics: A Comparative Analysis of Their Central Ideas. *Constructivist Foundations*, 5(2):75–85.

- Froese, T., Iizuka, H., and Ikegami, T. (2014a). Embodied social interaction constitutes social cognition in pairs of humans: A minimalist virtual reality experiment. *Scientific Reports*, 4(3672).
- Froese, T., Iizuka, H., and Ikegami, T. (2014b). Using minimal human-computer interfaces for studying the interactive development of social awareness. *Frontiers in Cognitive Science*, 5:1061.
- Froese, T. and Stewart, J. (2010). Life After Ashby: Ultrastability and the Autopoietic Foundations of Biological Autonomy. *Cybernetics & Human Knowing*, 17(4):7–49.
- Fuchs, A., Jirsa, V., Haken, H., and Kelso, J. A. S. (1995). Extending the HKB model of coordinated movement to oscillators with different eigenfrequencies. *Biological Cybernetics*, 74(1):21–30.
- Fuchs, C. and Hofkirchner, W. (2010). Autopoiesis and Critical Social Systems Theory. *Advanced Series in Management*, 6:111–129.
- Fuchs, T. (2011). The Brain—A Mediating Organ. *Journal of Consciousness Studies*, 18(7-8):196–221.
- Fuchs, T. and Jaegher, H. D. (2009). Enactive intersubjectivity: Participatory sense-making and mutual incorporation. *Phenomenology and the Cognitive Sciences*, 8(4):465–486.
- Gallagher, S. (2007). Logical and Phenomenological Arguments Against Simulation Theory. In Hutto, D. D. and Ratcliffe, M., editors, *Folk Psychology Re-Assessed*, pages 63–78. Springer Netherlands.
- Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition*, 17(2):535–543.
- Gallese, V. and Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2(12):493–501.
- Gelder, T. v. (1998). The Dynamical Hypothesis in Cognitive Science. *Behavioral and Brain Sciences*, 21(5):615–28.
- Gerbaudo, P. (2013). Online aggregation in the 'mass web'.

- Gerbaudo, P. (2014). The persistence of collectivity in digital protest. *Information, Communication & Society*, 17(2):264–268.
- Gerbaudo, P. and Treré, E. (2015). In search of the ‘we’ of social media activism: introduction to the special issue on social media and protest identities. *Information, Communication & Society*, 0(0):1–7.
- Gibson, J. J. (1966). *The senses considered as perceptual systems*. Houghton Mifflin, Oxford, England.
- Gibson, J. J. (1986). *The Ecological Approach to Visual Perception*. Routledge.
- Gilden, D. L. (2001). Cognitive emissions of 1/f noise. *Psychological Review*, 108(1):33–56.
- Glanville, R. (2002). Second order cybernetics. *Encyclopaedia of Life Support Systems*.
- González-Bailón, S., Borge-Holthoefer, J., Rivero, A., and Moreno, Y. (2011). The Dynamics of Protest Recruitment through an Online Network. *Scientific Reports*, 1.
- Gottlieb, G. (2007). Probabilistic epigenesis. *Developmental Science*, 10(1):1–11.
- Gray, C. M., König, P., Engel, A. K., and Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 338(6213):334–337.
- Grossberg, S. (1988). Adaptive pattern classification and universal recoding. I.: parallel development and coding of neural feature detectors. pages 243–258.
- Gutiérrez, B. (2013a). Spain’s Micro-Utopias: The 15m Movement and its Prototypes (Part 1) | Take The Square.
- Gutiérrez, B. (2013b). Spain’s Micro-Utopias: The 15m Movement and its Prototypes (Part 2) | Take The Square.
- Haken, H. (1978). *Synergetics: An introduction : nonequilibrium phase transitions and self-organization in physics, chemistry, and biology*. Springer-Verlag, enlarged 2nd edition.
- Haken, H. (1995). Some Basic Concepts of Synergetics with Respect to Multistability in Perception, Phase Transitions and Formation of Meaning. In Kruse, D. P. and Stadler, P. D. M., editors, *Ambiguity in Mind and Nature*, number 64 in Springer Series in Synergetics, pages 23–44. Springer Berlin Heidelberg.

- Haken, H., Kelso, J. A. S., and Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, 51(5):347–356.
- Hanczyc, M. M. and Ikegami, T. (2010). Chemical Basis for Minimal Cognition. *Artificial Life*, 16(3):233–243.
- Hardt, M. and Negri, A. (2005). *Multitude: War and democracy in the age of empire*. Penguin.
- Harvey, C. D., Coen, P., and Tank, D. W. (2012). Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature*, 484(7392):62–68.
- Harvey, C. D., Collman, F., Dombeck, D. A., and Tank, D. W. (2009). Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature*, 461(7266):941–946.
- Harvey, I., Di Paolo, E., Wood, R., Quinn, M., Tuci, E., and Iridia, E. T. (2005). Evolutionary Robotics: A New Scientific Tool for Studying Cognition. *Artif. Life*, 11(1-2):79–98.
- Hausdorff, J. and Peng, C. K. (1996). Multiscaled randomness: A possible source of 1/f noise in biology. *Physical Review E*, 54(2):2154–2157.
- Hebb, D. O. (1952). *The organisation of behaviour: a neuropsychological theory*. Wiley.
- Heims, S. J. (1993). *Constructing a Social Science for Postwar America: The Cybernetics Group, 1946-1953*. The MIT Press, Cambridge, Mass.
- Held, R. and Hein, A. (1963). Movement-produced stimulation in the development of visually guided behavior. *Journal of Comparative and Physiological Psychology*, 56:872–876.
- Heylighen, F. (2001). The science of self-organization and adaptivity. *The encyclopedia of life support systems*, 5(3):253–280.
- Hill, K. (2015). Anabaptism and the World of Printing in Sixteenth-Century Germany. *Past & Present*, 226(1):79–114.
- Holland, J. H. (1992). Complex adaptive systems. *Daedalus*, pages 17–30.
- Holst, E. v. (1973). *The behavioural physiology of animals and man: the collected papers of Erich von Holst*. University of Miami Press.

