OUTAGENCIES
Varieties of autonomous agency across living, humanimal and technical systems

Project Description

The project "OUTAGENCIES: Varieties of autonomous agency across living, humanimal, and technical systems", lead by Xabier E. Barandiaran, with ID PID2023-147251NB-I00 was awarded by the Spanish Ministry of Science and Innovation for the period 01/09/2024-31/08/2028 within the 2023 Call for "R&D&I Projects," as part of the framework of the State Programs for the Generation of Knowledge and Scientific and Technological Strengthening of the R&D&I System Oriented to the Challenges of Society.

IAS-Research
Centre for Life, Mind, and Society

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN

Table of Contents

SUMMARY **3**
Summary of the project 3
Keywords 4
Summary of expected impact 4
1. Background and state of the art **7**
1.1. Motivation and Relevance 7
1.2. Working hypothesis and previous contributions of the research team 9
2. Objectives **13**
2.1. WP1: Living matters 13
2.2. WP2: Humanimal varieties 14
2.3. WP3: Artificial Agencies 14
2.4. WP4: Acts of integration 15
3. Methodology **17**
4. Work Plan **19**
5. Previous results of the team in the theme of the proposal **23**
6. Human, material and equipment resources **25**
7. Expected impact of the results **27**
7.1. Expected impact 27
7.2. Social and economic impact of the expected results 28
References **31**

SUMMARY

Summary of the project

The concept of "agency" currently emerges as a central theme across a number of scientific and technical disciplines, as well as in the philosophy of science, and poses new sociotechnical challenges. The shift definitely marks a move from reactive to active paradigms in science, technology and philosophy. In biology, there's a growing emphasis on the agent perspectives, from the study of protocells and prebiotic evolvability to multicellular systems and relational views of reproduction. In evolutionary theory, the concept of agency is also increasingly central, viewing biological entities as active subjects of their own evolution and development, active constructors of their environments, including ultimately the planetary scale. For cognitive sciences, agency is pivotal in discussions on embodiment intentionality, and social interaction. Artificial Intelligence research is recently focusing on the creation of digital autonomous agents. Lastly, the impact of digital technologies on human agency and autonomy is a growing area of concern in social and political sciences.

The "OUTAGENCIES: Varieties of autonomous agency" project will follow a naturalised transdisciplinary strategy, adding complex systems' tools to understand the varieties of agency present in different scientific and technical disciplines (complementing more traditional metaphysical or philosophy of language approaches to agency). The departure **hypothesis** is that "*Autonomous agency, in its variety of manifestations, is a unitary phenomenon that springs from material living organisational principles, and evolves through different transitions that explain the observed variety*". Three corollaries would follow: 1. There can be no autonomous agency without deep material continuity (contra computationalist agency attributions). 2. Not all matter is autonomously agential (in dialogue with new materialism). 3. Humans do not constitute a privileged standard to ground agency (rationality needs naturalisation). The central problem of agency's unity involves exploring its relationship with materiality, the transitions between types of agency, the conflicts arising between various agency manifestations, and the sociotechnical impact of artificially created "agents".

The project comprises 4 work packages. "WP1: Living Matters" will analyse protocell organisation and evolution, multicellularity, reproductive agency, explore ecological contexts of agency, and its implications for des-extinction. "WP2: Humanimal Varieties" will bridge reflective and unreflective agency, conceptualise how the environment traverses organisms, and examine autonomy and agency in health-related contexts. "WP3: Artificial Agencies" will focus on clarifying the ontological status of AI, analysing how digital technologies affect human agency, and developing accountability frameworks for artificial agencies. Finally, "WP4: Acts of Integration" will put together the project's main findings, confronting the hypothesis of agency's unity in variety, facilitating dialogues between different philosophical

schools on agency, and developing a framework for understanding agency in the metaphysics of sexgender.

Academic impact is expected in the fields of philosophy (general), philosophy of biological and health sciences and social sciences as a result of 14 papers in indexed journals, 2 monographs and 2 workshops. Social impact is expected in gender and environmental issues and artificial agency related policymaking.

Keywords

Philosophy of Science, Philosophy of Biology, Philosophy of Cognitive Sciences, Philosophy of technology, New Materialism, Artificial Intelligence, Gaia

Summary of expected impact

Substantiating (or discarding) the main hypothesis of the project by clarifying (or proving wrong) the unity of the diverse manifestations of agency in our world would be a central contribution to contemporary philosophy (and philosophy of science in particular) with important ontological, metaphysical and epistemological implications. A positive answer would be an important contribution to the organisational and enactive approaches to life, mind, and society. It would also have a deep impact on current debates in philosophy: first, to enrich new materialist approaches to agency with autonomous organisational accounts; second, to set some limits to the computationalist program to digitalise autonomous agency; finally, to support the naturalising efforts that are already in place in rationalistic approaches to human agency. These general implications of the project can be spelled down to the philosophy of specific sciences and sociotechnical domains, along its main research lines.

Regarding the origins of life question, finding out whether autonomous agency is the result of a unified principle, is key to frame research pathways, articulated as protocellular transitions, in the next decades. Concerning ecological and planetary dimensions of agency, we shall contribute to better framing the interdisciplinary efforts to integrate different disciplines into the understanding of the Earth system, while posing the question of the role of human agency towards a sustainable future. Understanding the reproduction of living and artificial (biomimetic) systems requires introducing relational agency at the core of contemporary theoretical efforts, in order to overcome the merely formal (copying or replication) and individualistic models. Additionally, identifying and characterising the unitary nature of autonomous agency would also help to clarify and redirect causal democracy claims regarding the understanding of organisms.

In exploring the continuity of varieties of agency in cognitive sciences, we expect to contribute to a more nuanced understanding of how reflective and non-reflective capacities interact. Moreover, the development of a model of intentional agency will make possible to bridge the gap between sensorimotor and linguistic approaches in 4E philosophy of mind and provide an important contribution to current attempts to naturalise rationality. Rethinking this relationship is likely to have profound consequences for anthropology, comparative psychology, and psychiatry, as well as ethical questions concerning responsibility, inclusion and animal rights. New breakthroughs in the way in which varieties of agencies and varieties of environments intertwine will also have implications for evolutionary theory through organismic and agential analysis of evolutionary processes, including niche construction. Finally, this agential focus will contribute to explain interactive aspects of placebo effects in the field of psychotherapy, on the one hand, and menstrual agencies, on the other hand.

Finally, the project would enrich debates in philosophy of technology and provide a much-needed clarification in the discourse around AI and the place of human autonomy and agency in complex digital environments. The results should clarify the specific sense in which living materiality is often (yet uncritically) considered as a necessary condition for genuine forms of intentional or conscious agency in AI systems and how to conceive AI as a form of agency in a larger (human-)live-involving manner. Moreover, the project outcomes can substantially contribute to the philosophical (cognitive, ethical and sociotechnical) debate around the prospects of human autonomy in a digital world that is actively adapting to capture and steer human behaviour.

1. BACKGROUND AND STATE OF THE ART.

1.1. Motivation and Relevance.

Like the notions of particle, structure, system, or network in the past, the notion of “agency” is playing an increasingly relevant role in various scientific and technical disciplines and, thereby, it is also becoming a key concept in philosophy of science, and, more generally, in a number of challenges within the sociotechnical domain.

As the philosopher William Bechtel advances, **biology** and other sciences are moving from a "reactive" to an "endogenously active" scientific paradigm (Bechtel & Bollhagen, 2021). We can see signs of this shift in various aspects of science and philosophy, both in terms of evolved entities and artefacts. The main dimension of this project concerns the varieties of agency that we can identify and study within this framework, some of which would be at the basis of what we call autonomous agency. Historically, **living matter** was differentiated from inert matter by its autonomous activity, whereas non-living materials were considered to lack any inherent activity (Etxeberria, 2012; Grote, 2019; Keller, 2016). According to Evelyn F. Keller the field of "active matter" originates in physics yet extends naturally towards biology, conceiving and characterising living cells as material entities while, at the same time, addressing physical dynamics as an important element to understand their behaviour.

