Collaborative Project

__________________________

eSMCs

Extending Sensorimotor Contingencies to Cognition

FP7-ICT-2009-6 Grant Agreement No. 270212

Deliverable D1.6: Agency and eSMCs

Milestone: MS4
Workpackage number: 1
Lead beneficiary number: 6 (UPV)
Estimated person months: 11
Delivery date: 46
Actual submission date: 30.03.2015
Dissemination level: PU

Contributors: Thomas Buhrmann
Ezequiel Di Paolo
Xabier Barandiaran
Hanne De Jaegher

PROPRIETARY RIGHTS STATEMENT
This document contains proprietary information of the eSMCs consortium. Neither this document nor the information contained herein shall be used, duplicated, or communicated by any means to any third party, in whole or in parts, except with prior written consent of the eSMCs consortium.
## Document history

<table>
<thead>
<tr>
<th>Revision</th>
<th>Date</th>
<th>Modification</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24.03.2015</td>
<td>Initial version</td>
<td>Thomas Buhrmann</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ezequiel Di Paolo</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Xabier Barandiaran</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hanne De Jaegher</td>
</tr>
</tbody>
</table>
Agency and eSMCs
An examination of the relation of eSMCs and autonomous agency and the conditions for the spontaneous emergence of a sense of agency

Two trends permeate contemporary approaches to agency in a manner relevant for sensorimotor theory: the rupture of the dichotomy between “reasons” (mental) and “causes” (physical) as a departure point to approach the notion of agency; and the neurodynamic and experimental approach to the “sense of agency”. Both trends open the way for a naturalized conception of agency and the experience thereof. As we will argue in this report, such a conception can be rooted in the basic properties of organizations sustaining life and/or adaptive sensorimotor organizations (“mental life”).

Table of Contents

Document history

1. Introduction: SMCT and the need for a theory of agency 4
2. Theories of agency: A short overview 6
3. Minimal requirements for enactive agency 8
   3.1 Identity 8
   3.2 Interactional asymmetry 9
   3.3 Normativity 11
   3.4 Defining agency 12
   3.5 Autonomy and adaptivity 13
   3.6 Glossary 15
4. SMCT and agency requirements 17
   4.1 Biological autonomy 17
   4.2 Sensorimotor autonomy 20
      4.2.1 Habits and Mental Life 21
      4.2.2 Habitual SMCs and the conditions for agency 26
      4.2.3 Neuroscience 31
5. Higher forms of agency 32
   5.1 Sense of agency 33
   5.2 Social agency 40
6. Models 45
7. Discussion and conclusions 49
References 51
1. Introduction: SMCT and the need for a theory of agency

The notion of agency has subtly different connotations in the fields of biology, psychology, sociology, philosophy, neuroscience etc., some of which we will sketch in section 2. But few would disagree with a definition of agency as “the capacity of a subject to act on its own behalf”. Unfortunately, such a definition carries no meaning whatsoever, unless its component terms, such as “subject”, “act”, or “behalf” are also given a naturalized interpretation. The problem of agency thus boils down to the question of how to define, in a non-circular manner, what a subject is, i.e., what it means to be a well-defined system having its own perspective on the world, and what allows such subjects to behave intentionally, i.e., a concern for achieving certain self-related outcomes. That this is not a trivial endeavor is exemplified by a recent target article in a volume on the psychology of agency (Gruber et al., 2015, ch. 4), where a loose shopping list of inconsistent and seemingly circular agency definitions is provided. For example, for the author of the article (and many similarly confused attempts can be found in the literature on the subject), agency is constituted by the acts a person performs (and a reflex is considered as the most basic “act”). But what is an act, as opposed to mere movement, if not the kind of thing an intentional agent does?

It is exactly the same problem sensorimotor contingencies theory (SMCT) faces if it is supposed to serve as a theory of cognition in general, rather than of perceptual experience only. For example, O’Regan and Noë (2001) argue that you are perceptually aware of your situation, or simply perceive your situation, when you not only exercise your knowledge of the relevant sensorimotor contingencies (SMCs), but also integrate this exercise with processes of action-guidance. In other words, SMCs need to be an integral part of intentional actions. But sensorimotor theory provides no explanation of where this intentionality originates, or who the system is exercising knowledge of SMCs. Hence, as Thompson (2005) has pointed out, SMCT as originally conceived only accounts for the “easy” problem of experience, namely that of identifying what distinguishes the phenomenology of different sensory modalities or instances of perception. It does not, however, fully answer the question of why the enactment of SMCs is accompanied by any experience at all, or what characterises the kinds of systems that can have them. Clearly, something crucial is missed, when according to the original formulation of SMCT any machine can be said to have perceptual experiences as long as it uses internal processes involving SMCs to choose between different movement outcomes—as e.g. O’Regan and Noë’s visually-guided missile (2001).

A similar argument can be made for the sense of agency. According to contemporary explanations, it arises, broadly speaking, from reflections on internal representations of matches between motor intentions and outcomes. This is incompatible with SMCT, according to which perception is constituted by a relation between agent and environment, not by internal representations in the brain (see Deliverable 1.7 on virtual actions for further discussion on this issue). But SMCT mostly deals with transitive experiences, i.e. experiences of objects in the world, while the the basic sense of agency, as we will explain in more detail later on, is an intransitive experience, an experience of how or who, rather than what.
We argue in this report that the enactive approach may be able to complement sensorimotor theory and fill the explanatory gaps identified above. We believe it can do so by providing a naturalized account of normativity and subjectivity in terms of an organism’s organizational properties. To achieve this integration of sensorimotor theory with the enactive account of agency, after providing a brief overview of the different theories in section 2, we start in section 3 by summarising the requirements for minimal agency (Barandiaran et al., 2009), which are: i) sustained individuality through a process of self-maintenance under precarious conditions; ii) asymmetry in the interaction with the environment, i.e. being able to “act” by modulating environmental conditions and not being driven exclusively by the situation; and iii) normativity guiding action toward sustaining (and improving) the agent’s viability.

Section 4 then evaluates the usefulness of minimal agency for sensorimotor theory when grounded in biological autonomy (4.1), and proposes an independent level of sensorimotor autonomy (4.2) that may prove more fruitful for SMCT. Our interpretation of sensorimotor autonomy (or mental life) also tries to synthesise the idea of minimal agency with earlier work of ours on the dynamical interpretation of SMCs (Buhrmann et al., 2013), as well as our Piagetian development of principles underlying the adaptive organisation of an ecology of SMC-based skills (Di Paolo et al., 2014).

Section 5 then considers higher levels of agency, such as the sense of oneself as an agent (5.1), and socially constituted agency (5.2). With respect to the former we argue, for the first time in this area, that a sensorimotor-based account of the sense of agency consists in nothing other than applying the pragmatic turn once more to the experience of one’s own activity. In other words, the sense of agency, from a sensorimotor perspective, cannot be based on internal representations of one’s own actions (such as forward models, emulators, or “comparators”), as contemporary cognitive theory suggests, but just as other forms of perceptual experience, must be found in the sensorimotor relation between agent and environment. We thus develop in 5.1 an account of how the sense of agency may arise from modes of adaptive regulation in an (enactively conceived) sensorimotor agent (as developed in 4.2).

Concerning the question of social forms of agency (section 5.2), we examine how situations involving the presence of other active agents may affect the account we offer in this report. We demonstrate in which important ways social agency differs from the “additive sum” of the sensorimotor activity of two or more coupled agents. This is due to the mutuality of their actions and intentions as well as to an additional factor: the autonomy of relational interaction patterns which may sometimes involve agents in actions that they do not individually intend explicitly. We describe how social agency can acquire different complex forms in terms of this tension between the individual and the interactive order of normativity.

Section 6 then discusses a (non-comprehensive) list of models addressing different aspects of minimal agency. While no model so far covers all requirements, it shows that the three conditions can be operationalised and are amenable to empirical testing.

We conclude by examining the relation between the proposed agentic grounding of sensorimotor theory and previous accounts of eSMCs as developed in the project.
2. Theories of agency: A short overview

The problem of agency emerges in different disciplines and scientific fields, from biology to philosophy, from engineering to social sciences. The focus of attention is different for each field and the tools and conceptual distinctions made are also specific to the problem domain under study. Nevertheless some general features remain common, particularly among naturalistic approaches that don’t demand a clear-cut distinction between reasons and causes (or between reasons as causes and other kinds of causes) and thus make possible a continuity solution for a concept of agency that spans from cellular motility to collective or institutional agency. Here we summarise very briefly some of the conceptions and debates concerning the question of agency in philosophical psychology, biology, and robotics.

**Philosophy** is the field in which the concept of agency has attracted most attention and where more explicit theories of agency have been developed. Moreover, theories of agency have been developed from specific branches of philosophy, e.g. philosophy of biology, philosophy of mind, moral philosophy, etc. often merging their contributions with that of scientific fields. We shall here briefly focus on the most purely philosophical contributions. Most notably, what we might label as a “rationalistic” school, has taken agency to fix a dualistic divide between actions and events (Davidson 1963), or to say it more accurately that the phenomenon of agency ultimately refers to some primary reasons that cannot be reduced to events. Reasons are causes of actions, whereas events (other than actions) have other events as causes. Davidson defines reasons as a combination of a pro-attitude (e.g. a desire) towards an end and the belief that the action performed is a means to that end. In his account, agency demands high cognitive capacities, whereby beliefs about means-ends relationships between actions and goals need to be explicitly justifiable to the agent, it is the propositional content of the agents attitudes (beliefs and desires) that causes the action. Discussions along this tradition include questioning the nature of this causation, the relationship between intentions and actions, the right conditions under which intentions, actions and outcomes need be aligned for an event to be considered a genuine action and the cognitive nature of intentional representations and their causal status (Wilson and Shpall, 2012). The cognitivist evolution of the analytic Anglo-American philosophical tradition has since conceived agency as tied to the capacity of a cognitive system to decide upon the different courses of action based on a representation of oneself and of the outcomes of such actions. And discussion of agency often conflates with discussions about autonomy and free will or more fundamental or basic discussions concerning the nature of mind and cognition. Also along this tradition, some philosophers have lowered the cognitive weight for agency, Frankfurt (1978) opened the notion of agency to more basic goal oriented behaviour, whereby spiders for example, or other non-human sub-rational systems, could be considered agents. Interestingly enough, and perhaps in an attempt to avoid the controversial dispute between cyberneticians and phenomenologists concerning the nature of purposive behaviour (see Rosenblueth, Wiener & Bigelow 1943 and critiques from Taylor 1950 and Jonas 1953), among the conditions stated by Frankfurt for genuine agency is that the system need be somehow identified with the mechanisms guiding behaviour (this would leave out of the agentic category things like pupil dilatation). We will turn to this connection with identity latter. For now it suffices to state that when analytic philosophy has dealt with the issue of agency it has either restricted the phenomenon to higher level human reasoning capacities, or, when opening up the spectrum to encompass a wider set of agentic phenomena, issues of control, goal-directedness and
living purposefulness (irrespective of human rationality) enter the scene. And this brings us down to biology (and its interface with physics and chemistry), where the concept of agency has and still plays a relevant role.

In biology the question of agency arises at the origins of living organization (Kauffman 2000, Kauffman 2003, Moreno and Etxeberria 2005, Barandiaran, Di Paolo & Rohde 2009, Skewes and Hooker 2009). Whereas inanimate forms and structures passively suffer the consequences of other forces, or appear governed by statistical uniform laws, living systems are highly structured organizations where energy is channeled to produce work in a manner that is both evolutionarily and self-organized oriented to self-maintenance and adaptively channeled to manage system-environment relationships “so as to” satisfy a set of viability constraints. In other words if one is to observe the behaviour of bacterium and compare it with the behaviour of a rock, the tendency of the former to adaptively change its behaviour to improve its survival conditions stands out in comparison with the passive or blind course of the behaviour of the latter. Sensors, transducers, and mechanically articulated movement make possible to talk about actively sustained and guided interactions between what, in comparison to a rock, an atom or a drop of water, can be genuinely called “an agent”. Problems addressed by the biological conception of agency include the problem boundaries and membranes, biochemical organization of energy flows (Kauffman 2000), self-maintenance and viability (Di Paolo 2005, Barandiaran and Egbert 2014), self-organization and evolution of adaptation, etc. Overall, two major schools of biological thinking dominate agency theories in biology: those that link agentic capacities to the fact that certain behaviour generating mechanisms have been selected for the behavioural goal they achieved in the past (Millikan, 1989), and those that look at the systemic properties of the organism and justify the purpose of an action in terms of how it is initiated and evaluated within the physiological and environmental context.

Another research field where agency becomes an issue is autonomous robotics. The study of adaptive behavior using situated robot models in the 1990s has led to an explosion of proposed definitions of agency (Wooldridge & Jennings, 1995). However, notions of agency in this field remain bound with local and practical considerations and are sometimes circular. For instance, Russell and Norvig in their classical AI handbook (1995, p. 33) propose that “an agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through effectors.” Maes (1994, p. 136), on the other hand defines an agent as “a system that tries to fulfill a set of goals in a complex, dynamic environment”; Beer (1995, p. 173) considers an agent “any embodied system [that pursues] internal or external goals by its own actions while in continuous long-term interaction with the environment in which it is situated,” while Smithers (1995, p. 97) states that “agent systems are systems that can initiate, sustain, and maintain an ongoing and continuous interaction with their environment as an essential part of their normal functioning.” After an extensive review of different definitions of agency, Franklin and Graesser (1996, p. 25) conclude that “an autonomous agent is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future.”

While certainly relevant to a sensorimotor approach, these approaches to agency in philosophical psychology, biology, and robotics need to be somehow integrated and contrasted with each other using
a framework that attempts to resolve the question of agency from first principles. We propose such a framework in the remaining sections of this report.

3. Minimal requirements for enactive agency

As discussed in the introduction, one goal of the enactive approach to agency is to provide a naturalized theory of what it means for "a system to act on its own behalf". Specifically, it is the attempt to unpack the previous phrase by answering the following three questions:

- How do we distinguish a “system” from the background network of complex interacting processes in which it is embedded – in operational terms?
- What constitutes an “act”, above and beyond physical movements that the system is undergoing?
- What does it mean for a system to have a “behalf”, or concern, in contrast to systems simply executing a program, for example?