- Husbands, P. and Holland, O. (2012). Warren McCulloch and the British Cyberneticians. *Interdisciplinary Sciences Journal*, 37(3).
- Huygens, C., Hollandsche Maatschappij der Wetenschappen, and Vollgraff, J. A. (1888). *Oeuvres complètes de Christiaan Huygens*, volume 5. La Haye : M. Nijhoff.
- Iberall, A. S. (1977). A field and circuit thermodynamics for integrative physiology. I. Introduction to the general notions. *The American journal of physiology*, 233(5):R171–180.
- Iberall, A. S. and McCulloch, W. S. (1969). The Organizing Principle of Complex Living Systems. *Journal of Basic Engineering*, 91(2):290–294.
- Ihlen, E. A. F. and Vereijken, B. (2010). Interaction-dominant dynamics in human cognition: Beyond 1/f $\alpha$  fluctuation. *Journal of Experimental Psychology: General*, 139(3):436–463.
- Iizuka, H., Ando, H., and Maeda, T. (2009). The Anticipation of Human Behavior Using Parasitic Humanoid. In Jacko, J. A., editor, *Human-Computer Interaction. Ambient, Ubiquitous and Intelligent Interaction*, number 5612 in Lecture Notes in Computer Science, pages 284–293. Springer Berlin Heidelberg.
- Iizuka, H. and Di Paolo, E. A. (2007). Toward Spinozist Robotics: Exploring the Minimal Dynamics of Behavioral Preference. *Adaptive Behavior*, 15(4):359–376.
- Iizuka, H. and Di Paolo, E. A. (2008). Extended homeostatic adaptation: Improving the link between internal and behavioural stability. *Media*, 5040:1–10.
- Iizuka, H., Marocco, D., Ando, H., and Maeda, T. (2013). Experimental study on co-evolution of categorical perception and communication systems in humans. *Psychological research*, 77(1):53–63.
- Iizuka, H. and Paolo, E. D. (2007). Minimal Agency Detection of Embodied Agents. In Costa, F., Rocha, L. M., Costa, E., Harvey, I., and Coutinho, A., editors, *Advances in Artificial Life*, number 4648 in Lecture Notes in Computer Science, pages 485–494. Springer Berlin Heidelberg.
- Iizuka, H. H. and Ando, H. (2012). Emergence of communication and turn-taking behavior in nonverbal interaction. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, J95-A:165–174.

- Ikegami, T. (2007). Simulating active perception and mental imagery with embodied chaotic itinerancy. *Journal of Consciousness Studies*, 14(7):111–125.
- Izhikevich, E. M. (2006). *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. The MIT Press, 1 edition.
- Jaegher, H. D. and Paolo, E. D. (2007). Participatory sense-making. *Phenomenology and the Cognitive Sciences*, 6(4):485–507.
- James, W. (1879). Are We Automata? *Mind*, 4(13):1–22.
- James, W. (1896). Reflex action and theism. In *The will to believe and other essays in popular philosophy*, pages 111–144. Longmans, Green and Co, New York, NY, US.
- Jensen, H. J. (1998). *Self-organized criticality: emergent complex behavior in physical and biological systems*. Cambridge Univ. Press, Cambridge [u.a.
- Jensen, O., Kaiser, J., and Lachaux, J.-P. (2007). Human gamma-frequency oscillations associated with attention and memory. *Trends in Neurosciences*, 30(7):317–324.
- Jirsa, V. K., Friedrich, R., Haken, H., and Kelso, J. A. S. (1994). A theoretical model of phase transitions in the human brain. *Biological Cybernetics*, 71(1):27–35.
- Juarrero, A. (1999). *Dynamics in Action: Intentional Behavior as a Complex System: Alicia Juarrero: 9780262600477: Amazon.com: Books*. MIT Press, Cambridge, MA, USA.
- Juris, J. S. (2008). *Networking Futures: The Movements against Corporate Globalization*. Duke University Press Books, Durham, N.C.
- Juris, J. S. (2012). Reflections on #Occupy Everywhere: Social media, public space, and emerging logics of aggregation. *American Ethnologist*, 39(2):259–279.
- Kant, I. (1987). *Critique of Judgment*. Hackett Publishing.
- Kantelhardt, K. J. (2008). Fractal and Multifractal Time Series.
- Kaplan, D. M. and Bechtel, W. (2011). Dynamical Models: An Alternative or Complement to Mechanistic Explanations? *Topics in Cognitive Science*, 3(2):438–444.
- Kauffman, S. (2003). Molecular autonomous agents. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 361(1807):1089–1099.