In **origins of life** research, although traditional approaches have mostly focused on the chemical (abiotic) synthesis of biomolecules (Patel et al., 2015), awareness is increasing about the importance of taking a dynamic ‘systems’ perspective that puts protocells (complex, organised, self-assembled and metabolising units) at centre stage (Ruiz-Mirazo et al., 2017, p. 17; Shapiro, 2007), which opens the field to consider how autonomous agents could have emerged from self-organising phenomena and evolved under prebiotic conditions (Ruiz-Mirazo et al. 2020).

In the related field of **synthetic biology**, engineering and technologically driven approaches (e.g. Li, 2023) bring about a number of difficulties and incompatibilities with the actual living processes (Kwok, 2010), reflected also as a strong tension between the autonomy and agency of the artificial systems –or the hybrid constructs– thus created and the natural ones (Ruiz-Mirazo & Moreno, 2013).

Multicellular systems face the challenge of controlling the behaviours of individual cells and groups of cells that are themselves agents to achieve integration through division of labour and minimize conflicts (Bich et al., 2019). Importantly, multicellular systems are themselves agents that recruit the activities of their component cells to interact as integrated wholes with their surroundings (Arnellos & Moreno, 2015). Agency in **development** is crucially manifested in how interaction with the environment triggers functional and structural transitions (Gilbert & Epel, 2015). Neo-Darwinian evo-

lutionary theory and molecular biology with a focus on genetics pushed a notion of **reproduction** which is formal (rather than materially grounded) and individualistic (emphasising self-reproduction). New perspectives on reproduction need to be developed to stress its material and developmental organisation, and its agential and relational features (Etxeberria, 2023; Nuño De La Rosa, 2023). Some of the work done on the viviparous mode of reproduction and pregnancy needs to be expanded in this sense (Fusco & Minelli, 2023; Nuño De La Rosa et al., 2021).

In **evolutionary theory,** agency is a category that is becoming increasingly central. An initial impetus to this idea came from Richard Lewontin, who wrote that evolving entities are not just objects of change but subjects (Lewontin, 1983). Since then, many authors have argued that agency in evolution is linked to the idea that the evolutionary units are agent organisms and not genes (Walsh, 2015; Tomasello, 2022; Ruiz-Mirazo et al., 2000; Okasha, 2023).

The **ecological** agential perspective on organisms as agents involves considering how their activities shape the niches in which they live and determine the selective pressures to which they and their offspring are exposed (Cortés-García & Etxeberria, 2023). This agential perspective implies that, rather than passive units shaped by natural selection, organisms construct their environment, which affects the course of their own evolution. Niche Construction Theory elaborates the relationship between organisms as agents and environments (Laland et al., 1999, 2016).

A further and more general issue involves dealing with the prospects and specificities of what is considered the **interactionist consensus** (Kronfeldner, 2009; Schaffner, 2016; Sterelny & Griffiths, 1999) in order to face what Keller (Keller, 2010) has dubbed The Mirage of a Space (between Nature and Nurture) still in place.

The centrality of climate change in contemporary politics has raised new questions about the attribution of agency to the **Earth System**. On one side, philosophers have revitalised the concept of Gaia (Latour, 2017; Stengers, 2017) as a way of exploring the role of agents external to the human in the realm of history, emphasising the active nature of our environment. On the other, debates about governance of the Earth System have placed an important emphasis on the definition of agency, agentiality and actors in the Earth System framework and the strive for a sustainable future (Dellas et al., 2011).

The relevance of agency does not stop in the biological sciences. One of its most prominent classical manifestations is in the cognitive and behavioural sciences. The concept of agency lies at the heart of many current debates in **cognitive science** and philosophy of mind regarding intentional action (Bratman, 2007; Ford et al., 2011) embodiment (Barandiaran, 2017; Chemero, 2009; Di Paolo et al., 2017), active inference (Friston et al., 2010), rationality (Viale et al., 2023), or social interaction and collective agency

(Gallagher, 2020; Goodwin, 2019). In **archaeology**, material agency challenges anthropocentric, internalist cognitivist or abstract symbolic views of cultural evolution, highlighting both the intricate materiality of human agency and the agentic capacity of material culture itself (Knappett & Malafouris, 2008).

In recent developments of **Artificial Intelligence**, the notion of autonomous agency is playing a central role as a reachable milestone toward Artificial General Intelligence and a powerful deployment of Large Language Model capabilities to complement and/or substitute human agency (Morris et al., 2023; Wang et al., 2023). Living and socioeconomic conceptions of materiality have been argued to play an important role in the very possibility of artificial systems with agential capacities (Pasquinelli, 2023; Seth, 2021; Smith, 2019).

An increased interest is recently observed, as well, in exploring the impact of hyper-designed **digital environments** on human autonomous agency (Clowes, 2019; Marin, 2022; Mathur et al., 2021; Pérez-Verdugo, 2022; Pérez-Verdugo & Barandiaran, 2023), and, particularly, in relation to Artificial Intelligence (Andrada et al., 2022; Prunkl, 2022; Vaassen, 2022). Whereas human autonomous agency is jeopardized by increasingly complex and adaptive environments, the power of so claimed artificial autonomous agency increases.

What do all these conceptions of agency have in common? How can progress in our understanding of unicellular forms of agency help us grasp the challenges of artificial digital agents? How can the study of sociotechnical agencies shed light on the evolutionary effect of organismic scaffolding? How should living and artefactual forms of agency be merged to rethink and transform ourselves as agents?

1.2. Working hypothesis and previous contributions of the research team

The main aspect underlying all these questions has to do with the different approaches to the **unity of the concept of agency** (or lack of it) as it is used in various scales and scientific domains (as just outlined above). This is a problem that was identified throughout our previous project ("Outonomy: fleshing out autonomy beyond the individual" to be completed by February 2024) regarding the integrative, interactive, extended and sustainable nature of autonomy (as thematized in different disciplines). We identify at least three possible positions to respond to this issue:

No-unity: There is a *polysemy* of autonomy and agency, and there is no underlying unifying principle or phenomenon to which the different scientific uses of the term "autonomous agency" refer.

Unity by continuity: There is a unique type of autonomous agency that complexifies through biological and sociotechnical evolution – varieties of autonomous agency being nothing but different elaborations of this unity.

Unity by organising principle: Varieties of autonomous agency are the manifestation of an underlying general principle (e.g. operational closure, free-energy dissipation, active materiality, information processing, etc.). There are levels or types of agency that manifest differently the same underlying logic in different domains: biological, neurocognitive, sociolinguistic, technological, etc.

Intertwined with this central problem of unity, we also found a number of recurrent threads: (i) the relationship of this hypothetical unity with **materiality**; (ii) the **transition** between different types of autonomous agencies (e.g. the emergence of neurocognitive autonomy within biological evolution, from individual to social, etc.), (iii) several agential **conflicts** that appear while studying different types or manifestations of agency (e.g. mother-foetus in pregnancy, cell-organism, organism-population in evolution, etc.), and (iv) the ontological status, epistemic role and deep agency-transforming capacities of **artificial** attempts to implement, model or enhance autonomous agency (particularly in the fields of Artificial Life and Artificial Intelligence).