In the next sections we summarise how the enactive approach aims to answer these questions (following Barandiaran et al. 2009).

3.1 Identity

In many everyday uses of the word “agent”, e.g. when referring to persons or institutions, what distinguishes the agential system from its surroundings seems to be unproblematic and is rarely questioned. Yet on reflection it becomes clear that even these use-cases are far from simple. For example, does a library as a system include the people employed in it? Or is the internet part of the library, since it connects to it via its computer terminals? Similarly, do we include in a person system the approx. $10^{14}$ bacteria co-inhabiting his or her body, or the tools he or she uses in order to achieve a particular task? Where do we draw the line in defining which astronomical objects make up the planetary system? Similar questions can be asked for chemical reaction networks in the prebiotic soup or for robots and their environments. Often what distinguishes a system from its environment is a decision made by us as observers, using arbitrary criteria such as the function it may serve for us. From a social sciences perspective one might say that being agential is a status attributed to systems by people, and therefore a subjective matter.

The enactive approach, in contrast, suggests that there are systems that do not depend on their utility to observers for their individuality. It suggests that such systems define themselves as individuals, and may be identified as such in an observer independent fashion. Only systems that manage to distinguish themselves from their surroundings, and in doing so define an environment in which their activity is carried out, are considered as candidate agents in the strong sense in this approach.

Consider as one example of such systems an autocatalytic set of chemical reactions, i.e. a set of molecules each of which can be created catalytically by other chemicals within the set, such that as a whole, the set is able to catalyze its own production. When embedded in a soup of other chemicals, such a set may also engage in other reactions that do not contribute to catalyzing its own production.
While it may seem that thereby the boundary between the set and its environment is getting blurred, at any moment there is only a single well-defined set that catalyzes its own production. In this way the autocatalytic set defines its own identity without requiring an observer drawing arbitrary boundaries.

This self-distinction is the first requirement for agency as understood enactively. Note that what defines the system as a distinct unit, i.e. what defines its boundary conditions, is determined by a functional relation that holds between and is maintained by its component processes (e.g. the catalyzing reactions), but not other processes (in the environment). In certain, usually more complex, cases this unit may further separate itself by producing a physical boundary (in addition to the functional “boundary”), but this is not a requirement for self-defining identity (see Figure 1 for a schematic abstraction).

**Figure 1:** Self-constitution of a system undergoing energetic and material exchange with its environment. The system’s activity defines its own boundary conditions, which may or may not involve an actual physical boundary (such as a semi-permeable membrane).

3.2 Interactional asymmetry

Like the autocatalytic set described above, any agent that has carved itself out from an environment will nevertheless be subject to exchanges of matter and energy with its surroundings. Note, however, that in ordinary language, when we say that an agent *performs an act* or *engages in an action*, we do not normally mean that the agent is merely 'suffering' physical exchanges within its environment. Instead, we imply that the agent somehow is *responsible for*, or *the agent of*, such exchanges. Rather than implying an equal and mutual coupling between agent and environment, we mean that the agent is the source of activity, not merely a passive sufferer of external forces. Somehow agents must therefore be able not only of coupling to their environment, but also of actively modulating this coupling from within\(^1\). This condition of *interactional asymmetry* is the second requirement for speaking of agency in the strong sense.

\(^1\) The environment of an agent, in contrast, is merely coupled to it, but cannot change the way in which it is coupled (unless it is an agent itself of course).
It is far from trivial, however, to given an operational definition of what exactly we mean by this interactional asymmetry. Intuitively it would seem right to equate it to agents being the 'cause' of certain events. In complex systems (such as our example autocatalytic set) causation itself, however, is a concept with a variety of different meanings and interpretations. Two possible approaches may shed some light on the problem.

Starting from energetic considerations, we could try to express the asymmetry requirement in terms of the capacity of a system to constrain energy flows to sustain coordinated processes, which are in turn reused by the system in a circular manner (Kauffman, 2000). Cells, for example, by coupling endergonic and exergonic reactions and channeling energy flows, produce work and can thereby maintain themselves far from thermodynamic equilibrium. This perspective matches an intuitive notion of action and agency: the system is the energetic drive of an otherwise neutral or spontaneous coupling with its environment (actively pumping ions or performing chemotaxis, as opposed to the passive suffering of an osmotic burst or being moved by currents or local fluctuations in a pond). Though intuitive, one can imagine cases where explanations of this kind are difficult to uphold, say for example, in the case of a gliding bird that is being carried by the wind without having to exert energy. We would argue that gliding is certainly an action that a bird can engage in (with the purpose of travelling while conserving strength), yet on grounds of energetic considerations alone this seems difficult to defend.

Alternatively one may try to use statistical considerations to assess the extent to which one system (e.g. a candidate agent) influences another (its environment). For example, one may identify a system as being the agent of an interaction when changes in its behaviour-generating mechanism precede environmental changes in a statistically significant manner. Apart from certain mathematical complexities involved (such as non-stationarity and non-linearity of the data), we again find cases that seem to be difficult to capture this way. For example, to a first approximation, it would seem that the event of somebody falling off a cliff, and the action of somebody taking a dive into the ocean, are difficult to distinguish on the basis of correlations between the moving body and its environment alone.

We propose that a dynamical systems perspective may be used to capture the intuitive notion of interaction asymmetry. Consider a system (S) and its environment (E) as two coupled dynamical systems of the form

\[
\frac{dS}{dt} = F_Q(S, E) \\
\frac{dE}{dt} = G_Q(S, E) \\
p \subset Q, \Delta p = H_T(S)
\]

where the coupling is represented by the functions F and G, parameterised by a set of conditions and constraints Q. Asymmetric modulation of this coupling is described by the system's influence on a subset p of these constraints, which it controls via the function H during the time interval T.

Consider again the cliff diver. Before his jump, the diver is interacting with the ground underneath his feet. A sequence of muscle movements (changes in S) results in a dramatic change in the constraints that modulate the coupling with the environment (\(\Delta p\)) leading the system to engage in free-fall dynamics. Here we must, of course, notice that were the diver not poised at the edge of a cliff, the same
sequence of muscle movements resulting in jumping forwards would produce a very different effect. We therefore emphasize the fact that actions are i) contextual (on Q) and temporally extended (T).

Figure 2: The self-constituting system exerts control over its structural coupling to the environment (green arrows).

With this dynamical interpretation in mind, we can define an agent as a system that systematically and repeatedly modulates its structural coupling with the environment (see Figure 2). Of course such agents are not necessarily able to modulate all conditions of their coupling all the time. Our definition only requires that a system is capable of engaging in some modulations of the coupling and that it does so at least in certain cases.

3.3 Normativity

One dimension is still missing in our account of agency. Actions performed by an agent are commonly understood to be different from mere movements in the sense that they are performed according to the agent's goals and norms. The spasms of someone suffering Parkinson's disease, for example, are not usually considered actions, even though the person in question is undoubtedly a self-distinguishing identity capable of modulating its environmental coupling in other cases, i.e. an agent of many other possible actions. It is obviously not enough that a self-distinguishing individual be itself the active source of modulation of its coupling with the environment. For such a system to be the agent of proper actions, this modulation also has to be carried out in relation to the agent's goals or desires.

When systems actively modulate their interactions with respect to norms, we speak of regulations of their coupling. Such regulations introduce a normative dimension. The system's actions can succeed or fail. We call this the normative condition, and it is the third requirement for agency.

But what is the origin of these norms, and how can an agent take them into account in its actions? Like the requirement for (self-determined) individuality, the normativity in question cannot be the result of observers making judgments on behalf of the agent about the “adequacy” of its behavior in relation to some of their own norms, standards, or goals. Rather, in our naturalized approach, we need to justify this normativity based on the very nature of the agent itself. For example, with the identity requirement
in mind, certain regulations may support the processes that distinguish the agent from its environment and define it as an identifiable unit. Other regulations, in contrast, may interfere with these processes and threaten to break down the system as a unit. At least one source of intrinsic norms thus originates in the very organization of the system that maintains itself as a unit, and in this sense an agent's actions, and environmental events, can be good or bad for its continued existence.

3.4 Defining agency
The three requirements of individuality, asymmetry, and normativity seem to capture most senses of the notion of agency in common use. Neither of them is sufficient on its own, but the set as a whole seems to be necessary and sufficient.

We can now provide a generative definition of agency, i.e. a description of an organization capable of generating and satisfying the three requirements. In short, an agent is defined as an autonomous organization capable of adaptively regulating its coupling with the environment according to the norms established by its own viability conditions. Unpacking this definition further:

A system S is an agent for a particular coupling C with an environment E iff:

1. S is an open autonomous system in an environment E, meaning that:
   a. 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, so that isolated from the network any component process would tend to run down or extinguish;
   b. the set of processes (not belonging to S) that can affect S and are affected by S defines S's environment (E); and
   c. S depends on certain conditions (specified by S) that in turn depend on E.
2. S modulates the coupling C in an adaptive manner:
   a. where modulation indicates an alteration (dependent on S) in the set of constraints that determine the coupling between S and E;
   b. and adaptive means that the change in the coupling C contributes to the maintenance of some of the processes that constitute S.

This definition has the advantage of being at the same time generative (i.e. our three requirements for agency follow from it), and non-circular (it does not rely on terms presupposing the notion of agency). Statement 1 captures the requirement for individuality. An important aspect here is the precariousness of the system. The organization not only defines the system, but it is also thanks to it that the system endures in time, as without it the component processes would run down (i.e. the system is self-sustaining). Normativity in turn emerges from exactly this property, i.e. from the manner in which specific interactions can either help support or threaten to break down the system's self-maintenance. Statement 2 captures the requirement of interactional asymmetry. There is a specific sense in which the system is the source of actions, for not only is it modulating the coupling, but it is doing so in relation

---

2 I.e. we cannot think of any examples of systems that satisfy all three conditions without considering them agents, nor can we think of any empirical agents that fail to satisfy any one of these conditions.
to intrinsic norms. In other words, it is the organization of the system (from which the norms emerge) that is determining the modulation of the coupling.

Although the set of criteria and the generative definition proposed can already be put to use to determine whether or not a system is an agent, it would be desirable to further ground the concept in the specific organizational properties that minimal examples of agency have to possess in order to meet our definition. In the next chapter we will look at two candidates: agency grounded in the biological organization of living systems; and agency grounded in the autonomy of sensorimotor organizations.

3.5 Autonomy and adaptivity
The notion of agency summarised in the previous sections is closely related to other non-trivial concepts such as autonomy and adaptivity (Di Paolo, 2005). The condensed definition of agency after all defines agents essentially as adaptive autonomous systems.

As we have noted above, autonomy captures the requirement that agents actively produce and maintain themselves as self-distinct units. This implies a degree of closure of the system’s internal processes. The agency definition identifies this closure as

$$\forall p \in S : r \rightarrow p \rightarrow q \mid r \in S \land q \in S$$

which defines the autonomous unit S as the network of processes p which both depend on (at least) one other process in S and enable (at least) one other process in S (see Figure 3). Note that this is not complete closure in the algebraic sense. Processes p in the system S are not necessarily such that all their “products” q are also processes in S. In this sense the system is also open: it can be affected by processes outside of itself, and in turn affect such processes. In other words, it is (energetically, materially, dynamically) open to interactions with its environment.

The nature of the abstract relation $r \rightarrow p$ depends on the domain in which the system is realised. In principle it could be that of a catalyzing reactions between molecules in a chemical domain, as well as dynamical or causal dependence in other networks of interacting processes. However, the autonomy definition further requires that component processes in isolation tend to run down (precariousness, see e.g., Di Paolo, 2009). The closure required for autonomy therefore is of a special kind, namely one that has to be actively maintained by the same component processes. Certain systems (e.g., crystals) may therefore be said to exhibit closure of component processes (in the crystal’s molecular interactions), but not autonomy (since the closure is and need not be actively maintained).

The network definition of autonomy also allows for a graduated interpretation (Collier, 2002). At minimum it requires that each process is dependent on, as well as a precondition for, another process. But this leaves open how strong or manifold the dependencies in the network are, especially relative to the system’s interactions with the environment. The internal identity-reinforcing tendencies might be relatively strong or weak compared to the potentially destabilising tendencies an agent’s surroundings may exert. One can therefore in principle distinguish different levels of autonomy, even though it may not always be easy to quantify this.
Figure 3: An open autonomous network of interacting processes. Each process in the closed system (black) both depends on and enables another process in it. Gray nodes indicate other processes that the system may interact with (green arrows).

Adaptivity introduces another gradual dimension to being an agent. Without adaptivity the definition of agency would include only rather static entities. Agents would be more or less robust to environmental perturbations, but at any point the system would be either a self-maintaining agent or dead, without any room for agents to be sensitive to their current (or future) conditions, and to act such as to improve their prospects. This in turn would imply a rather poor normativity: an encounter would be good if it does not endanger an agent’s self-maintenance, and otherwise bad (and leading invariably to the system’s demise). Clearly, agents usually need to be capable of more nuanced evaluations to be viable, e.g. by being sensitive to the risk of disintegration, or to gradients and directions in their viability conditions (e.g. differential levels of sugar concentrations).

Di Paolo (2005) defines adaptivity as:

“A system’s capacity, in some circumstances, to regulate its states and its relation to the environment with the result that, if the states are sufficiently close to the boundary of viability,

1. tendencies are distinguished and acted upon depending on whether the states will approach or recede from the boundary and, as a consequence,
2. tendencies of the first kind are moved closer to or transformed into tendencies of the second and so future states are prevented from reaching the boundary with an outward velocity.”

---

Being adaptive, i.e. actively self-monitoring and regulating, allows agents to extend the all-or-nothing normativity originating in its self-production. Only adaptive agents have the capacity to improve their living conditions by assessing their current state relative to the norms given by self-maintenance, and acting on it in a graded and directed manner. This also allows adaptive agents to make much finer distinctions between situations that otherwise would be identical in terms of the consequences for current self-maintenance (e.g. different levels of sugar concentration surrounding a bacterium that are equally sufficient to maintain its metabolism). Adaptivity is thus a precondition for sense-making (Thompson, 2007).