- Kauffman, S. A. (2002). *Investigations*. Oxford University Press.
- Kello, C. T. and Van Orden, G. C. (2009). Soft-assembly of sensorimotor function. *Non-linear Dynamics, Psychology, and Life Sciences*, 13(1):57–78.
- Kelso, J. A., Holt, K. G., Rubin, P., and Kugler, P. N. (1981). Patterns of human interlimb coordination emerge from the properties of non-linear, limit cycle oscillatory processes: theory and data. *Journal of motor behavior*, 13(4):226–261.
- Kelso, J. A. S. (1995). *Dynamic Patterns: The Self Organization of Brain and Behaviour*. MIT Press.
- Kelso, J. A. S., de Guzman, G. C., Reveley, C., and Tognoli, E. (2009). Virtual Partner Interaction (VPI): Exploring Novel Behaviors via Coordination Dynamics. *PLoS ONE*, 4(6):e5749.
- Kelso, J. A. S., Del Colle, J., and Schöner, G. (1990). Action-perception as a pattern formation process. In *Attention and performance 13: Motor representation and control*, pages 139–169. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, England.
- Kitzbichler, M. G., Smith, M. L., Christensen, S. R., and Bullmore, E. (2009). Broadband Criticality of Human Brain Network Synchronization. *PLoS Comput Biol*, 5(3):e1000314.
- Kohler, I. (1962). Experiments with goggles. *Scientific American*, 206:62–72.
- Kreiter, A. K. and Singer, W. (1996). Stimulus-dependent synchronization of neuronal responses in the visual cortex of the awake macaque monkey. *The Journal of Neuroscience*, 16(7):2381–2396.
- Krueger, J. (2012). Seeing mind in action. *Phenomenology and the Cognitive Sciences*, 11(2):149–173.
- Kugler, P. N., Kelso, J. A. S., and Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. *Tutorials in motor behavior*, 3:3–47.
- Kuramoto, Y. (1975). Self-entrainment of a population of coupled non-linear oscillators. In Araki, P. H., editor, *International Symposium on Mathematical Problems in Theoretical Physics*, number 39 in Lecture Notes in Physics, pages 420–422. Springer Berlin Heidelberg.

- Kuramoto, Y. (1984). *Chemical oscillations, waves, and turbulence*. Springer Verlag.
- Kuramoto, Y. (2003). *Chemical Oscillations, Waves, and Turbulence*. Dover Publications, Mineola, N.Y, dover ed edition edition.
- Lachaux, J.-P., Rodriguez, E., Le Van Quyen, M., Lutz, A., Martinerie, J., and Varela, F. J. (2000). Studying single-trials of phase synchronous activity in the brain. *International Journal of Bifurcation and Chaos*, 10(10):2429–2439.
- Lakoff, G. and Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. Basic Books.
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., and Alstynne, M. V. (2009). Computational Social. *Science*, 323(5915):721–723.
- Le Van Quyen, M. (2003). Disentangling the dynamic core: a research program for a neurodynamics at the large-scale. *Biological research*, 36(1):67–88.
- Le Van Quyen, M. (2011). The brainweb of cross-scale interactions. *New Ideas in Psychology*, 29(2):57–63.
- Lee, D. N. (1976). A theory of visual control of braking based on information about time-to-collision. *Perception*, 5(4):437 – 459.
- Lenay, C., Stewart, J., Rohde, M., and Amar, A. A. (2011). “You never fail to surprise me”: the hallmark of the Other: Experimental study and simulations of perceptual crossing. *Interaction Studies*, 12(3):373–396.
- Levy, P. (1999). *Collective Intelligence: Mankind’s Emerging World in Cyberspace*. Perseus Books.
- Lewis, M., Fagg, A., and Solidum, A. (1992). Genetic programming approach to the construction of a neural network for control of a walking robot. In , *1992 IEEE International Conference on Robotics and Automation, 1992. Proceedings*, pages 2618–2623 vol.3.
- Linderman, M. D., Gilja, V., Santhanam, G., Afshar, A., Ryu, S., Meng, T. H., and Shenoy, K. V. (2006). Neural recording stability of chronic electrode arrays in freely behaving primates. *Conference proceedings: ... Annual International Conference of the IEEE Engineering*

*in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 1:4387–4391.

Linkenkaer-Hansen, K., Nikouline, V. V., Palva, J. M., and Ilmoniemi, R. J. (2001). Long-Range Temporal Correlations and Scaling Behavior in Human Brain Oscillations. *The Journal of Neuroscience*, 21(4):1370–1377.

List, C. and Pettit, P. (2011). *Group agency: The possibility, design, and status of corporate agents*. Oxford University Press Oxford.

Livingstone, M. S. (1996). Oscillatory firing and interneuronal correlations in squirrel monkey striate cortex. *Journal of Neurophysiology*, 75(6):2467–2485.

Llinás, R., Ribary, U., Contreras, D., and Pedroarena, C. (1998). The neuronal basis for consciousness. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 353(1377):1841–1849.

Llinás, R. R. and Iberall, A. (1977). A global model of neuronal command-control systems. *Bio Systems*, 8(4):233–235.

Longo, G. and Montevil, M. (2012). The Inert vs. the Living State of Matter: Extended Criticality, Time Geometry, Anti-Entropy - An Overview. *Frontiers in Physiology*, 3.