Our departure point is a mixture of positions 2 and 3 expressed above:

HYPOTHESIS: *Autonomous agency, in its variety of manifestations, is a unitary phenomenon that springs from material living organisational principles and evolves through different transitions that explain the observed variety.*

The hypothesis, if true, would have important philosophical and scientific consequences in the form of **corollaries**:

1. *There can be no genuine autonomous agency without deep material continuity with pre-existing forms.* This would set some important limits on some forms of **computational functionalism**, with important implications for some forms of digital transhumanism and "strong AI" research programs (Bostrom, 2017; Kurzweil, 2005). In particular, it would make possible to spell out the often claimed requirement for AI systems to be living systems in order to display genuine intentional agency (Seth, 2021)
2. *Not all matter is autonomously agential.* **New materialist** researchers (Barad, 2007; Coole & Frost, 2010) and object-oriented ontological conceptions of agency (Harman, 2017; Latour, 2017) often attribute agential capacities to items as diverse as electrons, chairs or Virgin statues. Consequently, there may be different conceptions of agency that encompass different capacities, some of which do not fit well with the autonomous agency of living beings, as these approaches emphasise activity, connectedness and interdependence rather than autonomy.

3. *No human-essentialism.* Human agency (not necessarily devoid of specificity) does not constitute a privileged standard for full-blown agency. Standard approaches in **action theory** treat non-human kinds of agency as conceptually derivative from 'full-blown' agency, dependent on reason-based intentions, conceptual thought and socio-olinguistic practices (Schlosser, 2019; Ford et al., 2011; cf. Brandom, 1994). In contrast, our hypothesis predicts a conceptual continuity and interdependence between human agency and other varieties of agency exhibited by human and non-human organisms.

The project distinguishes itself through its unique and specific approach within the **philosophy of science**. Unlike traditional metaphysics or philosophy of language perspectives, our approach emphasises **naturalisation**, grounding philosophical inquiry in real-world phenomena. This is achieved through an **inter- and multidisciplinary** effort, bridging various disciplines and scales to offer a holistic view. Central to our methodology is a focus on **complex systems**, utilising tools like simulation models, dynamical systems' theory, information theory, and network analysis to understand intricate interactions and dynamics. Additionally, our work is rooted in the **organisational** and **enactive** tradition (Di Paolo et al., 2017; Maturana & Varela, 1980; Moreno & Mossio, 2015) that has characterised the work of the group for the last 3 decades.

The connection between **biological and sociotechnical** conceptions of agency has attracted recent attention in philosophy of science (Moss, 2024), and comes justified by a number of philosophical approaches that have extended biocybernetic frameworks to sociotechnical systems, departing from the legacy of Maturana and Varela and Haraway (Hayles, 2017; Hui, 2019; Stiegler, 2020) or as a result of projecting biologically embodied accounts of cognition to technological domains (Clark, 2010; Pérez-Verdugo & Barandiaran, 2023).

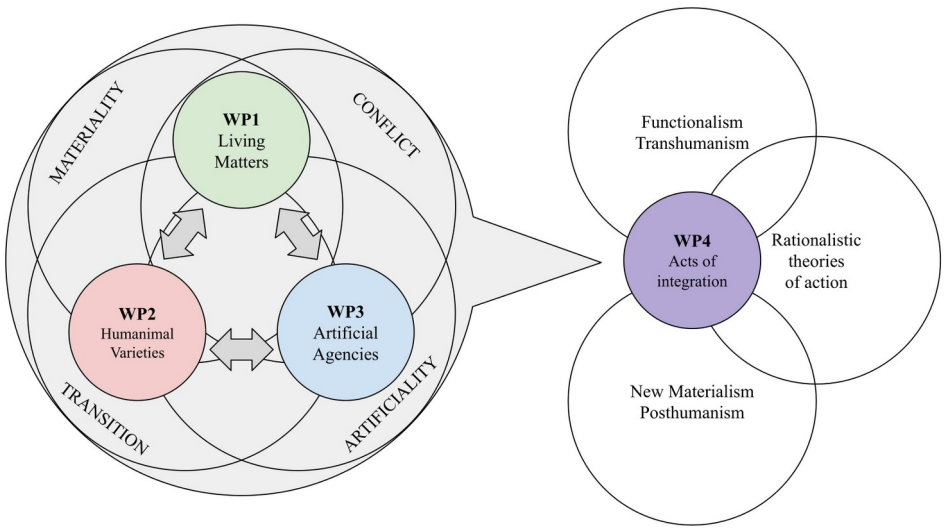

Figure 1: Diagram of the project's conceptual and organizational structure

2. OBJECTIVES

The general goal is to test the main hypothesis stated above. Accordingly, the project is structured in 4 work packages (WPs hereafter) encompassing specific tasks and their corresponding goals (see Figure 1 for an illustration). All tasks follow similar methodologies of conceptual analysis, synthesis, contrast and publication of a research output in the form of paper (or book chapter) as explained in section 3.2. The first three WPs cover different scales of autonomous agency: "**WP1**: Living Matters" will analyse protocell evolution and multicellularity, reframe reproductive agency, investigate ecological contexts of agency, and assess implications for des-extinction. "**WP2**: Humanimal Varieties" will explore the connection between reflective and unreflective agency, conceptualise environmental influences in biology and medicine, and examine autonomy and agency in health-related contexts. "**WP3**: Artificial Agencies" will focus on clarifying the ontological status of AI, analysing how digital technologies affect human agency, and developing accountability frameworks for artificial agencies. Finally, "**WP4**: Acts of Integration" will put together the project's main findings, confronting the hypothesis of agency's unity in variety, facilitating dialogues between different philosophical schools on agency, and developing a framework for understanding agency in the metaphysics of sexgender.

We here display the full list of work packages, tasks and goals. The members of the research team involved in each task are indicated with their initials (XB: Xabier Barandiaran, AE: Arantza Etxeberria, JU: Jon Umerez, KRM: Kepa Ruiz-Mirazo, JJB: Juan Bautista Bengoetxea, LB: Leonardo Bich), followed by the work-team members with their full name. In bold, the team leader for each task.

2.1. WP1: Living matters

T1.1. Biological origins [**KRM**, LB, Ben Shirt-Ediss, Arián Ferrero Fernández]

G1.1. To analyse critically different stages of protocell development and their prebiotic transition mechanisms in terms of their agential and autonomous capacities (e.g.: metabolic, behavioural, regulatory, informational, etc.).

T1.2. Collectivity and Evolvability [**KRM**, LB, Arián Ferrero Fernández]

G1.2. To study how microbial forms of individual and collective autonomous agency (including their ecological and phylogenetic-hereditary relationships) are essential for "open-ended evolvability".

T1.3. Reproductive agency [**AE**, David Cortés-García]

G1.3. To reframe reproductive agency as an organismic and relational capacity with goals that need to be realised in a distributed fashion among different agents and can only be fulfilled by establishing relationships with them.

T1.4. Ecologies and the earth system [**KRM**, JU, XB, Alex Merlo, AE]

G1.4. To explore how ecosystems and, ultimately, the planetary scale of material organisation (Gaia) has been and can be understood as the fundamental ecological context of all forms of agency, a mesh of agencies and, perhaps, an instance of agency itself.

T1.5. APPLICATION: Agency des-extinction [**JU**, Mikel Asteinza Arteche]

G1.5. To examine the claims of des-extinction techniques that disregard both specific organismic constraints and general environmental, niche, or ecological factors, granting agency only to genomes without any context; to assess the potential social impact of this view in reinforcing simple genetic engineering solutions over conservationist and sustainability policies.

2.2. WP2: Humanimal varieties

T2.1. Intentional agency, rationality and sociality [**XB**, Mirko Prokop, T. van Es]

G2.1 To provide a model of intentional agency that can bridge reflective (involving e.g. practical reasoning, inference and planning) and unreflective kinds of agency (involving e.g. sensorimotor skills and habits) and clarify its relationship with sociality.