By sense-making we refer here to the notion that objects or events become meaningful for an agent if they are involved in the normatively-guided regulation of the agent’s actions (e.g. by triggering or mediating it). An approaching stewardess, for instance, becomes meaningful in relation to the opportunity for asking for a glass of water. The glass of water, in turn, is meaningful in relation to one’s intention to drink and so on. The meaning that a particular event entails, can be equated to the regulative response it provokes. Hence two events that trigger the same action are not meaningfully distinguishable (from the point of view of the agent); and the severity of an event tending to destabilize an agent’s self-maintenance is nothing but the amount of regulative resources required to compensate it (Di Paolo, 2005). In short, adaptive regulation is a precondition for agents’ having a subjective outlook on the world. Without it the only meaningful events an agent would ever encounter would be those leading to its death. Without sense-making an agent would not be able to appreciate his standing relative to its own viability conditions while it is still alive. There would be no sense in which an agent could be under stress, malfunctioning, or ill (as from its own perspective it would only ever be dead or alive). Equally he would not be able to avoid potentially risky situations, or seek preferable ones (as long as he is currently alive).

3.6 Glossary
Here we provide a short summary of the more important notions in the enactive approach to agency.

**Individuality / Identity**
Given a particular functional relationship between processes, the network S of such processes in which every process has both an incoming as well as outgoing relationship with other processes in the network.

**Interactional asymmetry**
A self-defining system S modulating the set of constraints determining its coupling with the environment.

**(Intrinsic) Normativity**
“Value” derived from interactions either helping to maintain or tending to perturb a system’s self-maintenance.
(Open) Autonomy
A self-defining network of component processes in which the activity of the network is required to maintain itself against the material and energetic tendency to break down (precariousness).

Adaptivity
Asymmetric modulations of system-environment interactions that follow the system's intrinsic normativity.

Agency
The property of being an (open) autonomous and adaptive system.

Meaning
The normatively guided regulation an event provokes in the agent.

Value

Sense-making
The evaluation by the agent of environmental aspects and events in terms of the normatively guided regulations they provoke.

Intentionality
Actions are intentional because they are normatively directed regulations of an agent's interaction with the environment (not just symmetric exchanges between the two). Their purpose, or goal, derives from an agent's adaptive response to changes in its viability.

Viability
The condition of a system remaining the sort of system it is, i.e., conserving its individuality/identity, whether in organizational or functional terms.
4. SMCT and agency requirements

As already mentioned, we can in principle envision different domains in which the conditions of agency may be satisfied. Here we illustrate two, namely the domains of biological and sensorimotor autonomy.

4.1 Biological autonomy

The enactive definition of agency described above, while deliberately abstracted from the medium in which it may be realised, can in fact be considered a distillation of system-theoretic approaches to understanding minimal living (biological) organizations (see Barandiaran et al. 2009, for a short review). A common view has emerged, for example, of the living cell as a complex system situated far from thermodynamic equilibrium, at its core a metabolic network of chemical reactions that produces and repairs itself, including the generation of a membrane that encapsulates the reaction network while actively regulating matter and energy exchanges with the environment. Exchange through its semipermeable membrane is necessary to maintain a sufficient level of precursor chemicals required to fuel the cell’s metabolism, as well as to expel its waste products. The cell’s metabolism, by determining e.g. which precursor molecules it requires for self-maintenance, or which chemicals in the environment are detrimental to it, brings forth a particular (chemical) environment for the particular cell, and with it a basic normativity of what is good or bad for it. It easy to see that such a system fulfills the agency conditions of individuality and normativity. But in which sense do basic living organizations, such as single cells, adaptively regulate their environmental interactions?

One example is the regulation of osmotic pressure within cell walls. The natural tendency for a cell interior composed of macromolecules and surrounded by a semipermeable membrane is to create a solute gradient with respect to the exterior and thereby drive water through the membrane and into the compartment. To prevent bursting, the cell can regulate properties of its cell wall (e.g. the activity of ion channels) and thereby stabilise the osmotic pressure. Ruiz-Mirazo and Mavelli (2008) have shown a similar mechanism at work in a proto-cell model that achieves control of internal pressure resulting from buildup of waste products. Here the same metabolic network responsible for creating and maintaining the cell, if necessary, produces peptides that are incorporated into the cell wall and thereby increase its elasticity and/or permeability. The proto-cell thus regulates its own boundary conditions, as depicted schematically in Figure 2.

At this point one may wonder to what extent the normal unfolding of behaviours can be distinguished, formally, from their adaptive regulation? Or, in other words, is the cell really adaptive, or simply robust to a certain set of possible perturbations? Although there is no final answer yet, as briefly indicated above, various approaches may be used to distinguish the two modes of activity. From a statistical point of view, for example, the normal goings-on might be associated with stationarity of the system, while regulation introduces an element of non-stationarity; energetic considerations may uncover that the system needs to perform work in order to achieve regulation (by creating constraints coupling exergonic and endergonic reactions), while regular activities are driven by spontaneous exergonic reactions alone; viewed dynamically the regulation may correspond to a parametric change in the system, with the resulting qualitative change in attractor dynamics; also, a regulatory subsystem may be structurally distinguishable from the primary self-maintaining system. But independent of the chosen level of description, it is important to note that the conditions for agency, though not subjective, are
associated with a certain timescale and spatial granularity. For example, what counts as variables, parameters or constraints in a dynamical description depends on the chosen timescale (and on large enough timescales no system is truly adaptive, as eventually they must all succumb to the second law of thermodynamics). Equally, an autocatalytic set achieves its functional closure at the level of molecules and their relations only, and no such closure can be found at the atomic or quantum level. It is thus perfectly possible to miss the presence of agency altogether when considering different levels of description. But once a level is chosen, there will be signs (like the ones listed above in the paragraph) of whether adaptive regulation is happening.

A case where the separation between regulated and regulating subsystems can be appreciated rather easily is that of the Lac-Operon mechanism in *E. coli* bacteria (Jacob and Monod, 1961). Under normal conditions *E. coli* metabolizes glucose. But when availability of this sugar is low while another (lactose) is abundant, certain normally inactive genes will be expressed that enable a new metabolic pathway allowing for the processing of the new sugar. In effect, the bacterium detects a change in environmental conditions that jeopardizes its self-maintenance and reacts by modifying the internal processes underlying its self-construction. It is easier here to distinguish the regulation of behaviour from its normal execution, as the normally dormant genes that are activated contingent on specific environmental conditions do not take part in the on-going self-sustaining processes of the organism.

But the adaptive regulation required for agency need not be of internal processes (such as the metabolism in the previous example). That which is regulated can also be the system’s mode of interaction with the environment. Consider, for example, a simple motile cell exposed to toxic chemicals. One mechanism for dealing with this situation could involve internal processes of rendering the dangerous chemicals harmless, or expelling them at sufficient rates. Another option, however, would be for the cell to simply move away from the source of the danger. Motility thereby allows for regulation with respect to variables that the agent has no direct control over (unlike, say, the buildup of internal waste materials). Moreover, in motile systems that can affect as well as sense changes in their environments, the circular interaction itself can become a target for modulation (Figure 3).
One example of this is the chemotactic behaviour exemplified by *E. coli*. The mode of movement of this bacterium involves straight forward motion (“running”), interspersed with random changes of direction (“tumbling”). When the organism detects a positive gradient of attractants (chemicals that under normal circumstances it can metabolise) it performs running motion and as a result moves towards higher concentrations of the attractant. If it detects a negative gradient, in contrast, tumbling frequency is increased in search for more profitable locations. While the ability to sense gradients itself depends on adaptive processes (see e.g. Alon et al., 1999), what is perhaps more interesting is the regulation of the chemotactic behaviour. In principle chemotactic behaviour can be carried out independently of the organism’s metabolism; i.e. sensed changes in the level of attractants are coupled directly to the motion of the flagellum responsible for the bacterium’s motion. However, a number of chemical pathways seem to exist in *E. coli*, which link this basic sensorimotor loop with processes of metabolic regulation. It has been shown, for example, that

- sensitivity to attractants is increased after starvation
- mutants that cannot metabolize a specific attractant do not show chemotaxis toward that particular sugar, even though the mutation has no effect at all on the function of the sensory receptors or other elements of the pathway underlying the sensorimotor loop
- the presence of a metabolizable chemical prevents chemotaxis to all attractants
- inhibiting the metabolism of a chemical stops chemotaxis towards that particular attractant (and only to it)
• some bacteria show a direct correlation between the efficiency of a chemical as a growth substrate and sensitivity to it.

Alexandre and Zhulin (2001) have demonstrated that these metabolism-dependent modulations of chemotaxis rely on the bacterium’s ability to monitor its own metabolism (e.g. by sensing metabolites in the internal milieu directly, or the flow through the electron transport system). As a result, chemical substrates can be “dynamically interpreted” as attractants (or repellents/neutral), as a function of whether they’re metabolizable at any given moment in time.

The view that emerges thus of motile adaptive agents is that of two interacting circular processes: firstly the cycle of self-construction underlying the agent’s basic autonomy; and secondly the somewhat independent sensorimotor cycle. Coupling of the two, as exemplified in *E. coli*’s metabolism-dependent chemotaxis, implies that sensorimotor behaviours can be regulated with respect to the norms originating in the agent’s self-maintenance. And they represent, therefore, the most basic form of actions.

While examples such as these provide clear illustrations of agency in action, it is not necessarily the case that agency needs to be based in living organization (in biological media). What is essential is that the conditions for agency, namely individuality, asymmetry and normativity, can be grounded purely in the particular form of organization of the system in question. We will next investigate to what extent they may be found in complex sensorimotor structures. Specifically, the question we like to answer is whether sensorimotor loops have to obtain their normativity directly from the system’s self-maintenance to be considered actions, or whether we can imagine other types of autonomy where what is preserved, and therefore what grounds normativity, is not the material basis of the organism.

4.2 Sensorimotor autonomy

Despite the undeniable biological embodiment of cognitive systems, extending sensorimotor contingencies to cognition demands, at the level of agency, a consideration of the transition to a new form of autonomy (i.e. of identity and normativity) that has a status of its own and serves as a bridge between lower level (biological) agency and higher order (rational, linguistically articulated and self-reflexive) agency.

We have shown so far how even the simplest organisms are capable of performing normative actions, by coupling sensorimotor loops more or less directly to processes of self-maintenance. It is clear, however, that much of our everyday behaviour (and the accompanying neural activity in the brain), while taking place within the constraint of metabolic viability, is strongly underdetermined by it. Many of our actions, for example, acquire intrinsic value on top of their primary functionality: movements can be dexterous, postures awkward, a walk elegant, and so on. The question hence arises of whether stable behavioural patterns reliably enacted in given situations, which are constituted by SMCs, can be elevated to the status of normative actions by grounding them in their own form of autonomy. Since SMCs are agent-environment relational structures, the right level to look for this kind of autonomy is the behavioural domain. The enactive proposal (Barandiaran, 2007) is that SMCs (of a certain kind) can
form complex networks of interdependencies that at least in principle could constitute the substrate in which individuality, asymmetric regulation, and normativity may arise. The broad idea in what Barandiaran (2007) calls “mental life” is that nervous systems are sufficiently complex and decoupled from metabolism to create their own mode of agency, one which results from adaptive mechanisms subserving the preservation of internal behavioural stability and coherence.

4.2.1 Habits and Mental Life
Barandiaran (2007) and Di Paolo (e.g. 2005) suggest that the behavioral analogue to living agency is a network of dissipative but interactively self-sustaining sensorimotor structures, namely “habits”, which emerge in the domain of the continuous reciprocal interaction between brain, body and environment. The notion of habit employed here is rather different from (at least some) traditional and contemporary interpretations. Barandiaran and Di Paolo, (2014) offer a succinct history of the concept (see Figure 5). They broadly sketch two large historical trends: an associationist trend that sees habits as the linking of pre-existing elements (ideas, stimuli and responses, etc), and an “organicist” trend that considers habits as dynamic wholes where parts mutually shape each other. Habits, for our purposes, are not static stimulus-triggered response automatisms, which are typical of how the notion was studied in behaviourism. Rather, following some of the organicist thinkers, they are plastic, situated sensorimotor structures nested and composed at different temporal and spatial scales. Moreover, they are characterised by a particular form of self-maintenance, in that they are both cause and effect of their own occurrence. Habits of this kind can be defined as

*self-sustaining and plastic patterns of behaviour* that are formed when the stability of a particular structure of sensorimotor correlations is dynamically coupled with the stability of the mechanisms generating it.
Figure 5. Conceptual genealogy of the habit concept (Barandiaran and Di Paolo, 2014). The plot at top-left corner of the map displays the n-gram of the terms “habit” and “representation” in non-fiction literature (in English) published between 1850 and 2008 and scanned by Google. Green circles indicate positive contributors to the concept of habit. Red stars indicate breaks in the development or significance of the habit concept. A high definition version of the map can be found here: http://barandiaran.net/design/habit-map.

According to this definition, once a habit is formed, its recurrence is a condition for its own continuation – the habit “calls” for its exercise and its exercise reinforces its permanence (Figure 6). One way of understanding this circularity is to say that the stability or recurrence of the behaviour that the habit involves (smoking, reading, jogging) both depends on and reinforces the mechanisms that give rise to it. Habits therefore imply plasticity, which involves the capacity of a habit to be shaped or deformed under sensorimotor or neurodynamic “pressure” (meaning perturbations or variations in its stability conditions) and to maintain this deformation as its new form. This plasticity can be manifested at the neural level (e.g., changes in synapses, changes in white matter), in the body (e.g., muscle tonality, callosities), or in the environment (niche, organised spatial patterns, interactions with others, etc). A habit also implies precariousness, in so far as it will cease to assert itself when its supporting conditions do not present themselves with sufficient frequency.
Figure 6: biologically-grounded action (A) and sensorimotor autonomous habit (B). In the former case, adaptive regulations of the sensorimotor loop by processes involved in the organism’s biological self-maintenance underlie its normativity. In the latter case, the organism creates a partially decoupled subsystem (e.g. central nervous system) subserving agent-environment interactions. When the stability of a particular mode of sensorimotor engagement is dynamically coupled with the stability of the mechanisms generating it, a habit is formed, i.e. self-sustaining behavioural pattern (represented by the circular nature of the sensorimotor loop).