Lotka, A. (1920). Analytical note on certain rhythmic relations in organic systems. *Proceedings of the National Academy of Sciences of the United States of America*, 6(7):410.

Luhmann, N. (1986). The autopoiesis of social systems. *Sociocybernetic paradoxes*, pages 172–192.

Luisi, P. L. (2003). Autopoiesis: a review and a reappraisal. *Naturwissenschaften*, 90(2):49–59.

Lutz, A., Lachaux, J.-P., Martinerie, J., and Varela, F. J. (2002). Guiding the study of brain dynamics by using first-person data: Synchrony patterns correlate with ongoing conscious states during a simple visual task. *Proceedings of the National Academy of Sciences*, 99(3):1586–1591.

Maes, P., editor (1991). *Designing Autonomous Agents: Theory and Practice from Biology to Engineering and Back*. A Bradford Book, Cambridge, Mass, mit press ed edition edition.

- Malo, M. and Pérez, D. (2012). Latidos: el 15m y la revuelta. In *Democracia Distribuida. Miradas de la Universidad Nómada al 15M*.
- Markram, H. (2006). The Blue Brain Project. *Nature Reviews Neuroscience*, 7(2):153–160.
- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Henry Holt and Co., Inc., New York, NY, USA.
- Maturana, H. R. (1970). Biology of Cognition. In *Autopoiesis and cognition: The realization of the living*, pages 1–58. Dordrecht, Holland: Kluwer Academic.
- Maturana, H. R. and Varela, F. J. (1980). *Autopoiesis and Cognition: The Realization of the Living*. D. Reidel Publishing Company, 1st edition.
- Mazzocchi, F. (2012). Complexity and the reductionism–holism debate in systems biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 4(5):413–427.
- McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133.
- McDonald, K. (2002). From Solidarity to Fluidarity: Social movements beyond 'collective identity'—the case of globalization conflicts. *Social Movement Studies*, 1(2):109–128.
- McNeill, W. H. (2008). *Keeping Together in Time: Dance and Drill in Human History*. ACLS Humanities E-Book, New York, USA.
- Medina, E. (2006). Designing Freedom, Regulating a Nation: Socialist Cybernetics in Allende's Chile. *Journal of Latin American Studies*, 38(03):571–606.
- Merleau-Ponty, M. (1942). *The structure of behavior*. Beacon Press.
- Milner, P. M. (1974). A model for visual shape recognition. *Psychological Review*, 81(6):521–535.
- Mingers, J. (1992). The Problems of Social Autopoiesis. *International Journal of General Systems*, 21(2):229–236.
- Mirowski, P. (2001). *Machine Dreams: Economics Becomes a Cyborg Science*. Cambridge University Press.

- Moioli, R. C., Vargas, P. A., and Husbands, P. (2010). Exploring the Kuramoto model of coupled oscillators in minimally cognitive evolutionary robotics tasks. In *2010 IEEE Congress on Evolutionary Computation (CEC)*, pages 1–8. IEEE.
- Mongillo, G., Barak, O., and Tsodyks, M. (2008). Synaptic Theory of Working Memory. *Science*, 319(5869):1543–1546.
- Monterde, A., Calleja-López, A., Aguilera, M., Barandiaran, X. E., and Postill, J. (2015). Multitudinous identities: a qualitative and network analysis of the 15m collective identity. *Information, Communication & Society*, 18(8):930–950.
- Monterde, A. and Postill, J. (2014). Mobile ensembles: The uses of mobile phones for social protest by Spain’s indignados. In *The Routledge companion to mobile media*. New York: Routledge, pages 429–438. Goggin, Gerard; Hjorth, Larissa.
- Moreno, A., Etxeberria, A., and Umerez, J. (1994). Universality without matter? pages 406–410. MIT Press.
- Moreno, A., Ruiz-mirazo, K., and Bar, X. (2011). The Impact of the Paradigm of Complexity on the Foundational Frameworks of Biology and Cognitive Science. In *Philosophy of Complex Systems*. Elsevier.
- Morozov, E. (2012). *The Net Delusion: The Dark Side of Internet Freedom*. PublicAffairs.
- Murray, L. and Trevarthen, C. (1985). Emotional regulation of interactions between two-month-olds and their mothers. In *T. M. Field & N. A. Fox (Eds.), Social Perception in Infants*. Norwood, NJ, pages 177–197.
- Myin, E. (2003). An account of color without a subject? *Behavioral and Brain Sciences*, 26(01):42–43.
- Müller, V., Sängler, J., and Lindenberger, U. (2013). Intra- and Inter-Brain Synchronization during Musical Improvisation on the Guitar. *PLoS ONE*, 8(9):e73852.
- Nadel, J., Carchon, I., Kervella, C., Marcelli, D., and Réserbat-Plantey, D. (1999). Expectancies for social contingency in 2-month-olds. *Developmental Science*, 2(2):164–173.
- Nolfi, S. and Floreano, D. (2000). *Evolutionary robotics: the biology, intelligence, and technology of self-organizing machines*. MIT Press, Cambridge, Mass.