T2.2. Environments through agents [**AE**, JU, LB, Alberto Monterde Fuertes, Laura Menatti]

G.2.2. To establish what kind of conceptualisation of the environment is required in current biological and medical views (including epigenetics, symbiosis or holobionts), as in some current ontologies of biology agents are not merely surrounded, but also *traversed* by environments in many ways.

T2.3. Agencies in health and care [**AE**, Izar Agirresarobe, Enara García Otero]

G. 2.3. To examine the complex interrelations between varieties of autonomy and agency in giving rise to psychic conflict and resilience in health-related phenomena (e.g. placebo effect, menstruation, etc.).

2.3. WP3: Artificial Agencies

T3.1. Artificial Intelligence [**XB**, JU, JBB, AE, Lola Almendros]

G. 3.1. To clarify the ontological status of Artificial Intelligence systems that are considered "autonomous systems" in computer science literature, with an eye on the contemporary debates on agency, autonomy and materialism.

T3.2. Cyborg autonomies [**XB**, JBB, Marta Pérez Verdugo, Maite Arraiza]

G3.2. To analyse the way in which digital-technological environments (particularly AI and platforms) affect and transform human autonomous agency at individual and collective scales.

T3.3. APPLICATION: Policy making for artificial agents [**JBB**, XB, Lola Almendros]

G3.3. To assess how material and organisational principles of autonomous agency in artificial environments can be used to develop an accountability framework for decision-making, regulation and public policy.

2.4. WP4: Acts of integration

T4.1. Unity in Variety [**XB**, KRM, JU, LB, Andrea Gambarotto]

G4.1. To confront the general hypothesis of the project regarding the unity of the phenomenological variety of agency, bringing together results of the project.

T4.2. Dialogues in Action [**XB**, AE, JU, JBB]

G4.2. To bring research results into a dialogue regarding the autonomy of agency with existing philosophical schools, particularly with the *interactionist consensus*, with new materialism (posthuman), rationalist theories of action, and functionalism (transhuman).

T4.3. APPLICATION: Metaphysics of sexgender agencies [**AE**, Miren Lizarzaburu, Maite Arraiza]

G4.3. To develop a conceptual framework to address agency in the metaphysics of sexgender, integrating material, humanimal and technical dimensions and its impact in society.

3. METHODOLOGY

The main philosophical methodology of the project involves naturalistic conceptual analysis and synthesis based on an active dialogue with empirical research, computational and mathematical models and theoretical concepts, as well as discussion and confrontation with alternative theories. In particular, this methodological approach involves the following steps:

1. Conceptual **analysis** as a result of
    a. Literature review for the philosophical and theoretical aspects of the problem involved: (i) Mapping of the main positions or distinctions in the theoretical debate, and (ii) Identification of specific problems or tension within or between approaches.
    b. Literature review for the relevant empirical and, in general, scientific case studies and contributions that can shed light on the theoretical problem(s) under investigation.
2. Conceptual **synthesis** of new or conflicting categories (autonomy, agency, action, norm, etc. in different domains) or philosophical and theoretical accounts that are needed to address the problem.
    a. This conceptual synthesis is sometimes expressed in terms of *minimal models* or minimal theoretical and philosophical accounts built upon basic case studies.
    b. The development of conceptual simulations or *computational models* (mostly using network science tools, agent-based modelling, numerical calculus for differential equation modelling and optimization methods, like genetic algorithms, for parametric fitting and multivariable regression methods for pattern identification). [Apply only to T1.1, T1.2, T2.1]
    c. Adjustment of the model to the empirical literature to evaluate the plausibility and relevance of the synthesis.
3. **Contrast and discussion** at different scales:
    a. With scientists to establish a constant feedback between philosophical and scientific research, and to pursue active collaborations on common issues (see stays planned in sect. 3.3 below).
    b. With other philosophers and theoreticians, searching both collaboration and constructive discussion between opposing approaches.
    c. In wider academic conferences and symposia, with reviewers, etc.
4. **Revision** of the conceptual synthesis or problem definition in the light of discussion or new empirical findings or challenges.
5. **Publication** of final results.

Specific methodologies will be applied to some tasks. Computer simulations of the ecological and evolutionary dynamics of self-reproducing protocellular systems will be used for T1.1, T1.2 and robotic agents in T2.1. Task 4.1 will include quantitative meta-analysis techniques of scientific literature (in combination with the use of qualitative methods) in search of the different uses of the term "agency" and "autonomous agency" across the fields of Biology, Cognitive Science and Artificial Intelligence. T3.1 and T3.2 involve systematic studies and user experience with Large Language Models and Generative AI.

Interdisciplinary collaborative methodologies will be used for team meetings, seminars and workshops, including design-thinking methodologies and collaborative writing techniques. These methodologies will also include participation and discussion protocols aimed at neutralising the gender and minority domination structures that are usually imposed in scientific and academic meetings.

4. WORK PLAN

The research methodology will not be carried out individually but at different levels of cooperation between research-team and work-team members of the project. In general, each task/goal will be pursued by at least one interdisciplinary research-crew made of, at least, one senior researcher, a junior or predoctoral philosophy researcher, supported when needed by a scientist or modeller. These crews constitute the nodes of a collaboration network. Crews will host **regular meetings** to share knowledge, references, analysis and to make progress on the synthetic proposal. Every two months a session of the **IAS-Research seminars** (a regular series of research seminars that have been taking place for the last 8 years) will be devoted to this project (overall 3 seminars per research line during the duration of the project). In general, seminars will involve both external-open sessions and internal-closed sessions. Internal sessions will be used to contrast and call for feedback on the analytic or synthetic achievements or work in progress on the goals, whereas external sessions will involve inviting external collaborators to expand on their areas of expertise in order to provide valuable input for the goals. They will also serve to attract wider audiences and to carry out dissemination activity.

Every two months, there will be a **coordination meeting** between the PI and the postdoc (when hired) in order to review the state of the project. Every six months, there will be a meeting of the research-team to keep track of the progress regarding each goal. Meetings with the whole network of collaborators (the work-team) will take place 5 times during the project duration: 1. Kickoff meeting, 2. Internal project workshop, 3. Public external workshop, 4. Progress report meeting and 5. Evaluation meeting. The **internal workshop** (2025Q3, see Figure 2) will make the outcomes of the first part of the project available to be discussed and reviewed by the rest of the project members. During this gathering, **drafts of research outcomes** (goal specification, progress report, working hypothesis, achieved and expected publications) will be discussed and evaluated collectively between research and work team members.