It is important to note that we do not restrict the notion of habit to its common everyday sense. Looking at a handle just before we hold it to open the door and look forward right after, walking at a certain pace, breathing diaphragmatically, etc. are all examples of habits as much as smoking with a pipe or reading the newspaper on Sundays. Babies can be said to be born equipped with “innate” sensorimotor habits (mostly involving self-reinforcing patterns of stereotypical movements and correlated proprioceptive and tactile feedback) resulting from the exploration of sensory and motor spaces in the uterus (Kuniyoshi & Sangawa 2006). A process of self-organized development of habits and fixation of sensorimotor contingencies starts right after (Marquez et al., 2013); always within an ecological context that provides stable regularities and remains open to the creation of new habits. Relevant here is Piaget’s description of the transitions from reflexes to abstract sensorimotor schemas (Piaget, 1963), which we will come back to.

Habits do not stand in isolation as egotistically self-sustaining patterns of behavior. On the contrary, as a result of inter-active development habits appear nested in a hierarchical and sequential manner. Together with their precarious nature, this explains why habits may be more malleable and adaptive than the traditional picture that associates them to automatisms. The point was already recognized by John Dewey, whom we quote at length:

“Strict repetition and recurrence decrease relatively to the novel. Apart from communication, habit forming wears grooves; behavior is confined to channels established by previous behavior. In so far the tendency is towards monotonous regularity. The very operation of learning sets limits to itself, and makes subsequent learning more difficult. But this holds only of a habit, a habit in isolation, a non-
communicating habit. Communication not only increases the number and variety of habits but tends to link them subtly together, and eventually to subject habit-forming in a particular case to the habit of recognizing that new modes of association will exact a new use of it. Thus habit is formed in view of possible future changes and does not harden so readily”, (Dewey, 1925, pp 280–281).

Many complex actions are composed of webs of sensorimotor habits whose deep structure does not reveal itself until strongly disrupted. Such is the case, for example, in visual adaptation to inverted goggles reported by Ivo Kohler (1964), who points to the idea that “habits exist in all areas of human personality” (Kohler 1964: 137) and that only through a strong process of re-habituation “do we notice what habit is, and to what extent we consist of many and strong habits” (Kohler 1964: 138).

Barandiaran (2007) envisions the development of such bundles of habits along the following lines. At first, neurodynamic structures (such as dynamic cell assemblies, Varela, 1995; Tsukada et al., 1996) are created that sustain different sensorimotor couplings with the environment. The formation of these structures might originate in early architectural constraints (like retinotopic wiring, Kirkby et al., 2013) or the fixation of self-organized patterns (like spinal circuits driven by spontaneous firing, Marques et al., 2013), and possibly modulated by biologically adaptive signals (pain, pleasure etc.). Next, interactive stability dependencies are created, both between neurodynamic structures and the behaviours they sustain (habits), as well as between neurodynamic structures themselves. Consider as a potential example of the former case the circular nature of retinotopic maps being tuned by neural activity in the retina (Kirkby et al., 2013); activity which depends on movements of the eyes, which are themselves controlled with contribution by the same retinotopic maps. Over time, a nested web of neurodynamic structures and corresponding sensorimotor coordinations appears, which becomes progressively more independent from biological adaptive signals and innate architectural constraints and more dependent on: (a) higher order stability dependencies between individual habits, and (b) the interactions that they altogether sustain with the environment. Eventually, the adaptive regulation of behaviour to preserve the bundle of habits becomes the main organizational principle of brain activity and behaviour. At this point we may speak of mental life (Figure 7).
Figure 7: Mental Life (reproduced with permission by X. Barandiaran): a network of mutually enabling and sustaining sensorimotor organizations (SM schemes), the regulation of which is directed at the preservation of the network as a whole. This creates a new closure in the sensorimotor domain, decoupled but not independent from the agent’s primary biological self-production. This domain is relational, it exists in the space of agent-environment interactions, not, say, in the brain.
4.2.2 Habitual SMCs and the conditions for agency

**Individuality**

In the enactive account of biological agency an identity is created through a kind of organizational closure, in which the collective activity of a network of interdependent reactions is necessary to maintain the component processes. The same could be said about habits and their underlying neurodynamic structures.

Firstly, habitual sensorimotor coordinations are precarious and need to be reinforced through their exercise in order to be retained (much like the extinction that occurs in operant conditioning when learned responses are no longer rewarded sufficiently). But they are also part of a greater network, and depend on other behaviours as preconditions for their exercise. It is thus the complex network of interdependent habits, each helping to sustain the others, that forms the agential unit (in analogy to chemical reactions in an autocatalytic set).

Just like biological agents, the self-sustaining network of interacting habits also brings forth its own environment. Since habitual SMCs are closed sensorimotor loops, from the point of view of the organism’s biological agency they already cross the agent-environment boundary. As a result, what is the biological agent’s environment, is different from that which is the environment for the sensorimotor agent. For example, we could argue that the flu virus is part of the environment of the biological agent, in so far as it triggers in the organism immune reactions and other defense processes. But the same virus, as such, cannot be said to be a direct part of the agent’s sensorimotor environment; we don’t perceive the virus floating in the air, although its effect on the organism do influence our sensorimotor agency and become manifested at this level too. To differentiate between the two kinds of environment, we call all external aspects that can influence the web of sensorimotor habits, other than those aspects already constitutive parts of the component habits, its context.

Finally, just as biological agents are open autonomous systems that maintain themselves through exchange of energy and matter with their environment, a sensorimotor flow through the network of habits is required for their exercise. In fact, it is a constitutive part of the habit network, as both neurodynamic patterns and environmental coupling are required to close a habit’s sensorimotor loop.

**Interactional asymmetry**

One would expect most early sensorimotor habits to be rather stereotypic, being called for in a more or less circumscribed set of situations, and unfolding largely in the same manner (i.e. reflex-like). But as the agent’s bundle of habits matures, he will gain more and more fine-grained control over a) the situations he exposes himself to, thereby pre-selecting the possible sensorimotor structures to enact, b) which SMCs to associate with a given context and c) when to modify or generalize previous SMCs or create new ones, very much in accordance with Dewey’s description quoted above. The agent thus comes to regulate the activity of its networked habits with the goal of improving the internal coherence and stability of its sensorimotor repertoire. This implies of course, that a bundle of habitual SMCs can be incoherent or unstable to varying degrees. It is easy to see how in early developmental stages infants may make mistakes of enacting the wrong sensorimotor coordination (see e.g. the dynamical systems
analysis of preservative reaching, the A-not-B error, in infants by Smith and Thelen, 2003), or of possessing conflicting sensorimotor schemes (Piaget, 1981, relates the story of children’s confusion when first trying to recognise letters that look identical in the mirror and those that do not). The idea behind behavioural agency is that adaptive regulations at the network level allow the agent to resolve such internal tensions. How this might work we have recently sketched in a dynamical re-interpretation of Piaget’s theory of equilibration (Piaget, 1963, 1975, 1981; Di Paolo, et al., 2014), which we can here only sketch.

One of Piaget’s main proposals was that abstract and explicit (e.g., conceptual, formal, rational) human capacities of understanding stem from early and less explicit forms of sensorimotor schemes. By “scheme” Piaget referred to closed sensorimotor structures into which environmental aspects are assimilated when they contribute to successful completion of the loop. His idea was that we can only perceive that which can be assimilated into already existing sensorimotor coordinations. Whenever not-yet-assimilated environmental aspects are encountered, adaptive processes of accommodation transform existing schemes such as to resolve the internal tensions resulting from the perturbation to the stability of the organization of interrelated schemes. So when, for example, a baby that has successfully learned the sensorimotor scheme of suckling is first introduced to a bottle, the existing scheme may be ill-suited to this new situation. However, in a process of equilibration, i.e. through successive assimilation and accommodation attempts, the old sensorimotor sucking pattern may be tuned to the new reality, or a new scheme for feeding from a bottle may be created instead.

In a recent re-interpretation of equilibration (Di Paolo, et al., 2014) we have equated Piaget’s sensorimotor schemes with our previously operationalized notion of sensorimotor strategies, which are organized sets of sensorimotor coordinations (see Buhrmann et al. 2014). The latter can be defined as stable sensorimotor patterns that are reliably used in performing a specific task, and which constitute a specific type of SMC that can be identified in closed-loop embodied agents. And the former, the sensorimotor strategies, are sets of sensorimotor coordinations which are organized normatively (for instance, in sequence, in parallel, in their timing, and so on). Our dynamical operationalization of Piaget’s concepts helps filling in some detail about how the process of equilibration might work in practical terms, and hence shed light on how agents may adaptively regulate their environment interactions from within.

In our operationalization, individual sensorimotor schemes, or SMCs, are integrated into a greater closed sensorimotor organization that is associated with the execution of a complex tasks (so an infant’s feeding behaviour, e.g., may involve an initial orienting response to face the mother, followed by a latching on to the mother’s breast, followed by rhythmic sucking and so on). Now, we say that a specific sensorimotor coordination in class \( A \) assimilates an environmental feature or process that contributes to producing environmental time-varying states \( a' \in A' \) when the following two conditions apply (see Figure 8, left):

1. Stability condition: a sensorimotor pattern \( a = a(t), a \in A \) occurring in conjunction with an environmental pattern \( a' = a'(t), a' \in A' \) are mutually stabilized, i.e., the full agent-environment coupling does not produce sensorimotor or environmental states outside the respective sets.
2. Transition condition: if any combination of trajectories \( a \) and \( a' \) in the coupled system is produced such that \( a \in A \) and \( a' \in A' \), then this leads in time to the production of sensorimotor pattern \( b = b(t), b \in B \) in the agent and the production of states \( b' = b'(t), b' \in B' \) in the environment, where \( B \times B' \) is the next stage in the greater sensorimotor organization.

![Figure 8. Left: Illustration of the two conditions describing assimilation. Condition 1: stability. A trajectory \( a(t) \) in the projection of sensorimotor space (SM) belongs to a set \( A \) (sets are represented by gray bands), which is mutually stabilized in coupling with a trajectory in the relevant projection (E) of environmental variables \( a'(t) \) that belongs to region \( A' \). In other words, the SM trajectories in the upper panel and the environmental trajectories in the lower panel are both projections of the whole coupled system onto the respective subspaces during SM engagements of type A and B. Condition 2: transition. Trajectory \( a(t) \) in coupling with \( a'(t) \) lead respectively to \( b(t) \in B \) and \( b'(t) \in B' \), the next stage in the sensorimotor organization O. Right: Projection of O onto sensory (S) and motor (M) coordinates.](image)

Now, accommodation is triggered when one or both of these conditions are violated, and assimilation by the agent has thus failed in the given context. Note that the “distal reason” for this failure is invisible to the agent, only its effect is manifested as a disruption of the sensorimotor scheme: the loss of “control” over a previously stable sensorimotor coupling or the failure of an effectively achieved coupling to lead to its usual result. If this is the case, qualitative changes in the dynamics underlying the sensorimotor schemes are triggered until the sensorimotor organization (the loop which we associate with the idea of sensorimotor strategy) can once more be successfully completed (see Di Paolo et al., 2014 for further details).

Kohler’s experiments with visual inversions provide an illustrative example of adaptive regulation along the lines just summarised. Kohler observed that re-adaptation of the visual field always occurs only partially and step-wise. For example, bodily skills such walking, avoiding obstacles, reaching etc. are usually the first to be re-mastered, while the visual experience itself is still judged to be non-veridical.
Equally, visual experience itself often remains fragmented for some period: objects might appear in the correct place, for instance, yet remain mirrored in orientation. This fragmentation of the skill of seeing led Kohler to interpret his experiments as a means to probe the transformation of “the structure and mutual interweaving of [perceptual] habits” (Kohler, 1964, p. 139), and the process of adaptation as one of rehbituation.

Now, identifying Kohler’s “perceptual habits” (ibid. p. 140) with SMCs, and rehabilitation with the process of equilibration, we can derive some further implications for how a coherent bundle of habits is adaptively maintained. For example, new attempts at equilibration always depart from what is already known, i.e., from existing SM coordinations. During rehabilitation this results in a failure to grasp objects or move in the intended direction, since there is now a violation of the conditions for assimilation that involves vision, proprioception and intended movement. Kohler reports that only repeated and essentially random reach attempts eventually lead to the gradual adjustment of movements and the co-occurring visual experience, indicating that some level of randomness is required in the exploration of new SM coordinations. Secondly, adaptations always seem to be task-specific, i.e., progressive trials are guided towards achieving closure of some kind (e.g., reaching the object), and equilibration in one task (re-learning of left-right symmetry in cycling) does not necessarily transfer to others (left-right symmetry in reading letters). Each perceptual skill is thus mastered by undergoing an equilibration process of its own. But only when all tensions within and between individual SM schemes are resolved, i.e., after appropriate re-equilibration of the network of perceptual habits as a whole, is the world perceived as coherent one again, i.e., correct vision re-established.

Normativity

As with the previous two conditions, behavioural agency requires that we can define an intrinsic normativity that is not to be found at the level of biological self-maintenance, but at the level of networked habits. As mentioned above, the enactive approach proposes that such norms are related to what is required to sustain a stable and internally coherent behavioural organization. That norms governing behavioural agency are (or can be) distinct from metabolic ones, can be seen in the fact that the failure of a behavioural performance (say a misstep in a dance routine) does not necessarily imply a failure for biological adaptation; and conversely, success in cognitive performance (winning a chess match) does not necessarily involve biological success. Norms at the level of the organism act therefore as constraints to sensorimotor agency, but don’t fully determine it. What is good or bad for the behavioural agent, now can also be determined by what creates tension or conflict in the sensorimotor repertoire. And since the kind of engagements a sufficiently complex behavioural agent is capable of is practically unlimited, this mode of agency equally encompasses a much wider range of normative dimensions (not just, say, efficiency, effectiveness, robustness, adequacy etc., but in the case of social agents also aesthetic and moral norms, for example).