- Noë, A. (2004). *Action in perception*. MIT Press.
- O’Keefe, J. and Recce, M. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus*, 3(3):317–330.
- Oliveira, H. M. and Melo, L. V. (2015). Huygens synchronization of two clocks. *Scientific Reports*, 5:11548.
- O’Regan, J. and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24:939–1031.
- Oullier, O. and Kelso, J. A. S. (2009a). Coordination from the perspective of Social Coordination Dynamics. *Phase Transitions*, (Umr 6149):1–31.
- Oullier, O. and Kelso, J. A. S. (2009b). Social Coordination, from the Perspective of Coordination Dynamics. In Ph. D, R. A. M., editor, *Encyclopedia of Complexity and Systems Science*, pages 8198–8213. Springer New York.
- Ozouf, M. and Sheridan, A. (1991). *Festivals and the French Revolution*. Harvard University Press.
- O’Regan, J. K., Myin, E., and Noë, A. (2005). Skill, corporality and alerting capacity in an account of sensory consciousness. In Laureys, S., editor, *Progress in Brain Research*, volume 150 of *The Boundaries of Consciousness: Neurobiology and Neuropathology*, pages 55–592. Elsevier.
- Panksepp, J. (1997). The periconscious substrates of consciousness: affective states and the evolutionary origins of the self. *Journal of Consciousness Studies*, 5(5-6):566–582.
- Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford University Press.
- Panzarasa, P. and Jennings, N. R. (2001). Negotiation and joint commitments in multi-agent systems. *Sozionik aktuell*, 3:65–81.
- Peng, C. K., Hausdorff, J. M., and Goldberger, A. L. (2000). Fractal mechanisms in neuronal control: human heartbeat and gait dynamics in health and disease. In *Self-Organized Biological Dynamics and Nonlinear Control*. Cambridge University Press.

- Peña-López, I., Congosto, M., and Aragón, P. (2014). Spanish Indignados and the evolution of the 15m movement on Twitter: towards networked para-institutions. *Journal of Spanish Cultural Studies*, 15(1-2):189–216.
- Phillips, W. and Singer, W. (1997). In search of common foundations for cortical computation. *Behavioral and Brain Sciences*, 20(4):657–722.
- Pickering, A. (2010). *The cybernetic brain: Sketches of another future*. University of Chicago Press.
- Pigliucci, M. (2014). Between holism and reductionism: a philosophical primer on emergence. *Biological Journal of the Linnean Society*, 112(2):261–267.
- Pinker, S. (1997). *How the Mind Works*. W. W. Norton & Company, New York.
- Pittendrigh, C. S. (1960). Circadian Rhythms and the Circadian Organization of Living Systems. *Cold Spring Harbor Symposia on Quantitative Biology*, 25:159–184.
- Port, R. F. and Gelder, T. V. (1995). *Mind as motion: Explorations in the Dynamics of Cognition*. MIT Press.
- Powers, W. (1973). *Behavior: The Control of Perception*. Aldine Transaction, illustrated edition.
- Prigogine, I. and Nicolis, G. (1971). Biological order, structure and instabilities. *Quarterly Reviews of Biophysics*, 4(2-3):107–148.
- Prinz, J. (2006). Putting the brakes on enactive perception. *Psyche*, 12(1):1–19.
- Protevi, J. (2009). Beyond Autopoiesis: Inflections of Emergence and Politics in the Work of Francisco Varela. In Clarke, B. and Hansen, M. B. N., editors, *Emergence and Embodiment: New Essays on Second-Order Systems Theory*. Duke University Press.
- Prut, Y., Vaadia, E., Bergman, H., Haalman, I., Slovin, H., and Abeles, M. (1998). Spatiotemporal Structure of Cortical Activity: Properties and Behavioral Relevance. *Journal of Neurophysiology*, 79(6):2857–2874.
- Pérez, B. and Gil, J. (2014). ¿Mercantilización o Revolución? Reflexiones en torno a la figura del streamer como nuevo sujeto prosumidor. In *15MP2P. Una mirada transdisciplinar del 15M*, pages 294–301. IN3, Barcelona.

- Quinton, A. (1975). The presidential address: Social objects. pages 1–viii. JSTOR.
- Ratcliffe, M. (2006). Folk psychology' is not folk psychology. *Phenomenology and the Cognitive Sciences*, 5(1):31–52.
- Reatto, L. (2007). Phase transitions: A complex view of criticality. *Nature Physics*, 3(9):594–596.
- Reed, E. S. (1996). *Encountering the world: Toward an ecological psychology*, volume vii. Oxford University Press, New York, NY, US.
- Rheingold, H. (2007). *Smart Mobs: The Next Social Revolution*. Basic Books.
- Riesen, A. H. and Aarons, L. (1959). Visual movement and intensity discrimination in cats after early deprivation of pattern vision. *Journal of Comparative and Physiological Psychology*, 52(2):142–149.
- Riley, M. A., Richardson, M., Shockley, K., and Ramenzoni, V. C. (2011). Interpersonal synergies. *Movement Science and Sport Psychology*, 2:38.
- Rizzolatti, G. and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.*, 27:169–192.
- Rizzolatti, G. and Sinigaglia, C. (2007). *Mirrors in the Brain: How Our Minds Share Actions and Emotions*. OUP Oxford.
- Rodriguez, E., George, N., Lachaux, J.-P., Martinerie, J., Renault, B., and Varela, F. J. (1999). Perception's shadow: long-distance synchronization of human brain activity. *Nature*, 397(6718):430–433.
- Rohde, M. and Di Paolo, E. (2006). "Value Signals" and Adaptation: An Exploration in Evolutionary Robotics. Technical Report 584, University of Sussex, Brighton, UK.
- Rosenblueth, A., Wiener, N., and Bigelow, J. (1943). Behavior, purpose and teleology. *Philosophy of science*, 10(1):18–24.
- Rosslénbroich, B. (2014). *On the Origin of Autonomy*, volume 5 of *History, Philosophy and Theory of the Life Sciences*. Springer International Publishing, Cham.
- RTVE (2011). Más de seis millones de españoles han participado en el Movimiento 15m.