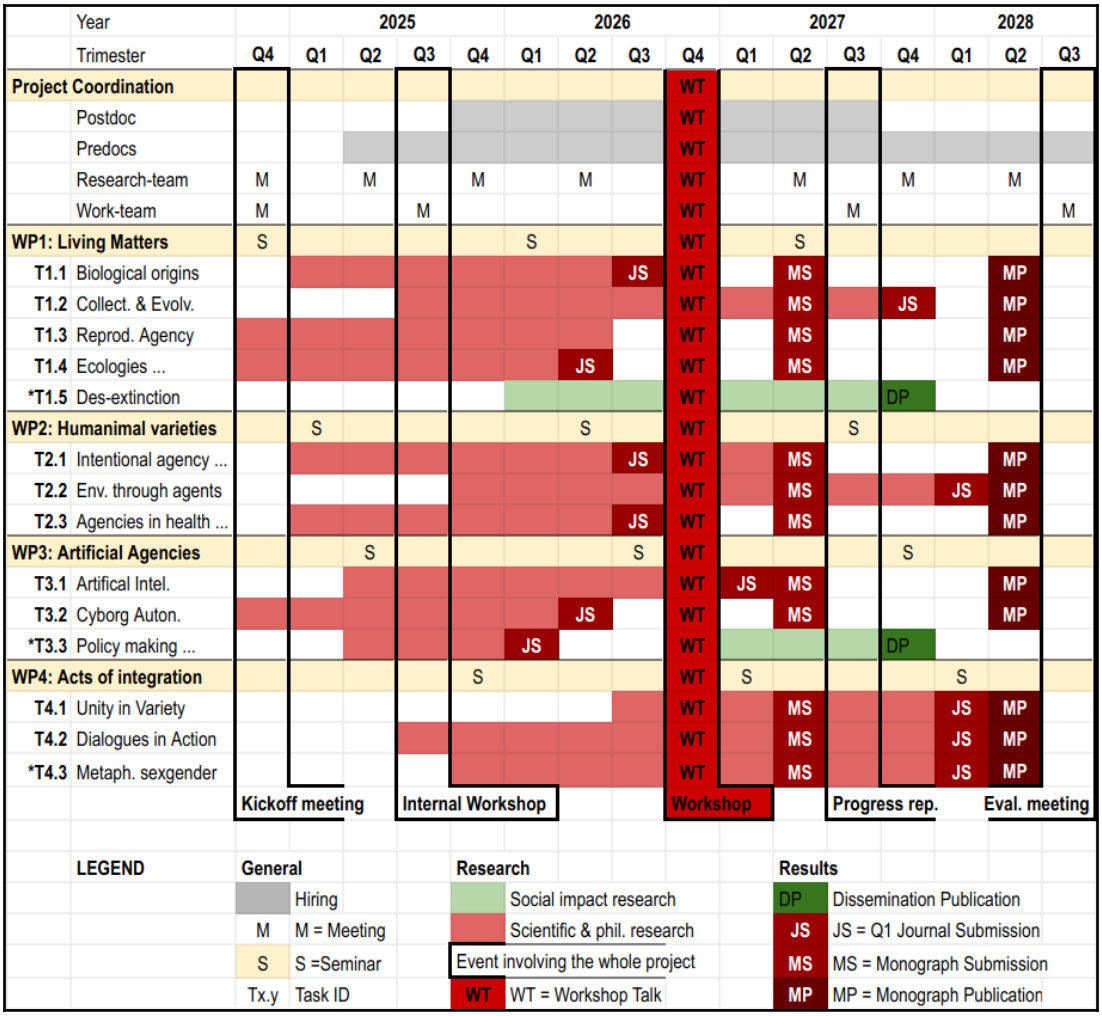

Figure 2: Gantt diagram depicting the temporal unfolding of the work plan. Important project events and activities and expected submission deadlines are shown (see legend). Each task follows the methodological steps described above. Research activities and submission data have been allocated to balance research-crew workload and theoretical interdependencies.

A **public workshop** will take place in November 2026 (in red in Figure 2). For this event, the research results are expected to be mature and open to revision within a wider community of contributors and experts. This workshop will be the most important mid-project milestone, bringing together key invited speakers for each research line and presentation of results for each subgoal (by members of the project). Speakers of the workshop will be expected to contribute an extended abstract to be peer-reviewed. A selection of the contributors to the workshop will be invited to submit a paper to the monograph. **Monograph** contributions are expected to be submitted 3–6 months after the workshop and to include the feedback from the workshop discussions. Peer reviewed revised versions of the monograph are expected to be published 4 months before the end of the project. Publication of the monograph will be followed by a communication campaign to ensure proper dissemination of the project results.

Parallel to the research process culminating in the (public) Workshop and the Monograph, research crews for each goal are expected to deliver a **research article in a specialised Q1 journal** and to present their work in at least one **international conference**. The overall approach is that this parallel research outcome is to be published and contrasted among a more specialised audience (e.g. Biology & Philosophy, or Minds and Machines) whereas the contribution to the monograph will be meant to achieve a wider level of generality and philosophical abstraction. In this way, interdisciplinary naturalism can contribute both to general philosophical problems and feed back to specific sciences and disciplines.

5. PREVIOUS RESULTS OF THE TEAM IN THE THEME OF THE PROPOSAL.

Xabier E. Barandiaran (PI) has made significant contributions to the conceptualisation of autonomous agency as a unified phenomenon (Barandiaran et al., 2009), specifically concerning biological or minimal cellular emergence of normativity in agency (Barandiaran & Egbert, 2013), most notably in sensorimotor agency (Barandiaran, 2017; Di Paolo et al., 2017) and, more recently, personal and political autonomous agency in digital environments (Barandiaran et al., 2024; Pérez-Verdugo & Barandiaran, 2023).

Arantza Etxeberria has worked on the organismic perspective and organicism, autonomous organisation and agency across living beings and technical domains (Etxeberria & Moreno, 2014; Etxeberria & Ruiz-Mirazo, 2009; Moreno & Etxeberria, 2005). Two aspects in particular need to be highlighted in her latest work on agency: one is the focus on issues related to reproduction, including pregnancy and menstruation, from an evolutionary perspective (Etxeberria, 2023; Nuño De La Rosa et al., 2021); the other concerns the contrast between conflict-based approaches and interdependencies, highlighting the relevance of relationality to address agency (Etxeberria et al., 2023).

Kepa Ruiz-Mirazo has devoted his academic career to understanding the emergence of biological complexity through the process of abiogenesis. In that context, he has contributed during the last 25 years to developing the protocell camp, putting forward both in vitro and in silico computational models of simplified but non-reducible, prebiotic versions of current prokaryotic cells. The organisational and evolutionary roots of autonomous agent behaviour in this naturalised scenario has been one of his central research interests (Ruiz-Mirazo et al., 2017; Ruiz-Mirazo & Mavelli, 2008; Ruiz-Mirazo & Moreno, 1998, 2004).

Jon Umerez, has made several contributions to the definition of material conditions for grounding living or cognitive phenomena (Moreno et al., 1994, 1997; Umerez, 1994, 1998) as well as to the analysis of the concept of autonomous systems (Etxeberria et al., 2000; Moreno et al., 2008). Additionally, he has contributed to the development of the concept of constraint to account for the autonomy of living systems and, especially, to characterise control hierarchical interrelations of levels (Umerez, 2001, 2016, 2021) in view of critically analysing the so-called *interactionist consensus* (Ferreira Ruiz & Umerez, 2021).

Leonardo Bich has been focusing his research on the organisation of living systems. He has contributed to the development of the foundations of the autonomy framework (Bich et al., 2016), and of an organisational approach to the study of minimal cognition in living and artificial systems (Bechtel & Bich, 2021; Bich, 2020). He has also focused on fleshing out the notion of environment from an organisational perspective (Menatti et al., 2022) and the role of materiality in multicellular systems (Bich, 2024; Bich et al., 2019).

Juan Bautista Bengoechea has contributed to a philosophical analysis of scientific results and evidence-based tasks aimed at supporting decision-making and regulatory actions in the field of foods (health claims) (Bengoetxea & Todt, 2021; Luján et al., 2016). He has also developed a framework for the study of educational technological practices in the face of the rise of AI and in the projection of cyborg agencies (Bengoetxea, 2021).

6. HUMAN, MATERIAL AND EQUIPMENT RESOURCES

The research group can count on the infrastructure provided by the UPV/EHU in the Faculty of Educación, Filosofía y Antropología on the Campus of Gipuzkoa. All members of the group affiliated with the UPV/EHU have assigned (individual or shared) offices in the Faculty or in the **research facilities** of the Centro Carlos Santamaria on the same campus, where workspaces for visiting researchers are available and where the group library and seminar room are located. The latter is the space where meetings, seminars and complementary educational activities are carried out.