An example of how adaptation can occur in the absence of both internal signals related to metabolic needs and externally provided performance-based rewards, is the work by Kostrubie et al. (2012), who use dynamical systems analysis to describe different subjects’ strategies in learning a new sensorimotor skill (in the context of finger-tapping). In agreement with our account of sensorimotor equilibration, the authors find that the routes of learning, i.e., the dynamical adaptations involved, depend on the relevant SM coordinations each individual learner brings to the task. The observed adaptations are
either small incremental modifications of an existing SM strategy or otherwise abrupt bifurcations that qualitatively change the underlying SM repertoire and create novel forms of SM strategy. More importantly, regarding the norms guiding the observed adaptations, the authors show that it is the stability of the desired coordinations, rather than detected errors in performance, which serves to guide sensorimotor learning. While the specific neural mechanism underlying this form of adaptation remains to be elucidated in detail, these empirical findings lend support to the idea of adaptation with respect to intrinsic norms that are not directly derived from the agent’s metabolic requirements.

Levels of agency
Having established that there is a difference between biological or organismic agency and sensorimotor agency, we may sketch some of the relation between these different forms. This is in fact an issue that still deserves more theoretical elaboration so we cannot go further than some broad considerations.

The first thing to note is that there are not simply just two levels of agency. What we have broadly described as organismic agency, based mainly on the adaptive, self-sustaining and precarious identity of organizationally closed metabolic processes, is in itself already constituted by more than one level of autonomy. This has long been recognized by proponents of the organizational approach to the study of living processes. Varela (1979), for instance, proposes apart from autopoietic closure, the autonomy of the immune system, and of the nervous system, as well as other self-sustaining loops such as the relation between tissue cell activity and the extracellular matrix (Varela and Frenk, 1987). These levels of autonomy do not immediately translate into forms of agency, of course. Only one of the three requirements would seem to be warranted, that of identity. But some of them may be shown to fulfill the requirements of asymmetry and normativity too (we could speculate for instance that self-affirming immune networks may be a potential candidate, see e.g. Varela et al., 1988; Stewart and Varela, 1991; Varela and Coutinho, 1991).

Similarly, as we have seen, sensorimotor agency involves habits as self-sustaining sensorimotor strategies which are themselves equilibrated in higher order organizations. At this stage we couldn’t say that such higher level bundles of habits are necessarily unitary. On the contrary, it may well be the case that we enact different forms of sensorimotor agency for instance in different social situations (what we normally describe as roles) and that at the psychological level these are seen as different, sometimes conflicting aspects of a personality.

In short, there are many forms of organismic and sensorimotor agency. What can we say about their relation? The general answer is that there isn’t a general answer, and that the relation between forms of agency depends on the specific relations between the processes involved.

We can make a few broad observations, though. It would seem that in some sense, organismic identity is more fundamental that sensorimotor identity, at least as presented to us empirically by every example of agency we know of. Thus, organic norms play the role of constraints on sensorimotor norms, because it is the organic process of self-construction which (among others) enables the existence of the sensorimotor agent. This enabling, as we have already noted, is not the same a full determination. There are many ways in which organic constraints can be met, but only some of the seem to be preferred, a fact already noted by neurophysiologist Kurt Goldstein in his classical study of abstract and concrete
behaviour (Goldstein, 1934/2000). There are therefore some degrees of freedom within which sensorimotor agency establishes its own normativity. These sensorimotor norms can even be in tension with organic norms (eating junk food, smoking, and other habits that put organic health at risk). This tension may be sustained in time. What cannot happen is that sensorimotor agency breaks the ultimate barrier of organic norms, ie., the organism’s viability. At that point both forms of agency cease to exist.

The above describes one “arc” of the relation between organic and sensorimotor agency. But the description wouldn't be complete if we left it at that. In many ways, once sensorimotor agency is at work, it offers the organic body many possibilities to fulfill its viability that may not have been available to it before. Sensorimotor agents explore their environments and modify them, changing in so doing the environments that affect their organic selves. And in turn this can lead to the organic identity also changing in accordance to new possibilities afforded by sensorimotor agency. At some point in this history of mutual changes, the organic level may not only be “helped” by the sensorimotor level, but actually become dependent on it. Such is the case with practically all animal life. The organismic energy budgeting of animals (different from that of plant life) allows the development of nervous systems that control rapid movement that permit mobile forms of sensorimotor agency where the whole body is able to uproot itself in controlled manners. This in turn affords a variety of possible sources of nutrition. But this “gain” in opportunity is at the “cost” of organismic dependence on sensorimotor success because the energy levels require to sustain mobile agency cannot be achieved in any other way. We see then that while it is clear that organismic agency enables sensorimotor agency, the historical/evolutionary development between the two forms can reach situations where sensorimotor agency in turns also enables organismic agency.

The relationship between organismic and sensorimotor agency is not exhausted by the mutual constraining and enabling described above. In their call for a radically embodied approach to the neuroscience of consciousness, Thompson and Varela (2001) highlight the fact that the brain is not only the mediator of an agent’s sensorimotor embodiment, but also the point where behaviour is organismically regulated. The organismic body and the sensorimotor self in fact engage in direct and mutual interactions at various levels, mediated by links between the autonomic nervous system and the limbic system via the hypothalamus, as well as through contributions of the endocrine and immune system. Emotional states and homeodynamic metabolic processes thus regulate not only specific functions such as sleep, wakefulness or arousal. As Panksepp (1998) notes, the interaction between the neurodynamics of basic emotional circuits and neural schemas related to bodily action plans might also be the origin of affective feelings of pleasure, anger, desire and so on. Finally, organismic regulation may even underlie a “basic self-affection or a core consciousness of one’s bodily self-hood. Thus, processes of life and processes of mind are inseparably linked. Every conscious state is rooted in the homeodynamic regulation between brain and body, and, in a sense, integrates the present state of the organism as a whole” (Fuchs, 2009).

4.2.3 Neuroscience
So far we have concerned ourselves only with self-maintaining sensorimotor structures at the behavioural level, without considering what is required on the brain side to sustain a bundle of habits. While at this stage this is an area of much speculation, several streams of neuroscientific investigation
seem to promise at least compatibility with the idea of behavioural agency as laid out so far. Here we mention a few cases of relevant work.

Firstly, the “bundle of habitual SMCs” hypothesis would predict that there are neurodynamic patterns (as opposed to neuroanatomical structures) underlying the formation of stable sensorimotor coordinations, which are organized into an interconnected network organization. Since SMCs are task-specific, rather than, say, associated with a particular modality, one would expect this to involve the dynamic synchronization of neural ensembles distributed over large parts of the brain. In other words, one would expect the transient integration of local neural clusters into a large-scale synchronized network, which depends on the task, the anatomical disposition, endogenous brain activity and sensorimotor context. The bundle of habits approach therefore is more congruent with neuroscientific theories that emphasise the synchronising dynamics of highly interconnected brain regions (Buzsáki, 2006; Fries 2009; Siegel et al. 2012; Engel et al. 2013), rather than their anatomical or functional modularity. Among these are theories concerned with meaning and intentionality in the dynamic regimes of large scale chaotic brain activity (Freeman, 2000; Tsuda, 2001); approaches to conscious awareness that emphasise the integration (coherence) of functionally differentiated neural ensembles (Edelman & Tononi, 2000); the dynamic core hypothesis (Varela, 1995; 2001); or dynamic cell assemblies (Tsukada, 1996), to name just a few.

Regarding the development of habitual SMCs we have earlier pointed to the activity-driven self-organization of neural circuits (such as retinotopic maps or spinal reflexes) as possible precursors (Kirkby et al., 2013; Marquez et al. 2013). While research in this area is currently focused on the earliest adaptive wiring driven by spontaneous activity (retinal waves, SNA in the developing motor system, see e.g. Kirby et al., 2013), it would be interesting to investigate whether such activity-dependent circuit tuning continues also when the same circuits receive self-produced reafferent signals, i.e. when acting as closed loops (e.g. when spinal circuits receive muscle-related afferent feedback). It is easy to see how this could lead to a kind of habit formation, whereby initially spontaneous activity (motor babbling), along with the constraints of body morphology, produces correlated sensory activity, which in turn feeds back into and helps forming the original circuit that drives the body’s motion. Such self-organized modes of behaviour can in fact be observed in simulations closing the loop between adaptive cortical maps and body-environment dynamics (Kuniyoshi and Sangawa, 2006).

5. Higher forms of agency

Often, the question of agency appears loaded with what we will call “higher order” requirements or definitional constraints that are applicable only to certain aspects of human (individual or collective) behaviour. Depending on the domain of discussion, higher levels of agency might require specific forms of consciousness, reflexive judgement and interpretation, conditions for responsibility attribution, or specific social or inter-subjective bootstrapping of forms of action or forms of justifying or structuring behaviour. Questions of volition, responsibility or free-will are often associated with these forms of agency.
In a sense, it is possible to conceive higher ‘levels’ of agency as a layered superposition of autonomy or self-reflexiveness. One can distinguish self-generated movement from externally imposed movement. Among self-generated movements it is possible to distinguish intentional movement from that produced by a reflex. A movement can be intentional yet not freely voluntary (if, e.g., forced by an external threat). Even if freely voluntary the movement might not be subject to moral responsibility by lack of contextual information. And even if the agent is morally responsible it might fall short of being virtuously motivated and thus proper of an authentic agent (in the sense of authenticity derived from a coherent personal history endowed with care for discipline and cultivated motivation).

There is no room here to develop at length a detailed analysis of the uses and requirements for all distinguishable layers or levels of agency, but we shall sketch some of the most relevant contributions and how they relate to SMCT. In particular we briefly touch upon two dimensions: the subjective “sense of agency” and the collective or intersubjective dimension of agency.

5.1 Sense of agency

**Multiple aspects in the sense of agency**

People generally have an intuition whether or not they are the author of an action. At closer inspection there are in fact several levels and modes at and in which one can be aware of one’s movements or actions. Gallagher (2000), for example, distinguishes two forms of minimal self-awareness: sense of ownership (SoO)—the sense that it is my body that is moving—and sense of agency (SoA)—the sense that I am the initiator or source of the action. Their difference can be appreciated easily in the case of involuntary movements, such as triggered reflexes, or when being pushed, where I can have the sense that it is me that is moving, while knowing that the responsibility for it lies with somebody (or something) else. Pathologies and phenomena such as delusions of control, auditory hallucinations or thought insertion also illustrate possible disassociations between different types of self-awareness (see e.g. Stephens and Graham, 1994; Frith, 2005).

Along another dimension one can further distinguish between a pre-reflective, non-conceptual self-awareness, and a more interpretative kind. The former Gallagher (2000) associates with the fact that when I perceive objects or movements in the environment, I also gain information about myself, e.g. in the form of a perspective that is unique to me. On an even more basic level, my first-order experience also already includes a basic feeling that it is I who is living through that experience. While this is more or less an instantaneous sense of self, the suggested role of the narrative self is to weave together a coherent story of ourselves over time. I.e. it is presumed to construct the sense of a unified and continuous self by reflectively explaining our actions to ourselves in terms of our beliefs and desires.

It seems logical to posit then that both SoO as well as SoA may come in various forms of explicitness, ranging from immediate sensorimotor situatedness, to pure rational reasoning about one’s actions. Gallagher (2007) already hinted at the possibility that SoA could come in (at least) two flavours: a pre-reflective, first-order experience that is linked to the intentional aspects of an action (task, goal etc.), and a second-order reflective attribution (allowing us to retrospectively explain our own behaviour). Each of the two is revealed by different pathologies and could therefore result from different mechanisms. Gallagher suggests, for example, that the perceived lack of concordance between one’s
intention and the perceived (distal) effects of action may explain the loss of the motor-intentional SoA, while problems with introspective higher-order cognition (as in advanced schizophrenia) may result in a loss of reflective SoA. Other authors (e.g. Synofzik, 2007) make similar distinctions, such as between a non-conceptual feeling of agency (FoA) and the interpretative judgement of agency (JoA).

Cognitive models
A common cognitive model for the sense of agency is based on the concept of efference copy, introduced in the early 19th century (Steinbuch, 1811/2012). This concept refers to the idea that appropriately transformed copies of motor commands can be used to modulate afferent signals, in order, for example, to cancel out sensory signals resulting from self-generated movements. An updated version postulates detailed internal models that “predict” the sensory consequences (proximal and distal) of an action, and which can serve in comparisons with either actual sensory feedback or the desired state of the agent. For example, Frith (2005) proposes that a match between predicted and desired state underlies a sense of being in control, while sensory events can be self-attributed (recognised as re-afference) via comparison between predicted and actual state. In this scheme, if the actual state is congruent with predictions, sensory events are attributed to one’s own agency; if sensory feedback is incongruent with sensory predictions, external causation of the sensory stimuli can be inferred. According to this view, the pre-reflective feeling of agency is a diffuse sense of the coherent flow of predictions and sensory feedback, which occurs on the subpersonal level and is phenomenological recessive, in the sense that one normally becomes aware only of disruptions of the process, i.e. when one’s anticipations are not met (Synofzik, 2007). According to Christoff et al. (2011) reafference mechanisms thus implicitly specify a perspective and the self as subject and agent of that perspective. According to the latter authors, this constitutes a basic self-experience that is independent of the content represented in the underlying computational architecture.

There is no consensus yet in the cognitive literature on how the pre-conceptual feeling of agency is supposed to relate to the interpretative judgement of agency. Synofzik et al. (2007), for example, propose that representations formed at the comparator level can be further processed using conceptual capacities and attitudes (e.g. beliefs, desires) to form an attribution of agency (judgement of agency). The extent to which the feeling and judgement of agency, respectively, contribute to the overall SoA, according to the authors, depends on the context and task requirements. For instance, if a mismatch is detected at the subpersonal level, interpretative processes may be triggered that use conceptual facilities to identify the best explanation, with the result of either concluding that I am the agent of an action, despite the low-level mismatch, or that somebody or something else is. Other research, in contrast, provides evidence for a more integrated sense of agency. Weiss et al. (2014), correlating a low-level indicator for the feeling of agency and verbal reports reflecting the judgement of agency in an experiment manipulating the perceived congruence between movements and their visual consequences, conclude that “agency-related judgements are not purely inferentially mediated reconstructions [...], but could be directly based on a ‘readout’ of information within the sensorimotor system” (p. 90).