- Rubinov, M., Sporns, O., Thivierge, J.-P., and Breakspear, M. (2011). Neurobiologically Realistic Determinants of Self-Organized Criticality in Networks of Spiking Neurons. *PLoS Comput Biol*, 7(6):e1002038.
- Ruiz-Mirazo, K. and Mavelli, F. (2008). On the way towards ‘basic autonomous agents’: Stochastic simulations of minimal lipid–peptide cells. *Biosystems*, 91(2):374–387.
- Ruiz-Mirazo, K. and Moreno, A. (2004). Basic autonomy as a fundamental step in the synthesis of life. *Artificial Life*, 10(3):235–259.
- Ruiz-Mirazo, K. and Moreno, A. (2011). Autonomy in evolution: from minimal to complex life. *Synthese*, 185(1):21–52.
- Ruiz-Mirazo, K., Peretó, J., and Moreno, A. (2004). A Universal Definition of Life: Autonomy and Open-Ended Evolution. *Origins of life and evolution of the biosphere*, 34(3):323–346.
- Ruiz-Mirazo, K., Umerez, J., and Moreno, A. (2007). Enabling conditions for ‘open-ended evolution’. *Biology & Philosophy*, 23(1):67–85.
- Rumelhart, D. E. and McClelland, J. L. (1986). Parallel distributed processing. *Explorations in the microstructure of cognition*, 2.
- Runciman, D. (1997). *Pluralism and the Personality of the State*. Cambridge University Press.
- Rybarsch, M. and Bornholdt, S. (2012). Self-organized criticality in neural network models. arXiv e-print 1212.3106.
- Sampedro, E. (2014). La transversalidad del del 15m entre la ciudadanía. In *15MP2P. Una mirada transdisciplinar del 15M*, pages 470–489. IN3, Barcelona.
- Santhanam, G., Linderman, M., Gilja, V., Afshar, A., Ryu, S., Meng, T., and Shenoy, K. (2007). HermesB: A Continuous Neural Recording System for Freely Behaving Primates. *IEEE Transactions on Biomedical Engineering*, 54(11).
- Santos, B., Barandiaran, X., Husbands, P., Aguilera, M., and Bedia, M. (2012). Sensorimotor coordination and metastability in a situated HKB model. *Connection Science*, 24(4):143–161.

- Santos, B. A., Barandiaran, X. E., and Husbands, P. (2011). Metastable dynamical regimes in an oscillatory network modulated by an agent's sensorimotor loop. In *2011 IEEE Symposium on Artificial Life (ALIFE)*, pages 124–131. IEEE.
- Schmidt, R. C., Carello, C., and Turvey, M. T. (1990). Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. *Journal of Experimental Psychology. Human Perception and Performance*, 16(2):227–247.
- Schreiber, T. and Schmitz, A. (1996). Improved Surrogate Data for Nonlinearity Tests. *Physical Review Letters*, 77(4):635–638.
- Scott, D. W. (2012). Multivariate Density Estimation and Visualization. In Gentle, J., Hardle, W., and Mori, Y., editors, *Handbook of Computational Statistics*, Springer Handbooks of Computational Statistics, pages 549–569. Springer Berlin Heidelberg.
- Serrano, E. (2012). New technologies, social networks and media: Spanish Revolution. *Romanian Journal of Journalism & Communication*, 7(3/4):27–34.
- Seth, A. K. (2007). Measuring Autonomy by Multivariate Autoregressive Modelling. In Costa, F. A. e., Rocha, L. M., Costa, E., Harvey, I., and Coutinho, A., editors, *Advances in Artificial Life*, number 4648 in Lecture Notes in Computer Science, pages 475–484. Springer Berlin Heidelberg.
- Seth, A. K., Izhikevich, E., Reeke, G. N., and Edelman, G. M. (2006). Theories and measures of consciousness: An extended framework. *Proceedings of the National Academy of Sciences*, 103(28):10799–10804.
- Shanahan, M. (2010). Metastable chimera states in community-structured oscillator networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 20(1):013108.
- Silberstein, M. and Chemero, A. (2011). Dynamics, Agency and Intentional Action. *Humana.Mente*, 15:1–19.
- Silberstein, M. and Chemero, A. (2012). Complexity and Extended Phenomenological-Cognitive Systems. *Topics in Cognitive Science*, 4(1):35–50.
- Simon, H. A. (1996). *The Sciences of the Artificial*. MIT Press.