The group has its own **specialised bibliographical resources** in addition to those provided by the university library. For this reason, the project requests a budget to keep these resources up to date. Regarding the technological equipment, the members of the group are equipped with **computers** connected to the UPV/EHU network. The budget includes funding for three new computers to replace the obsolete ones and to equip the new doctoral students. Due to the memory and processing computational demands of computer simulations and network analysis techniques, two powerful laptops will be acquired. The group also has **Virtual/Mixed Reality** Headsets (three 128GB Meta Quest 3 + one Link cable) which allows for an exploration of human agency within this novel kind of digital environment by various researchers simultaneously, either through existing apps or potentially through specific environments developed to carry out experiments.

Regarding the collaborative publishing of open data and code, if needed, the project will use the **public repositories** of the IAS-Research group on Github: https://github.com/IAS-Research. One of the most important digital infrastructures for this project will include the **bibliography and knowledge management software** Zotero, enabling effective management and sharing of bibliographical items, citations, documents and notes across users, devices and platforms. The program is free and open source and is widely used in the group to share references and bibliography.

7. EXPECTED IMPACT OF THE RESULTS.

7.1. Expected impact

Substantiating (or discarding) the main hypothesis of the project by clarifying (or proving wrong) the unity of the diverse manifestations of agency in our world would be a central contribution to contemporary philosophy (and philosophy of science in particular) with important ontological, metaphysical and epistemological implications. A positive answer would be an important contribution to the **organisational and enactive approaches** to life, mind, and society. It would also have a deep impact on current debates in philosophy: first, to enrich **new materialist approaches** to agency with autonomous organisational accounts; second, to set some limits to the **computationalist** program to digitalise autonomous agency; finally, to support the naturalising efforts that are already in place in rationalistic approaches to **human agency**. These general implications of the project can be spelled down to the philosophy of specific sciences and sociotechnical domains, along its main research lines.

Regarding the **origins of life** question, finding out whether autonomous agency is the result of a unified principle, is key to frame research pathways, articulated as protocellular transitions, in the next decades. Concerning **ecological** and planetary dimensions of agency, we shall contribute to better framing the interdisciplinary efforts to integrate different disciplines into the understanding of the Earth system, while posing the question of the role of human agency towards a sustainable future. Understanding the **reproduction** of living and artificial (biomimetic) systems requires introducing relational agency at the core of contemporary theoretical efforts, in order to overcome the merely formal (copying or replication) and individualistic models. Additionally, identifying and characterising the unitary nature of autonomous agency would also help to clarify and redirect **causal democracy** claims regarding the understanding of organisms.

In exploring the continuity of varieties of agency in **cognitive sciences,** we expect to contribute to a more nuanced understanding of how reflective and non-reflective capacities interact. Moreover, the development of a model of intentional agency will make possible to bridge the **gap between sensorimotor and linguistic approaches in 4E philosophy** of mind and provide an important contribution to current attempts to naturalise rationality. Rethinking this relationship is likely to have profound consequences for anthropology, comparative psychology, and psychiatry, as well as ethical questions concerning responsibility, inclusion and animal rights. New breakthroughs in the way in which varieties of agencies and varieties of environments intertwine will also have implications for **evolutionary theory** through organismic and agential analysis of evolutionary processes, including niche construction. Finally, this agential focus will contribute to explain interactive aspects of **placebo** effects in the field of psychotherapy, on the one hand, and **menstrual agencies**, on the other hand.

Finally, the project would enrich debates in **philosophy of technology** and provide a much-needed clarification in the discourse around AI and the place of human autonomy and agency in complex digital environments. The results should clarify the specific sense in which living materiality is often (yet uncritically) considered as a necessary condition for genuine forms of intentional or conscious agency in **AI systems** and how to conceive AI as a form of agency in a larger (human-)live-involving manner. Moreover, the project outcomes can substantially contribute to the philosophical (cognitive, ethical and sociotechnical) debate around the prospects of **human autonomy** in a digital world that is actively adapting to capture and steer human behaviour.

The **bibliometric impact** of the project includes 14 papers in specialised indexed journals (10 Q1), two monographs, and at least 16 communications in national and international conferences.

7.2. Social and economic impact of the expected results.

Social Impact of Understanding Agency in Digital Environments: A deeper understanding of agency within digital environments could have far-reaching social impact. This could influence how we interact with and govern digital entities, including AI and virtual agents, leading to more ethical and socially responsible digital ecosystems. It might also shape our approach to digital rights and privacy, fostering a more conscious and informed digital society.

Influence on Public Policy and Societal Well-being: By exploring the concept of agency in various contexts, including biological and artificial systems, the project could inform public policies related to technology regulation, environmental conservation, and healthcare. The findings could contribute to the development of policies that better align with human and environmental well-being, promoting a more equitable and sustainable society.

Economic Implications through Technological and Biomedical Applications: The research has potential economic impact through its applications in fields like synthetic biology and AI. Insights into autonomous agency could lead to innovations in these areas, potentially resulting in new industrial opportunities or enhancing existing ones.

Enhancing Social Inclusion and Addressing Gender and Disability Issues: The project's exploration of agency in diverse systems, including its "humanimal" focus, could lead to a better understanding of social inclusion, particularly in terms of gender,disability and cognitive diversity. This could influence social attitudes and lead to more inclusive practices in various sectors, enhancing societal cohesion and equality.

Contribution to Sustainable Development and Environmental Protection: The project's investigation into ecological systems and the concept of agency within these systems might provide valuable insights for environmental protection and sustainable development. This could help in formulating strategies that are more aligned with ecological balance, contributing to the long-term sustainability of our planet.

REFERENCES

Andrada, G., Clowes, R. W., & Smart, P. R. (2022). Varieties of transparency: Exploring agency within AI systems. *AI & SOCIETY*. https://doi.org/10.1007/s00146-021-01326-6

Arnellos, A., & Moreno, A. (2015). Multicellular agency: An organizational view. *Biology & Philosophy*, *30*(3), 333–357. https://doi.org/10.1007/s10539-015-9484-0

Barad, K. (2007). *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Duke University Press.

Barandiaran, X. E. (2017). Autonomy and Enactivism: Towards a Theory of Sensorimotor Autonomous Agency. *Topoi*, *36*(3), 409–430. https://doi.org/10.1007/s11245-016-9365-4

Barandiaran, X. E., Calleja-López, A., Monterde, A., & Romero, C. (2024). *Decidim, a Technopolitical Network for Participatory Democracy. Philosophy, Practice and Autonomy of a Collective Platform in the age of Digital Intelligence.* Springer. https://link.springer.com/book/9783031507830

Barandiaran, X. E., Di Paolo, E., & Rohde, M. (2009). Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-temporality in Action. *Adaptive Behavior*, *17*(5), 367–386. https://doi.org/10.1177/1059712309343819

Barandiaran, X. E., & Egbert, M. D. (2013). Norm-Establishing and Norm-Following in Autonomous Agency. *Artificial Life*, *91*(2), 1–24. https://doi.org/10.1162/ARTL_a_00094

Bechtel, W., & Bich, L. (2021). Grounding cognition: Heterarchical control mechanisms in biology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1820), 20190751. https://doi.org/10.1098/rstb.2019.0751

Bechtel, W., & Bollhagen, A. (2021). Active biological mechanisms: Transforming energy into motion in molecular motors. *Synthese*, *199*(5–6), 12705–12729. https://doi.org/10.1007/s11229-021-03350-x

Bengoetxea, J. B. (2021). Causalidad y regulación - Causality and Regulation: Algunas dificultades de los ensayos controlados aleatorizados. *Revista Portuguesa de Filosofia*, *77*(4), 1409–1438.