Critique of cognitive models
Despite its appeal, in providing a simple computational mechanism underlying sensorimotor self-awareness, the comparator model is most likely both insufficient and unnecessary to explain the feeling as well as judgement of agency (Synofzik et al., 2007). Firstly, subjects can attribute the same
comparator mismatch in some cases to themselves and in others to the world. In other words, any potential mismatch itself has to be registered and appropriately categorized by another process different from the comparator model. Also, in order to learn the required internal models, i.e. to learn the effect of its own movements, the system already has to somehow know which of its movements are caused by itself and which are not (i.e. before precise comparisons can be made). Moreover, based on cases of pathological loss of SoA, as well as neuroanatomical lesions in areas supposed to be involved in the comparator, Synofzik (2007) reaches the conclusion that a much less specific congruence between efferent and afferent signals in general (e.g. between an action intention and a distal sensory effect), alone or in combination with certain intermodal congruences, are sufficient to explain the SoA. This idea is supported by cases of FoA in patients with phantom limbs reported by Ramachandran et al. (1996) (i.e. in the absence of proprioceptive feedback and likely also efference copy). Similarly, in the “helping hands” pantomime task, passive subjects experience high degrees of agency for movements that are in fact performed by another agent, when only the other agent’s hands appears in the place where the subject’s hands would normally appear (Wegner and Sparrow, 2004), i.e. when there is most plausibly no efference copy and no comparator involved. Synofzik et al. conclude that it is not a specific, unique and accurate prediction scheme underlying different forms of SoA. Rather, all kinds of action-related perceptual and motor information, like efference copies, sensory feedback modalities and their congruence, is combined in a multifactorial weighting process at different levels of cognitive processing, where the importance of the different authorship cues may vary with task, context and person. Thus, in the absence of reliable proprioceptive cues—as in deafferented patients, for example—a person may be more dependent on the congruency between intention and visual action consequences to experience agency, while in other cases a feeling of agency may be related more to congruence between different sensory modalities.

Although some researchers may claim that SoA explained by reference to the comparator model is somehow an intrinsic property of action unfolding itself (Synofzik, 2007), because it arises from one of the processing stages in the ongoing computational control of movement, this is not so. The comparator model can only suggest a particular (neural) mechanism that may or may not be involved in SoA, in which certain variables correlate with the match between predictions and feedback. It does not, however, explain who consumes these correlations, or how it leads to any kind of experience at all.

From a sensorimotor perspective, authors such as Synofzik et al. (2008), or Christoff et al. (2011), correctly start with the general observation that the co-occurrence of actions and their typical sensory consequences (i.e. SMCs), as well as the registration of non-typical consequences, is a necessary precondition for SoA, as it enables a momentary and implicit distinction between self and environment. But current developments of this idea fall short in several conceptual ways. Christoff et al., for example, remain dedicated to simplistic computational metaphors (insufficient ones as mentioned above), and do not provide an account of how the experience of agency can derive from comparator-like computations. Synofzik et al., on the other hand, highlight the heterogeneous nature of SoA and its underlying mechanisms, but then rely on a representational account for explaining SoA based on notions such as intentions, beliefs etc. (themselves in need of explanation) that is incompatible with both the sensorimotor and the enactive approach. In addition, both accounts suffer from an implicit dualism that is never addressed. Both, more or less explicitly, assume a distinction between the brain-body system on the one hand, which produces movement and implements efferent-afferent computations,
A sensorimotor response

We propose, that such problems can potentially be avoided by extending the sensorimotor approach of perceptual experience to the particular experience of oneself as an agent. The main idea is this: just as perceptual experience, such as seeing an object, does not derive from internal representations in the brain, but is constituted by the rules governing our active perceptual exploration of the world, so is the sense of agency not something that needs to be derived from internal representations of one’s own actions, but rather is a particular mode of acting in the world. According to the SM approach, perceptual experience is a relation between agent and environment, not a property nor a part of one or the other. Sensory modalities, for example, differ, because there are different sensorimotor regularities involved in, say, seeing and hearing. Our claim is that the sense of agency is simply another type, or another dimension, of regularities present in our interactions with the world.

Before attempting to provide a more detailed explanation, we note that this idea alone would not be sufficient to address all problems mentioned above. The SM approach as such only solves the “easy” problem of perceptual experience. It may explain the origin of differences between sensory modalities or instances of perceptual experiences. But as Thompson (2006) has already noted, it does not address the Hard Problem (e.g. Chalmers, 1996) of why we have any experiences at all, who the system is having these experiences, or knowledge of SMCs, or in fact a (sensorimotor) environment. In the same way that O’Regan and Noë’s self-guiding missile (2001), despite their claim to the contrary, does not have genuine mastery of the skill of airplane-tracking, and hence no genuine perceptual experience of “seeing” airplanes (because it has no sensorimotor agency to start with, i.e. there is no “it” that could have any such thing), the sensorimotor account alone is insufficient to describe the kinds of systems that can experience themselves as agents. As Thompson concludes “a complete account of perceptual experience requires an account of non-intentional (intransitive, non-object directed), pre-reflective bodily self-consciousness. Although the dynamic sensorimotor approach has made significant progress in accounting for transitive perceptual bodily self-consciousness, further work needs to be done to address bodily self-consciousness” (Thompson, 2006, p. 14).

Some claim that for the SM approach the Hard Problem is not an issue. Hutto and Myin (2013), for example, rightly claim that “the phenomenological character of experiences must, ultimately, be understood by appealing to interactions between experiencers and aspects of their environment” (p. 176-77). By defining phenomenological experience, such as of oneself as an agent, exactly as a relation between the acting subject and its environment, they avoid having to explain the otherwise mysterious transition from mere physical happenings to experiential phenomena. But, this only reframes the Hard Problem as the question of what an experiencer is in the first place. We suggest that the enactive approach may provide a way to resolve this issue.

Enactive sense of agency
Let us summarize what we require of an explanation for the sense of agency from an enactive sensorimotor perspective.

Firstly, as mentioned above, the experience of agency, just like any other perceptual experience in the SM approach, needs to be constituted by structures present in our active perceptual exploration of the world, i.e. by properties or modes of the relation between agent and environment. This is already compatible with the main idea of enaction, according to which “a cognitive being’s world is not a pre-specified, external realm, represented internally by its brain, but a relational domain enacted or brought forth by that being’s autonomous agency and mode of coupling with the environment” (Thompson, 2006).

Secondly, the basic feeling of agency is a \textit{intransitive} experience, it is not an experience of something, but it is a particular way of experiencing (or acting in the world). It is an experience of “how” and not an experience of “what”. In SM terms, it hence cannot be related to the specific SM regularities associated with particular events or objects, but must be captured by more general properties of our coping with the world. It is this difference then, between specific object-related SMCs, and the more general modes of engagement, that explains why we experience our own agency differently from specific properties of our bodies or events and entities in the environment.

Thirdly, psychological and neurological studies seem to suggest that some form of congruence between different motor and sensory streams seems to be involved in the feeling of agency. Hence, whatever detailed form the modes of SM interaction underlying SoA takes, it should be characterised by the possibility of being more or less congruent, and by the possibility of breakdowns in this congruence.

Fourthly, studies also show that the sense of agency is in fact not a single sense (like seeing e.g.) but a heterogeneous collection of different ways of feeling in control that depends on context, the task, the person etc. It follows that if we want to explain SoA as a more or less congruent mode of SM interaction, then the way this congruence is established and disrupted must be dependent on factors such as the kind of interactions engaged in, the subject’s personal history etc.

Fifthly, the sense of agency seems to be consciously recessive, in the sense that under normal conditions we are primarily aware of \textit{what} we are doing, rather than the fact that it is \textit{we} who are doing it (see e.g. Gallagher, 2012). And often it is only when our intentions are thwarted or our SM engagements unexpectedly disrupted, that we become aware of the loss of agency.

Lastly, for the various modes of SM engagement to constitute genuine experiences, there must be a subject involved in these engagements, i.e. a true subject as understood enactively, invested in interactions with its own intrinsic norms, a subject whose stability can be challenged by disruptions of the congruence in its mode of interactions with the environment.

We suggest here then, that sensorimotor agency, or mental life, as described in section 4.2, may be a candidate approach that could fulfill all the above requirements. Specifically, we propose that the sense of agency is constituted by modes of equilibration in a network of mutually sustaining SM organizations. To further motivate this idea, consider the following parallel. The sense of agency is defined by many as a kind of monitoring, i.e. diffuse awareness of the effect of one’s actions on the
environment (as distinct from those effects or changes not due to one’s own actions). This is analogous to the idea of adaptivity in minimal agency, i.e. an agent’s monitoring and adaptive regulation of the consequences of its activities on its own viability. In the context of sensorimotor agency, we have developed Piaget’s theory of equilibration as a possible form of adaptivity (Di Paolo et al. 2014). Our suggestion hence is that the sense of agency derives from the modes in which a sensorimotor agent monitors and adaptively regulates its interactions such as to stabilise a precarious network of SM organizations.

Remember (from section 4.2; or Di Paolo et al. 2014) that equilibration is the adaptive regulation of sensorimotor organizations—i.e. closed sets of sensorimotor coordinations deployed to achieve a particular goal—whenever a given organization fails to assimilate an environmental situation. Assimilation here is defined by a) the stability condition, which expresses the coherence of agent internal and environmental states mutually stabilising each other such as to form a dynamically stable sensorimotor coordination; and b) the transition condition, which captures the requirement that a stable SM coordination naturally tends to lead to the enactment of the next SM pattern in the organization. An assimilation failure could hence result from either i) a violation of the stability condition, i.e. when something in the relation between environmental variables and the enacted sensorimotor coordination has failed where in the past it used to work (“obstacle” in Piagetian terminology); or ii) a violation of the transition condition, i.e. when something is manifestly unknown about the world since the presumed “right” handling of the situation does not lead “as expected” to the next coordination in the sensorimotor organization (“lacunae”). If either violation occurs, adaptive mechanisms are triggered that aim to re-equilibrate the network of SM organizations. This can occur at three different levels (Di Paolo et al. 2014):

1. Equilibrations that re-establish stability between coordination processes and environmental aspects
2. Those due to the reciprocal accommodation and assimilation between sensorimotor schemes, i.e. reducing tensions between incompatible SM schemes
3. Equilibrations that result from tensions between a particular scheme and the system’s totality, involving a hierarchical dimension of relationships among schemes or subsystems.

Our enactive hypothesis states that the different senses of agency may be related to the different levels of assimilation and accommodation. For example, the diffuse and attentively recessive feeling of agency could be considered the experiential consequence of a network of SM organizations successfully assimilating the current environmental context. In other words, the absence of any perturbations to my dynamic equilibrium, is the feeling that “everything is going according to (my) plan”. Since SM schemes are always enacted by the sensorimotor agent owning it, the “first-person giveness” is already implicitly given. In other words, whether or not I am the agent of my actions, at this level of equilibration, is not a question of reflection, since it always is and can only be I who enacts a SM scheme. Though phenomenologically recessive, we do become more consciously aware of situations where our stability is challenged and an assimilation may fail (i.e. in the case of lacunae or obstacles). The basic feeling of agency is thus best described perhaps as an absence. It is the feeling that my agency is not challenged.
Another level of agency awareness can be associated with active equilibration of SM schemes. Though failure of assimilation may in the first instance raise awareness that I am not in total control of my interactions, a second level of awareness then may arise from the effort I have to exert in order to maximise equilibration, i.e. to counter perturbations and ensure a “smooth” interaction. At this level of active adaptive regulation, senses of agency could be further differentiated according to the type of condition that is being challenged (lacunae or obstacles), or the level of equilibration (between SM schemes and environment, or between different SM schemes).

A prediction of this view would be that if we were to perform a very repetitive task (such as a long copy & paste sequence), our sense of agency should be greatest at the beginning, since we have to deploy more regulatory resources to establish a stable interaction to begin with. But as we repeat the movements, and as long as we do not suffer from fatigue, they would feel more and more automatic. Once the interaction is established, and not challenged by unexpected events, the initial strong sense of agency is thus replaced by a less conscious feeling of being in control (and though the actions may feel “automatic”, it never feels as if somebody else is in control). An analogous explanation may be given for Gallagher’s distinction between the sense of being the initiator of an action (associated with intentions and pre-motor activity), and the sense of being in control (ongoing motor-sensory matching). In terms of equilibration, the initiation of a new SM scheme is already an adaptive regulation, i.e. a response to a new external or internal challenge to stability. But once a new SM scheme has been engaged, it may, if indeed appropriately dealing with the environmental challenge, unfold without further effort.

This view addresses all of our requirements stated above. It solves the problem of who is experiencing by positing that there already exists a well-defined subject that is doing the sensing of its own agency, namely a sensorimotor agent constituted by a self-sustaining network of sensorimotor organizations. It accounts for the different levels of awareness by distinguishing successfully ongoing and already purposeful assimilation, from the effort required for the active regulation of sensorimotor interactions. It also accounts for the heterogeneity of breakdowns in the sense of agency associated with various pathologies, which reveal that the particular sensorimotor “congruencies” required depend on the task, context and person. Under the hypothesis of equilibration, the basic building blocks of a sensorimotor agent are individual task-specifying sensorimotor schemes. These, naturally, vary in terms of the sensory modalities involved, the balance between amounts of internal and external dynamic “processing” etc. It thus follows directly that the sense of agency for different kinds of interactions should depend to varying degree on modalities, participation of specific internal processes, properties of the task and so on. Lastly, the sense of agency conceived in terms of equilibration, being constituted by different modes of active regulation of the sensorimotor network, is an intransitive experience and relational in nature, as would be required by any enactive sensorimotor approach to SoA.