- Singer, W. (1993). Synchronization of Cortical Activity and its Putative Role in Information Processing and Learning. *Annual Review of Physiology*, 55(1):349–374.
- Singer, W. (1999). Neuronal synchrony: a versatile code for the definition of relations? *Neuron*, 24(1):49–65, III–125.
- Singer, W. (2007). Binding by synchrony. *Scholarpedia*, 2(12):1657.
- Skarda, C. A. and Freeman, W. J. (1987). How brains make chaos in order to make sense of the world. *Behavioral and Brain Sciences*, 10(02):161.
- Sole, R. and Goodwin, B. (2008). *Signs Of Life How Complexity Pervades Biology: How Complexity Pervades Biology*. Basic Books.
- Solé, R. V., Munteanu, A., Rodriguez-Caso, C., and Macía, J. (2007). Synthetic protocell biology: from reproduction to computation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 362(1486):1727–1739.
- Sornette, D. (2006). Endogenous versus exogenous origins of crises. In *Extreme events in nature and society*, pages 95–119. Springer.
- Spinoza, B. (1883). *Ethics Part III. On the Origin and Nature of the Emotions*. Library of Alexandria.
- Steels, L. (1990). Towards a theory of emergent functionality. In *Proceedings of the first international conference on simulation of adaptive behavior on From animals to animats*, pages 451–461, Cambridge, MA, USA. MIT Press.
- Strogatz, S. (2001). *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry and engineering*. Perseus Books Group.
- Strogatz, S. H. (2004). *Sync: How Order Emerges From Chaos In the Universe, Nature, and Daily Life*. Hyperion, New York, reprint edition edition.
- Suzuki, K., Wakisaka, S., and Fujii, N. (2012). Substitutional Reality System: A Novel Experimental Platform for Experiencing Alternative Reality. *Scientific Reports*, 2.
- Thelen, E. (1996). *A Dynamic Systems Approach to the Development of Cognition and Action*. MIT Press.

- Thelen, E. and Smith, L. (1994). *A dynamic systems approach to the development of cognition and action*. Number 2. MIT Press.
- Thompson, E. (2001). Empathy and consciousness. *Journal of consciousness studies*, 8(5-7):5-7.
- Thompson, E. and Varela, F. J. (2001). Radical embodiment: neural dynamics and consciousness. *Trends in cognitive sciences*, 5(10):418-425.
- Tognoli, E. and Kelso, J. A. S. (2009). Brain coordination dynamics: True and false faces of phase synchrony and metastability. *Progress in Neurobiology*, 87(1):31-40.
- Tognoli, E. and Kelso, J. A. S. (2013). Enlarging the scope: grasping brain complexity. *arXiv:1310.7277 [q-bio]*.
- Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, 5(1):42.
- Tononi, G. and Edelman, G. M. (1998). Consciousness and complexity. *Science (New York, N.Y.)*, 282(5395):1846-1851.
- Tononi, G., Sporns, O., and Edelman, G. M. (1994). A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences*, 91(11):5033-5037.
- Tononi, G., Sporns, O., and Edelman, G. M. (1996). A complexity measure for selective matching of signals by the brain. *Proceedings of the National Academy of Sciences*, 93(8):3422-3427.
- Toret, J., Calleja-López, A., Marín, O., Aragón, P., Aguilera, M., Barandiaran, X., and Monterde, A. (2015). *Tecnopolítica y 15M: La potencia de las multitudes conectadas*. Barcelona: Editorial UOC.
- Torre, K. and Wagenmakers, E.-J. (2009). Theories and models for 1/f(beta) noise in human movement science. *Human movement science*, 28(3):297-318.
- Treré, E. and Barassi, V. (2015). Net-authoritarianism? How web ideologies reinforce political hierarchies in the Italian 5 Star Movement. *Journal of Italian Cinema & Media Studies*, 3(3):287-304.

- Trevarthen, C. (1994). The self born in intersubjectivity: The psychology of an infant communicating. In *The perceived self*, Emory Symposia in Cognition. Cambridge University Press.
- Turvey, M. T. and Fonseca, S. T. (2014). The medium of haptic perception: a tensegrity hypothesis. *Journal of Motor Behavior*, 46(3):143–187.
- Usher, M., Stemmler, M., and Olami, Z. (1995). Dynamic Pattern Formation Leads to 1/f Noise in Neural Populations. *Physical Review Letters*, 74(2):326–329.
- Van Orden, G., Hollis, G., and Wallot, S. (2012). The blue-collar brain. *Fractal Physiology*, 3:207.
- Van Orden, G. C., Holden, J. G., and Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General*, 132(3):331–350.
- Van Orden, G. C., Holden, J. G., and Turvey, M. T. (2005). Human Cognition and 1/f Scaling. *Journal of Experimental Psychology: General*, 134(1):117–123.
- Van Orden, G. C. V. and Holden, J. G. (2002). Intentional Contents and Self-Control. *Ecological Psychology*, 14(1-2):87–109.
- Varela, F., Lachaux, J. P., Rodriguez, E., and Martinerie, J. (2001a). The brainweb: phase synchronization and large-scale integration. *Nature reviews. Neuroscience*, 2(4):229–239.
- Varela, F., Lachaux, J. P., Rodriguez, E., and Martinerie, J. (2001b). The brainweb: phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, 2(4):229–239.
- Varela, F., Maturana, H., and Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *Biosystems*, 5(4):187–196.
- Varela, F. and Thompson, E. (2003). Neural Synchrony and the Unity of Mind: A Neurophenomenological Perspective. In *The Unity of Consciousness*. Oxford University Press.
- Varela, F. J. (1978). The Cultural Contradictions of Power.
- Varela, F. J. (1979). *Principles of Biological Autonomy*. Appleton & Lange.
- Varela, F. J. (1981). Autonomy and autopoiesis. *Self-organizing systems*, pages 14–23.