Bengoetxea, J. B., & Todt, O. (2021). Decision-making in the nutrition sciences: A critical analysis of scientific evidence for assessing health claims. *Manuscrito*. https://doi.org/10.1590/0100-6045.2021.v44n3.jb

Bich, L. (2020). Interactive Models in Synthetic Biology: Exploring Biological and Cognitive Inter-Identities. *Frontiers in Psychology*, *11*. https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00682

Bich, L. (2024). Integrating Multicellular Systems: Physiological Control and Degrees of Biological Individuality. *Acta Biotheoretica*, *72*(1), 1. https://doi.org/10.1007/s10441-023-09476-4

Bich, L., Mossio, M., Ruiz-Mirazo, K., & Moreno, A. (2016). Biological regulation: Controlling the system from within. *Biology & Philosophy*, *31*(2), 237–265. https://doi.org/10.1007/s10539-015-9497-8

Bich, L., Pradeu, T., & Moreau, J.-F. (2019). Understanding Multicellularity: The Functional Organization of the Intercellular Space. *Frontiers in Physiology*, *10*. https://doi.org/10.3389/fphys.2019.01170

Bostrom, N. (2017). *Superintelligence*. Dunod.

Brandom, R. (1994). *Making it explicit: Reasoning, representing, and discursive commitment*. Harvard University Press.

Bratman, M. (2007). *Structures of agency: Essays*. Oxford University Press.

Chemero, A. (2009). *Radical embodied cognitive science*. MIT Press.

Clark, A. (2010). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension* (B14). Oxford University Press; B14.

Clowes, R. W. (2019). Immaterial engagement: Human agency and the cognitive ecology of the internet. *Phenomenology and the Cognitive Sciences*, *18*(1), 259–279. https://doi.org/10.1007/s11097-018-9560-4

Coole, D., & Frost, S. (Eds.). (2010). *New Materialisms: Ontology, Agency, and Politics*. Duke University Press Books.

Cortés-García, D., & Etxeberria, A. (2023). Ontologies in Evolutionary Biology: The Role of the Organism in the Two Syntheses. In J. M. Viejo & M. Sanjuán (Eds.), *Life and Mind*

(Vol. 8, pp. 185–205). Springer International Publishing. https://doi.org/10.1007/978-3-031-30304-3_9

Dellas, E., Pattberg, P., & Betsill, M. (2011). Agency in earth system governance: Refining a research agenda. *International Environmental Agreements: Politics, Law and Economics*, *11*(1), 85–98. https://doi.org/10.1007/s10784-011-9147-9

Di Paolo, E. A., Buhrmann, T., & Barandiaran, X. E. (2017). *Sensorimotor life: An enactive proposal* (First edition). Oxford University Press.

Etxeberria, A. (2012). Active Matter, Organisms and Their Others. *Ludus Vitalis*, *20*(37), 253–262.

Etxeberria, A. (2023). Jacob's Understanding of Reproduction: Challenges from an Organismic Collaborative Framework. *HOPOS: The Journal of the International Society for the History of Philosophy of Science*, *13*(2), 535–553. https://doi.org/10.1086/726256

Etxeberria, A., Cortés García, D., & Torres Aldave, M. (2023). Organisms, Life Relations, and Evolution: Inter-Dependencies after Kropotkin's Mutual Aid. *ArtefaCToS. Revista de Estudios Sobre La Ciencia y La Tecnología*, *12*(1), 179–204. https://doi.org/10.14201/art2023121179204

Etxeberria, A., & Moreno, A. (2014). La noción de autonomía en biología: Aportaciones, retos y discusiones. In A. Casado da Rocha (Ed.), *Autonomía con otros. Ensayos de Bioética* (pp. 95–110). Madrid: Plaza y Valdés.

Etxeberria, A., Moreno, A., & Umerez, J. (2000). Understanding Autonomous Systems. *Communication & Cognition - Artificial Intelligence*, *17*(3–4).

Etxeberria, A., & Ruiz-Mirazo, K. (2009). The challenging biology of transients. *EMBO Reports*, *10*, S33–S36.

Ferreira Ruiz, M., & Umerez, J. (2021). Interactionism, Post-interactionism, and Causal Complexity: Lessons From the Philosophy of Causation. *Frontiers in Psychology*, *12*. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.590533

Ford, A., Hornsby, J., & Stoutland, F. (Eds.). (2011). *Essays on Anscombe's Intention* (First Harvard University Press paperback edition). Harvard University Press.

Friston, K., Daunizeau, J., Kilner, J., & Kiebel, S. (2010). Action and behavior: A free-energy formulation. *Biological Cybernetics*, *102*(3), 227–260. https://doi.org/10.1007/s00422-010-0364-z

Fusco, G., & Minelli, A. (2023). *Understanding reproduction*. Cambridge University Press.

Gallagher, S. (2020). *Action and Interaction* (New product). Oxford University Press.

Gilbert, S. F., & Epel, D. (2015). *Ecological developmental biology: The environmental regulation of development, health, and evolution* (Second edition). Sinauer Associates, Inc. Publishers.

Goodwin, C. (2019). *Co-operative action*. Cambridge University Press.

Grote, M. (2019). *Membranes to Molecular Machines: Active Matter and the Remaking of Life*. University of Chicago Press.

Harman, G. (2017). *Object Oriented Ontology: A New Theory of Everything* (1st ed.). Penguin Random House. http://gen.lib.rus.ec/book/index.php?md5=fea84861b16ed65e52207269b0ad172c

Hayles, N. K. (2017). *Unthought: The Power of the Cognitive Nonconscious*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226447919.001.0001

Hui, Y. (2019). *Recursivity and Contingency*. Rowman & Littlefield Publishers.

Keller, E. F. (2010). *The mirage of a space between nature and nurture*. Duke University Press.

Keller, E. F. (2016). Active matter, then and now. *History and Philosophy of the Life Sciences*, *38*(3), 11. https://doi.org/10.1007/s40656-016-0112-3

Knappett, C., & Malafouris, L. (Eds.). (2008). *Material Agency*. Springer US. https://doi.org/10.1007/978-0-387-74711-8

Kronfeldner, M. E. (2009). Genetic Determinism and the Innate-Acquired Distinction in Medicine. *Medicine Studies*, *1*(2), 167–181. https://doi.org/10.1007/s12376-009-0014-8

Kurzweil, R. (2005). *The Singularity Is Near: When Humans Transcend Biology*. Penguin.

Kwok, R. (2010). Five hard truths for synthetic biology. *Nature*, *463*(7279), 288–290. https://doi.org/10.1038/463288a

Laland, K., Matthews, B., & Feldman, M. W. (2016). An introduction to niche construction theory. *Evolutionary Ecology*, *30*(2), 191–202. https://doi.org/10.1007/s10682-016-9821-z

Laland, K., Odling-Smee, F. J., & Feldman, M. W. (1999). Evolutionary consequences of niche construction and their implications for ecology. *Proceedings of the National Academy of Sciences*, *96*(18), 10242–10247. https://doi.org/10.1073/pnas.96.18.10242

Latour, B. (2017). *Facing Gaia: Eight Lectures on the New Climatic Regime*. John Wiley & Sons.

Lewontin, R. C. (1983). The Organism as the Subject and Object of Evolution. *Scientia*, *77*(18), 65.

Li, L. (2023). Next-generation synthetic biology approaches for the accelerated discovery of microbial natural products. *Engineering Microbiology*, *3*(1), 100060. https://doi.org/10.1016/j.engmic.2022.100060

Luján, J. L., Todt, O., & Bengoetxea, J. B. (2016). Mechanistic Information as Evidence in Decision-Oriented Science. *Journal for General Philosophy of Science*, *47*(2), 293–306. https://doi.org/10.1007/s10838-015-9306-8

Marin, L. (2022). How to Do Things with Information Online. A Conceptual Framework for Evaluating Social Networking Platforms as Epistemic Environments. *Philosophy & Technology*, *35*(3), 77. https://doi.org/10.1007/s13347-022-00569-5

Mathur, A., Kshirsagar, M., & Mayer, J. (2021). What Makes a Dark Pattern... Dark?: Design Attributes, Normative Considerations, and Measurement Methods. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–18. https://doi.org/10.1145/3411764.3445610

Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and cognition*. D. Reidel Publishing Company.