The modes-of-equilibration approach may thus serve as first steps in the direction of a naturalized account of the sense of agency as understood enactively. We believe it provides a plausible mechanistic underpinning for philosophical accounts of enactive agency such as this by Poortier (2013):

“We either establish a connection or we fail to do so when interacting with our environment. When we succeed, we feel a sense of coherence and continuity which can be identified as our sense of self. This
feeling does not have a qualitative component or content. One might even say that we do not specifically feel this continuity, just like we do not sense Heidegger’s hammer when we are successfully hammering a nail into the wall. And whenever this connection between us and the world is disrupted, we feel a sense of estrangement and confusion. Now the feeling of continuity, which is almost transparent when everything runs smoothly, comes to the foreground. Whenever this connection is disrupted, we feel a sense of estrangement and confusion. Now the feeling of continuity, which is almost transparent when everything runs smoothly, comes to the foreground. We will try to solve this confusion by moving around and re-establishing a meaningful relation between what we do and what we perceive. This is the point when we can re-insert the notion of intentionally. Our fundamental intentional directedness is always aimed at establishing and retaining this connection. When this connection is in place, our sense of self emerges. In the small interval in between connections, this intentionality, precisely in the sense of its failing, comes to the foreground.

Our bodily sense of self is not something we feel inside prior to acting. Rather, it emerges and re-emerges whenever a meaningful connection between our moving body and the world is made. We are mostly aware of our bodily selves in a negative way, when our matching activity falls short. It is then, in this interval of estrangement, that we realize that our intentional directedness consists in the way we actively engage our moving bodies with the world around us, and that only when we succeed in this connective activity, a continuous sense of self appears”.

By providing operational definitions for notions such as “coherence” or “continuity” (the assimilation conditions), disruptions, estrangement or confusion (lacunae and obstacles), “establishment of meaningful relations” (equilibration of task-specific sensorimotor organizations), “intentional directedness” (the normativity associated with goal-oriented SM organizations), or “we” and “our environment” (sensorimotor agents and their environments), the equilibration hypothesis may serve as a useful bridge between phenomenological-philosophical observations of agency experience on the one hand, and neuroscientific explanations on the other. How well it fulfills that role, however, will only be determined once the account is worked out in more detail, to the extent that specific predictions can be made and tested.

5.2 Social agency
The paradigmatic examples of SMCT concern interactions with passive environments, as when visually perceiving and reaching for an object. Such objects, moreover, tend to be exemplified as fully external (not subject to their own dynamics in response to the agent’s moves) and neutral (perception is explained in the absence of the agent’s own interests, goals and norms with respect to the object in question). The environment becomes more dynamic when the object the agent sees and reaches for also moves. In such a situation the target is dynamic and the agent adapts its reaching movements accordingly. The environment becomes even more dynamic when objects are responsive or when other agents are present, who move according to their own drives. Actions of others, moreover, cannot be taken as neutral for the agent as a default case. Are our foregoing considerations about sensorimotor agency in any way modified in such situations? We sometimes speak of situations of social agency. But is social agency in any way different from individual agency?

There is an important literature in psychology on joint action and joint attention which is in many ways relevant to this question (see e.g., Knoblich, Butterfill and Sebanz, 2011 and Racine & Carpendale, 2007, for recent reviews). In philosophy of mind we also find relevant input from research on
collective intentionality investigating shared intentions in terms of actions and commitments and the ontological status of collectives (e.g., Bratman, 1992, Searle, 1990, Tuomela 2013). In this short comments, we will limit ourselves to exploring the question of social agency within the framework presented above.

When the environment of an agent includes another agent (or several other agents), its agency may become significantly modulated not just by the presence of the other social agents, but also by their actions. The maintenance of the agent’s identity, at the biological level and at the sensorimotor level, can at any moment be disrupted, constrained, enabled or modulated by the others. If this is the case, agency is in effect negotiated between social agents, or even emergent out of the interactions between them. Thus, we can say that social agency is indeed different from individual agency. In which ways?

**Interaction dynamics and interactive autonomy**

In social situations, synergistic elements, i.e. all kinds of coordination (including lack of coordination and miscoordination) between the agents have different effects on individual agency. Such coordinations can occur at various levels: bodily, in terms of movements, relative position, posture, etc. (e.g. Schmidt and O’Brien 1997; Shockley, Santana and Fowler 2003; Schmidt and Richardson 2008), dialogical, e.g, in terms of the timing and turn organization of interaction moves (Sacks, Schegloff and Jefferson 1974; Kendon, 1990), emotional and cognitive, in terms affect, joint perception and joint action, and so on (see e.g.; Tronick, 2005; Oullier et al. 2008; Marsh, Richardson and Schmidt 2009; Laroche, Berardi, Brangier 2014).

Reed et al. (2006) demonstrate that collective activity in situations of social interaction is not simply constituted by the sum of individual acts. They had participants jointly move a heavy crank lying on a table, by moving its handle by hand to an instructed position while seated on opposite sides of the crank and not being able to see each other otherwise. Individual participants performed this task on average more slowly than when they did it in pairs (differences in inertia are compensated between the individual and the social situation, so that the crank should offer the same resistance to movement). In the pair condition, participants sat at the table opposite from each other, with a closed curtain between them. They moved the same crank, which had two handles, together towards an experimentally set target. Even if most participants reported feeling that the partner hindered their movement, pairs performed the task faster than individuals. The recorded force profiles showed that pairs developed a strategy that individuals alone could not have used: while one partner accelerated the crank, the other took care of decelerating it. Overall, this strategy allowed to perform the task faster than individuals could do it and this occurred reliably and without any previous arrangement between the participants, but as a synergistic effect of the interaction dynamics.

At the more complex end of synergistic coordination, we find acts that can only be understood as joint acts, i.e. acts that necessarily involve more than one agent. Consider the act of giving. The individual intention to give something to another (to hand over an object to the other) cannot be fulfilled – the act of giving cannot be completed – until the other takes the object. To put an object into somebody’s hands is not to give an object. To offer an object and wait for a response on the part of an indifferent other agent is also not (yet) an act of giving. The fulfillment of the intention that accompanies the act of giving (its normativity) is literally in the other agent’s hands and in her moves.
These and many other findings lead to one important conclusion: Social agency is not simply the sum of individual agencies, since agency in social situations is modulated by the other agents. The relation between the sensorimotor activity of more than one agent in interaction is not only nonlinear, but often – though not always – intentionally oriented by individual agents towards obtaining particular responses in others.

Moreover, it has been shown that the interaction process between the agents (i.e., the dynamical patterns of their co-modulated relation of coupling) can itself exert an influence on the individuals, and thus modulate individual agency (De Jaegher, Di Paolo, and Gallagher 2010). The interaction process as an ongoing relational process can take on a form of autonomy (see section 3.5). As a closed network of processes it is constituted by patterns of coordinations, breakdowns and recoveries from breakdowns. It can maintain itself, sometimes even against the intentions of the individuals involved, for instance in the common situation of bumping into someone in a narrow corridor, and not being able to pass by each other while each of you unwillingly keeps stepping in front of the other, before breaking the coordination spell and each being able to continue on your way. Situations like this everyday example, where an emergent coordination process takes over and (even if briefly) controls individual intentions and agency, have also been demonstrated empirically. Auvray et al. (2009) have also shown that performing a cognitive task can sometimes be explained by interactive dynamics (see also De Jaegher et al. 2010). In their experiment, participants had to find each other in a virtual environment, in which they were represented by an avatar whose movements could not be distinguished from that of its shadow. Participants reliably picked out the avatar, and not the shadow (even though their objective movements were indistinguishable). Analysis showed that this was due to the interactive coordination dynamics between participants trying to find each other being more stable than individual searching movements with a non-responsive object (Auvray et al. 2009).

In sum, social agency is not simply a more complex form of individual agency but for several reasons striking at the core of the enactive conception of agency, it is a different form of agency altogether, one where individual, asymmetric and normative modulations of the coupling with the world are modified, constrained, in conflict or coordinated, and sometimes even jointly achieved, by the modulations performed by other agents.

**Enactive intersubjectivity: participatory sense-making**

Now, how do these social interaction dynamics connect with individual agency and intentions – with individual cognition? Social agency is best understood from the point of view of the enactive account of social understanding, called *participatory sense-making*. Individual sense-making and agency can be modulated by engaging with other agents, i.e., by the other agents’ autonomous actions, as well as by the autonomy of the interaction processes they engage in. If our embodiment is fundamental in causal and constitutive ways to our agency and to how we understand the world, and the body is fundamentally restless and in need of ongoing relations with the world in order to maintain itself, as seen above, then a body needs to move in and through the world thus making sense of it in order to survive. The world then makes sense to embodied agents in that certain things are relevant to them in ways particular to their survival, or the maintenance of a specific level of autonomy, like sensorimotor autonomy. In this way, we can say that we know, perceive, and act in the world by and through moving
in it (Merleau-Ponty 1945-2012; Johnson 2007; Sheets-Johnstone 2009). On the other hand, it has been amply demonstrated that, in social situations, we extensively coordinate with each other at various levels, from physiological variables, posture, distance, gestures, speech acts, and affect (see e.g. Streeck, Goodwin, LeBaron 2011; Tschacher, Rees and Ramseyer 2014; Konvalinka et al. 2011; Abney et al. 2015). If both these statements are true, then we can conclude that in and through moving together, we understand each other and the world together. Thus, we literally participate in each other’s sense-making activities (De Jaegher and Di Paolo 2007).

Social agency and self-other contingencies
In this framework of participatory sense-making, then, social agency is characterized by what McGann and De Jaegher (2009) call “self-other contingencies”. Action and perception in the social domain are a matter of coordinating the behaviours, emotions, and intentions of the agents involved, in and through the coordination of movement (including utterances). Social skill then is the mastery of these social contingencies. Self-other contingencies are different from SMCs in a number of ways. Social interactions are interactions between subjects, each of whom is maintaining their own autonomy. This affects the asymmetry in how agents interact with the environment. Whereas, in interactions with objects, the agent effects changes in the environment, agents can intend to bring about changes in each other, i.e. to influence each other’s sense-making and autonomous organization. The asymmetry is thus more complex (able to change over time along different dimensions), and the regulation of social interactions is not completely down to either individual (De Jaegher & Froese 2009). Interactions with other social agents are far less predictable than those with (most) objects. Interactions between agents potentially transform not only each agent, but also their relationship itself. Forming and transforming social relationships is done in a participative, shared process structured by each agent’s normativity and their concomitant perspectives on the situation. Such situations are often flexible, but not always, as in the case of rusty relationships with family members, where certain patterns (e.g. fighting) are repeated without the participants wanting to do so (e.g. Granic 2000).

There are different forms of social agency. The most basic form results from the attempt to resolve the primordial tension in participatory sense-making (Cuffari, Di Paolo and De Jaegher 2015). When individuals encounter each other, the most basic tension between them is that between individual normativity and interactive normativity (since the interaction as a process is also autonomous). Even before social agents recognize each other as such, their individual normativities and that of the interaction process that emerges between them can be in tension with each other. Since both the individuals and the interaction process attempt to maintain autonomy, the normativities that exists at both levels can be in conflict with each other. Resolving this tension entails increasingly sophisticated forms of social agency. We can co-regulate interactions with others, by spontaneous social acts, such as when a parent offers something to his infant. The infant is at first not yet capable of accepting the object, but over time, it becomes possible for her to do so. Giving and receiving thus over time become structured as co-defined partial acts, and become a well-established way to regulate social contingencies. Infants soon even play with such interchanges, when they start to coyly pull back objects they hold out before the other person can grab them (Reddy 2001). Increasingly sophisticated ways of regulating social acts develop as these partial acts are themselves used to regulate other partial acts (such as when we use gestures to imply acceptance or rejection, yes or no, contingently on the partial acts of others). This increases the complexity of social agency by introducing its own social normativity (partial acts can
be appropriate or not, sufficient or not, but in addition they can be good or bad, clear or unclear regulators of other social acts). Eventually, the primordial tension takes on new forms which are resolved not simply at the level of local interactions but at the level of a community leading to the capacities of producing and interpreting utterances in regulated dialogical structures. Joint acts, normativity at community level, participation genres, social self-control (the application of social regulatory acts on oneself, as in behaviour directed by inner speech) and language all develop on the basis of this fundamental tension in participatory sense-making between individual and interactional autonomy (Cuffari et al. 2015).

In this context, it is unsurprising that our individual capacities and cognition are influenced and affected by social interaction history and skills. This happens for instance when a box looks lighter if you expect to lift it together with another person (Doerrfeld, Sebanz, and Shiffrar, 2012). Such interactional dynamics also modulate and constrain brain development and functioning across the life-span (e.g., Schilbach et al. 2006, 2013; Catmur et al. 2007). These questions have now begun to be examined in neuroscience using dual scanning techniques which permit to some extent a less constrained development of spontaneous interaction dynamics (e.g., Dumas et al. 2010, Lindenberger, et al. 2009, Konvalinka & Roepstorff, 2012). This “interactive turn” in social neuroscience has led to the proposal that skills acquired in situations of social agency shape and in some cases even constitute the skills that are deployed by a single agent when acting on her own. These are the developmental and the constitutive versions of the Interactive Brain Hypothesis proposed by Di Paolo and De Jaegher (2012).

How do these considerations impact on the current and future development of SMC theory? Gallagher (2009) has pointed out that the SMC theory has a tendency to a kind of “autism” if it doesn't take account of this fact of the social modulation of even individual cognition and perception. That is, Gallagher suggests that SMCT, on its own, cannot explain its own generation, if it cannot explain how we learn SMCs, namely, mainly from others, through processes of participatory sense-making. We could suggest here that understanding social agency is precisely a necessary requirement for understanding how we can ever make sense of the simplest example cases of SMC theory which involve a certain detached and objective attitude towards the object of perception (the tomato in its roundness and redness, as opposed to the tomato in the process of preparing a salad). Such detachment, which is assumed unproblematic in SMC theory, is in fact the result of having been able to experience a multiplicity of sensorimotor intentions towards a same object, many of them different from our own, in situations of social agency. In this way, much of our mastery of SMCs is acquired in deeply social situations, not only through social learning during development, but also through recurrent patterns of interactions throughout a lifetime. Most skills involving new forms of sensorimotor mastery and expertise (from carpentry to wine-tasting) are enabled and sometimes even constituted by participation in social agency.
6. Models

Our discussion of minimal agency has so far been mostly conceptual. What is needed still are specific models that help illustrate and further develop these ideas. We have previously shown that sensorimotor theory can be translated into operational terms and thereby serve as a basis for models that aid in deriving more detailed implications from the theory (Buhrmann et al, 2013; Di Paolo et al. 2014). Similar attempts have been made in the enactive literature. Though it has to be said upfront that there exists so far no model that incorporates all three requirements for minimal agency (individuality, asymmetry and normativity), a variety of work has been produced each spanning at least two of the requirements. These fall broadly into two categories, namely dynamical systems approaches and minimal cellular systems (see Table 1).