- Varela, F. J. (1988). Structural Coupling and the Origin of Meaning in a Simple Cellular Automation. In Sercarz, E. E., Celada, F., Mitchison, N. A., and Tada, T., editors, *The Semiotics of Cellular Communication in the Immune System*, number 23 in NATO ASI Series, pages 151–161. Springer Berlin Heidelberg.
- Varela, F. J. (1991). Organism: A Meshwork of Selfless Selves. In Tauber, A. I., editor, *Organism and the Origins of Self*, number 129 in Boston Studies in the Philosophy of Science, pages 79–107. Springer Netherlands.
- Varela, F. J. (1992). Making It Concrete: Before, During and After Breakdowns. In Ogilvy, J. A., editor, *Revisioning Philosophy*. SUNY Press.
- Varela, F. J. (1995). Resonant cell assemblies: a new approach to cognitive functions and neuronal synchrony. *Biological Research*, 28(1):81–95.
- Varela, F. J. (1997). Patterns of life: Intertwining identity and cognition. *Brain and cognition*, 34(1):72–87.
- Varela, F. J. (2000). *El Fenómeno de la vida*. Dolmen Esayo, Santiago, Chile.
- Varela, F. J. and Coutinho, A. (1991). Second generation immune networks. *Immunology Today*, 12(5):159–166.
- Varela, F. J., Thompson, E. T., and Rosch, E. (1993). *The Embodied Mind: Cognitive Science and Human Experience*. Mit Press Ltd, Cambridge, Mass., edición: new ed. edition.
- Virgo, N., Egbert, M. D., and Froese, T. (2011). The Role of the Spatial Boundary in Autopoiesis. In Kampis, G., Karsai, I., and Szathmáry, E., editors, *Advances in Artificial Life. Darwin Meets von Neumann*, number 5777 in Lecture Notes in Computer Science, pages 240–247. Springer Berlin Heidelberg.
- von der Malsburg, C. (1981). The Correlation Theory of Brain Function. Departmental Technical Report, MPI.
- von der Malsburg, C. (1995). Binding in models of perception and brain function. *Current Opinion in Neurobiology*, 5(4):520–526.
- Wagenmakers, E. J., Farrell, S., and Ratcliff, R. (2004). Estimation and interpretation of  $1/\alpha$  noise in human cognition. *Psychonomic Bulletin & Review*, 11(4):579–615.

- Wagenmakers, E. J., van der Maas, H. L. J., and Farrell, S. (2012). Abstract Concepts Require Concrete Models: Why Cognitive Scientists Have Not Yet Embraced Nonlinearly Coupled, Dynamical, Self-Organized Critical, Synergistic, Scale-Free, Exquisitely Context-Sensitive, Interaction-Dominant, Multifractal, Interdependent Brain-Body-Niche Systems. *Topics in Cognitive Science*, 4(1):87–93.
- Walter, W. G. (1963). *The Living Brain*. W. W. Norton and Company, Inc.
- Ward, L. M. (2011). The thalamic dynamic core theory of conscious experience. *Consciousness and Cognition*, 20(2):464–486.
- Warren, W. (2006). The dynamics of perception and action. *Psychological review*, 113(2):358.
- Weaver, W. (1991). Science and complexity. In *Facets of Systems Science*, pages 449–456. Springer.
- Welch, P. D. (1967). The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73.
- Werner, G. (2007a). Brain dynamics across levels of organization. *Journal of Physiology-Paris*, 101(4–6):273–279.
- Werner, G. (2007b). Metastability, criticality and phase transitions in brain and its models. *Biosystems*, 90(2):496–508.
- Wiener, N. (1948). *Cybernetics; or, Control and communication in the animal and the machine*. J. Wiley.
- Wiener, N. (1949). *Extrapolation, interpolation, and smoothing of stationary time series*, volume 2. MIT press Cambridge, MA.
- Wiener, N. (1965). *Cybernetics Or Control and Communication in the Animal and the Machine*. MIT Press.
- Wiener, N. (1988). *The Human Use of Human Beings: Cybernetics and Society*. Da Capo Press.
- Wiener, N. and Schadé, J. P. (1965). *Cybernetics of the Nervous system*. Elsevier.

Wilson, A. D. and Golonka, S. (2013). Embodied cognition is not what you think it is. *Cognitive Science*, 4:58.

Winfrey, A. T. (1967). Biological rhythms and the behavior of populations of coupled oscillators. *Journal of Theoretical Biology*, 16(1):15–42.

Wooldridge, M. J. (1992). *The Logical Modelling of Computational Multi-Agent Systems*.



**T**HIS THESIS WAS TYPESET using  $\text{\LaTeX}$ , originally developed by Leslie Lamport and based on Donald Knuth's  $\text{\TeX}$ . The body text is set in 11 point Egenolff-Berner Garamond, a revival of Claude Garamont's humanist typeface. Math text is set in Gara and figure captions are set in Lato. The above illustration, *Science Experiment 02*, was created by Ben Schlitter and released under [CC BY-NC-ND 3.0](#). The format and style of this PhD dissertation were adapted from the *Dissertate* template, released under the permissive [AGPL](#) license, which can be found online at [github.com/asm-products/Dissertate](https://github.com/asm-products/Dissertate).