Menatti, L., Bich, L., & Saborido, C. (2022). Health and environment from adaptation to adaptivity: A situated relational account. *History and Philosophy of the Life Sciences*, *44*(3), 38. https://doi.org/10.1007/s40656-022-00515-w

Moreno, A., & Etxeberria, A. (2005). Agency in Natural and Artificial Systems. *Artificial Life*, *11*(1–2), 161–175. https://doi.org/10.1162/1064546053278919

Moreno, A., Etxeberria, A., & Umerez, J. (1994). Universality without matter. *Artificial Life IV*, 406–410.

Moreno, A., Etxeberria, A., & Umerez, J. (2008). The autonomy of biological individuals and artificial models. *BioSystems*, *91*(2), 309–319.

Moreno, A., & Mossio, M. (2015). *Biological Autonomy: A Philosophical and Theoretical Enquiry*. Springer.

Moreno, A., Umerez, J., & Ibañez, J. (1997). Cognition and life: The autonomy of cognition. *Brain and Cognition*, *34*(1), 107–129.

Morris, M. R., Sohl-dickstein, J., Fiedel, N., Warkentin, T., Dafoe, A., Faust, A., Farabet, C., & Legg, S. (2023). *Levels of AGI: Operationalizing Progress on the Path to AGI* (arXiv:2311.02462). arXiv. https://doi.org/10.48550/arXiv.2311.02462

Moss, L. (2024). Concepts of Agency: Introduction to the Thematic Section. *Biological Theory*. https://doi.org/10.1007/s13752-023-00454-0

Nuño De La Rosa, L. (2023). Agency in Reproduction. *Evolution & Development*, *25*(6), 418–429. https://doi.org/10.1111/ede.12440

Nuño De La Rosa, L., Pavličev, M., & Etxeberria, A. (2021). Pregnant Females as Historical Individuals: An Insight From the Philosophy of Evo-Devo. *Frontiers in Psychology*, *11*, 572106. https://doi.org/10.3389/fpsyg.2020.572106

Okasha, S. (2023). The Concept of Agent in Biology: Motivations and Meanings. *Biological Theory*. https://doi.org/10.1007/s13752-023-00439-z

Pasquinelli, M. (2023). *The eye of the master: A social history of artificial intelligence*. Verso.

Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D., & Sutherland, J. D. (2015). Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nature Chemistry*, *7*(4), 301–307. https://doi.org/10.1038/nchem.2202

Pérez-Verdugo, M. (2022). Situating Transparency: An Extended Cognition Approach. *Teorema: Revista Internacional de Filosofía*, *41*(3), 7–24.

Pérez-Verdugo, M., & Barandiaran, X. (2023). Personal autonomy and (digital) technology: An enactive sensorimotor framework. *Philosophy & Technology*, *36*(4). https://doi.org/10.1007/s13347-023-00683-y

Prunkl, C. (2022). Human autonomy in the age of artificial intelligence. *Nature Machine Intelligence*, *4*(2), Article 2. https://doi.org/10.1038/s42256-022-00449-9

Ruiz-Mirazo, K., Briones, C., & De La Escosura, A. (2017). Chemical roots of biological evolution: The origins of life as a process of development of autonomous functional systems. *Open Biology*, *7*(4), 170050. https://doi.org/10.1098/rsob.170050

Ruiz-Mirazo, K., Etxeberria, A., Moreno, A., & Ibáñez, J. (2000). Organisms and their place in biology. *Theory in Biosciences*, *119*(3), 209–233.

Ruiz-Mirazo, K., & 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., & Moreno, A. (1998). Autonomy and emergence: How systems become agents through the generation of functional constraints. *Acta Polytechnica Scandinavica, Ma91*, 273–282.

Ruiz-Mirazo, K., & Moreno, A. (2004). Basic Autonomy as a Fundamental Step in the Synthesis of Life. *Artificial Life*, *10*(3), 235–259. https://doi.org/10.1162/1064546041255584

Ruiz-Mirazo, K., & Moreno, A. (2013). Synthetic Biology: Challenging life in order to grasp, use or extend it. *Biological Theory*, *8*, 376–382. https://doi.org/10.1007/s13752-013-0129-8

Schaffner, K. F. (2016). *Behaving: What's genetic, what's not, and why should we care?* (First edition). Oxford University Press.

Schlosser, M. (2019). Agency. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019 Ed.). https://plato.stanford.edu/archives/win2019/entries/agency/

Schwitzgebel, E., & Jennings, C. D. (2017). Women in philosophy: Quantitative analyses of specialization, prevalence, visibility, and generational change. *Public Affairs Quarterly*, *31*(2), 83–106.

Seth, A. (2021). *Being You: A New Science of Consciousness*. Faber & Faber.

Shapiro, R. (2007). A Simpler Origin for Life. *Scientific American*, *296*(6), 46–53. https://doi.org/10.1038/scientificamerican0607-46

Smith, B. C. (2019). *The promise of artificial intelligence: Reckoning and judgment*. The MIT Press.

Stengers, I. (2017). Autonomy and the Intrusion of Gaia. *South Atlantic Quarterly*, *116*(2), 381–400. https://doi.org/10.1215/00382876-3829467

Sterelny, K., & Griffiths, P. (1999). *Sex and death: An introduction to philosophy of biology*. University of Chicago Press.

Stiegler, B. (2020). Elements for a General Organology. *Derrida Today*, *13*(1), 72–94. https://doi.org/10.3366/drt.2020.0220

Tomasello, M. (2022). *The evolution of agency: Behavioral organization from lizards to humans*. The MIT Press.

Umerez, J. (1994). *Jerarquías Autónomas.-Un estudio sobre el origen y la naturaleza de los procesos de control y de formación de niveles en sistemas naturales complejos*. Ph. D. Dissertation, University of the Basque Country.

Umerez, J. (1998). The evolution of the symbolic domain in living systems and artificial life. In *Evolutionary Systems* (pp. 377–396). Springer.

Umerez, J. (2001). Howard Pattee's theoretical biology—A radical epistemological stance to approach life, evolution and complexity. *Biosystems*, *60*(1), 159–177.

Umerez, J. (2016). Biological Organization From a Hierarchical Perspective. Articulation of Concepts and Inter-level Relation. In N. Eldredge, T. Pievani, E. Serrelli, & I. Temkin (Eds.), *Evolutionary Theory. A Hierarchical Perspective* (pp. 63–85). The University of Chicago Press.

Umerez, J. (2021). Control Hierarchies: Pattee's Approach to Function and Control as Time-Dependent Constraints. In D. S. Brooks, J. DiFrisco, & W. C. Wimsatt (Eds.), *Levels of Organization in the Biological Sciences* (pp. 279–296). The MIT Press. https://doi.org/10.7551/mitpress/12389.003.0018

Vaassen, B. (2022). AI, Opacity, and Personal Autonomy. *Philosophy & Technology*, *35*(4), 88. https://doi.org/10.1007/s13347-022-00577-5

Viale, R., Gallagher, S., & Gallese, V. (2023). Editorial: Embodied bounded rationality. *Frontiers in Psychology*, *14*, 1235087. https://doi.org/10.3389/fpsyg.2023.1235087

Walsh, D. M. (2015). *Organisms, agency, and evolution*. Cambridge University Press.

Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., & Wen, J.-R. (2023). *A Survey on Large Language Model based Autonomous Agents*. arXiv. https://arxiv.org/abs/2308.11432v2