<table>
<thead>
<tr>
<th>Dynamical systems approaches in neuro-controlled agents</th>
<th>Minimal cellular systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-trivial self-production</td>
<td>Assumed / abstracted (e.g. bistable metabolic dynamics)</td>
</tr>
<tr>
<td>Asymmetric agent-environment coupling</td>
<td>Mathematically tractable analysis of means for sustaining norms of survival (externally imposed). E.g. models of Ashbyan ultrastability, habit-formation etc.</td>
</tr>
<tr>
<td>Normativity</td>
<td>Norm-following behaviour Quantitative account of how behaviour obeys organism’s intrinsic norms</td>
</tr>
</tbody>
</table>

**Table 1:** Kinds of models of minimal agency in relation to the three enactive requirements for agency.

The dynamical systems approach does not usually aim to model self-construction explicitly. Instead it looks for how the agent's modulation of the environmental coupling can effectively sustain some norms of survival, usually given by an externally imposed artificial selection process (e.g., Izquierdo et al., 2013). The advantage of this line of work is that it is more or less amenable to mathematical analysis. Work along the minimal cellular systems approach, in contrast, starts from a simplified notion of self-construction, e.g., in the form of an abstract model of autocatalytic chemical reactions that stand for the system's metabolism (Egbert & Di Paolo, 2009, Egbert, Barandiaran & Di Paolo, 2010, 2012). The aim of such models is to uncover the origins of intrinsic norms that emerge from the relation between self-constitution (metabolism) and behaviour. Other, more detailed, biochemical models of protocells focus on complex metabolic and compartmental stability and growth, but not often on the cell as behaving (i.e. as an agent). Exceptions are models that look specifically at regulatory events such as the modulation of ion-channels (Ruiz-Mirazo & Mavelli, 2008). These events can be considered to be part of minimal forms of agency.
Models of biological agency and SMCs
In this category, models aim to ground or investigate the normativity guiding sensorimotor interactions in an agent’s “biological” self-production or self-maintenance. The approach is exemplified in a series of simulations of bacterial metabolism-dependent chemotaxis (Egbert et al., 2010; 2011; Barandiaran & Egbert, 2013), in which the authors investigate the link between a precarious autocatalytic metabolism and the chemotactic behaviour towards the compounds sustaining it. This has yielded conceptual clarifications about the notion of self-maintenance (the first agency condition), such as the definition of precarious regions in state-space from which recovery can only happen through the intervention of agential regulation (see Figure 9). Another result is the identification of two aspects of intrinsic normative relations between an agent and its world: norm-establishing and norm-following aspects and behaviours (Barandiaran and Egbert, 2013).

![Figure 9: Regions in viability space (living, dead, viable, precarious, and terminal) in minimal agency model by Barandiaran & Egbert (2013). [A] represents the concentration of metabolites (proxy for cellular size) and [F] the concentration of nutrients. The dead region or state lies at [A]=0, above which the living region appears. Inside the living region three different subregions are distinguished: the viable region (light gray) where the system will remain alive if environmental conditions do not change, the precarious region (medium gray) where the system is still alive but tends toward death unless environmental conditions change, and the terminal region (dark gray) where the system will irreversibly fall into the dead region.](image)

Other work more explicitly attempts to model self-constitution at the level of molecular self-assembly, e.g. by using computational artificial chemistry to simulate the emergence of protocells with boundaries. Egbert & Di Paolo (2009), for instance, have modelled an autopoietic cell to investigate the limitations of behavioural modulation based on autopoietic efficacy when compared to a decoupling of behaviour generating mechanisms from processes of self-construction. Using a different simulation approach, Ruiz-Mirazo & Mavelli (2008) have studied the conditions for the appearance of the simplest forms of...
autonomy in the context of lipid vesicles that enclose an autocatalytic/proto-metabolic reaction network, and the mechanisms by which they could transform external material-energetic resources into their own means and actively regulate their boundary conditions (e.g., osmotic gradients, inflow/outflow of different compounds).

An example of experiments investigating mainly the second agency condition (asymmetry in the interaction with the environment) is a new dynamical model of Ashbyan ultrastability (Izquierdo et al., 2013), which relaxes some of the constraints originally proposed to be guiding the regulation of an agent’s essential variables through the modulation of agent-environment interactions.

Models of sensorimotor autonomous agency

In this category models aim to study behavioural regulation based on norms distinct from “biological” self-maintenance. The usual approach here is to require the stability of otherwise precarious sensorimotor interactions themselves to be that which is preserved as the agential unit. In this spirit, Di Paolo and colleagues have applied the Ashbyan idea of ultrastability to define a neural network model with local synaptic plasticity rules that only activate when the firing rate of post-synaptic neurons moves out of pre-established homeostatic bounds—thereby making the firing rates, rather than organismic states of survival, essential variables in Ashby’s terms. After the agent has learned a task, radical sensorimotor disruptions (such as left-right inversion of the visual field) are introduced. In a pattern similar to the behaviour of human subjects, the agents fail to perform the task for significant periods until the internal breaks in homeostasis that such failures produce eventually find new parametric combinations that allow agents to recover the original function (though the specific behaviours used to achieve the task is naturally different due to the sensorimotor disruption). Removing the disruption shows a similar pattern of re-adaptation. The same principles have been used to investigate the dynamics of spontaneous change in behavioural preference in a mobile agent (Iizuka and Di Paolo, 2007, Di Paolo and Iizuka, 2008). In all cases, agents follow an internal norm: what is good or bad, is good or bad only in terms of its contributions to homeostasis. The limitation here, of course, is that the experimenter arbitrarily decides on the acceptable ranges for essential variables, thus sidestepping the issue of self-constitution. In the latest addition to this body of simulation models (Aguilera et al., 2015), the authors show a network of homeostatic neural oscillators converging, through the learning of a behavioural preference task, on a regime of self-organized criticality as measured (amongst others) by 1/f noise, which is often interpreted as a sign of increased skill (hypothesised to be due to increased coordination between dynamics at all spatial and temporal scales). They show that both homeostatic plasticity as well as strong sensorimotor coupling are necessary, in their case, to observe the effects of self-organized criticality.

Egbert and colleagues explore sensorimotor agency from a slightly different angle, namely by studying directly the emergence of habits in what they call a “deforming sensorimotor medium” (Egbert & Barandiaran, 2014; Egbert and Canamero, 2014). The main idea of this model is that trajectories flowing through sensorimotor space deform that same space such that similar trajectories become more probable. The analogy is that of the paths made on patches of grass by people regularly walking along the same paths and thus inhibiting grass re-growth. It can be visualised e.g. as a vector field whose vectors are rearranged and re-oriented in the direction of trajectories passing by as illustrated in Figure 10 (the actual implementation efficiently represents only the actually visited areas of phase space).
Figure 10: Illustration of itinerant deformable sensorimotor medium (IDSM). First row: nodes (black arrows) are created as a record of previous trajectories (gray) taken through sensorimotor space. If a new trajectory approaches previously visited nodes it is more likely to follow the earlier route. Second and third row: nodes are precarious and degrade over time. But they can also be reinforced if re-visited. Last row: in the right conditions habits may form, where the exercise of a sensorimotor pattern becomes the condition for its own continuity (Figures with permission by M. Egbert from: “Habit Based Regulation of Essential Variables” talk given at ALife NYC).

Typical instances of this model when coupled to an agent and its environment are characterised by repeating, self-maintaining metastable movement patterns (“habits”). These patterns can occur both spontaneously, or as the result of teacher-forcing the system with imposed trajectories during a developmental phase. The particular SM patterns observed depend on the agent’s embodiment and environment and seem to vary in stability depending on how the pattern in the deformable medium
“resonates” with the behaviour in the environment. For instance, those patterns seem to be most stable that form a regular interaction with environmental features. Due to its metastability the system also seems to exhibit switching between “habits” and between periods of exploration and exploitation. Higher-level organizations of habits also seem to occur at times, e.g. the “super-habit” of switching between two “sub-habits” at more or less regular intervals.

Like the dynamic interpretation of equilibration this model treats closed-loop SM coordinations as the fundamental units to be learnt. Interestingly, it exhibits strong novelty-generation, by having SM units “emerge” from an initially undifferentiated medium. Another interesting aspect is that the process of carving out (or memorising) of SM trajectories is precarious, because the medium is constructed such that memory intrinsically fades away and needs to be continuously refreshed to persist. Due to this precariousness, formed habits can be seen as truly self-maintaining. Only through the continued enaction of the habit can it maintain itself. This active self-maintenance might also function as an internal evaluator of environmental perturbations, since these can be intrinsically good or bad for the continued stability of the habit. However, in contrast to the theory of equilibration, there is no possibility of directed adaptation. It is not only meaningful SM coordinations that are fixated in the deformable medium, but any and all metastable patterns; though one could imagine an extension where plasticity would be modulated by reward, with the aim of stabilizing only useful habits.

In conclusion, from the short and non-exhaustive list outlined here it should be clear that all aspects of minimal (enactive) agency can, in principle, be spelled out in sufficient operational detail to be tested in computer models. However, several challenges remain. While protocellular simulations are perhaps closest to covering all three agency requirements, a limitation so far has been that the complexity achieved by such models is not sufficient to produce cognitively interesting behaviours, thus limiting insights derived from such models to issues of self-production and regulation of boundary conditions. From the perspective of SMCT, dynamical systems models should be closer to the required domain of interest, as they usually start at the level of regulation of sensorimotor interactions. What is still missing in this case, however, is a clear illustration of a precarious dynamic entity that sustains itself via environmental exchanges (e.g. a network of mutually enabling SM organizations, as argued in section 4.2). In cases where this is more or less clear, on the other hand, as in the example of habit formation, aspects of self-production and normativity are present, but not the asymmetric, directed regulation of behaviour.

7. Discussion and conclusions
We set out to fill in what is perceived as a gap in sensorimotor theory, when taken as an account of cognition in general, by integrating it with an enactive grounding of subjectivity and normativity in the concept of minimal agency. To this end, after having summarised the three conditions for minimal agency (Barandiaran et al., 2009), we have evaluated two possible domains in which minimal agency may be relevant for sensorimotor theory: the “biological” domain (actions as adaptive regulations required to sustain an agent’s continued existence); and the sensorimotor domain (actions as adaptive regulations required to sustain a precarious network of mutually dependent sensorimotor organizations, decoupled from biological autonomy). To arrive at the latter, we have synthesised our previous work
on the dynamical interpretation of SMCs (Buhrmann et al., 2013), our development of a Piagetian framework for understanding the development of complex ecologies of SMC-based skills (Di Paolo et al., 2014), and minimal agency. We have then argued that such an approach is suited also for a conceptualization of the sense of agency that is compatible with the pragmatic turn at the heart of SMCs (in contrast to incompatible approaches based on internal representations). We have also provided examples of models that address various aspects of minimal agency and show that the concept can be operationalised and tested conceptually.

The account of minimal agency summarised in this report complements the notion of eSMCs already developed in this project. For example, in addition to the modality and object-related SMCs already covered by sensorimotor theory, in the project it has been suggested that sensorimotor correlation structures on larger time scales, covering more distal action effects, and more general properties of sensorimotor interactions, may be related to an agent’s intentional relation to the environment. From the sensorimotor agency perspective developed here this makes sense. As we have suggested above, Piagetian SM schemes, operationalized as closed organizations of simpler SM contingencies, form the basic building blocks of a network of mutually sustaining behaviours. By definition, these involve greater time scales, as well as intentional aspects. As we have also noted, a SM scheme’s closure is directed at the achievement of a particular goal, and ultimately, through its role in the greater network, this goal is related to the continued possibility of existence of the agent as a sensorimotor identity. Intention-related eSMCs as such, in this account, capture the mathematical regularities associated with the enactment of a particular SM scheme. But an account of agency is required to explain the origin of the scheme’s intentional aspects in the first place.

In other words, the agency approach provides one avenue for recovering an enactive notion of intentionality in sensorimotor theory, and for avoiding the representational pitfall left open in its original formulation by O’Regan and Noë (2001). Such a notion of intentionality would not be based on the idea of internal representations (of SMCs) being the “mental causes” of action (whatever that may mean), but rather stipulates intrinsically normative actions (normative because they contribute to the maintenance of a precarious biological or sensorimotor agent) as the primary source of intentionality, from which higher-level, attributed, “semantic” intentionality may or may not derive. Understanding intentionality thus, is simply taking the pragmatic turn seriously. As Brandom notes, “A founding idea of pragmatism is that the most fundamental kind of intentionality (in the sense of directedness towards objects) is the practical involvement with objects exhibited by a sentient creature dealing skillfully with its world” (2008, p. 178). The enactive account of agency summarised here naturalises what is meant by “creature” or “skillful”, and makes it clear that intentionality is not “in the head”, but involves the environment as much as the agent as constitutive parts of intentions in action. Gallagher and Miyahara (2011), further argue for (only) such an account of intentionality being compatible with enactive and extended theories of cognition, and in particular relate how even social normativity originates at the basic sensorimotor level. For example, the direct perception of other’s actions as affordances for further actions on my part, predates, ontogenetically speaking, more abstract, communicative social practices.

In SMCT it has also been proposed that knowledge of intention-related SMCs, when used offline (i.e. simulated), could be used in planning through predictive evaluation of different behavioural
alternatives, and be ultimately responsible for the sense of agency. In particular, it has been suggested that the experience of agency is “directly related to the mastery of action-effect couplings at the level of intention-related eSMCs”. Again, at a general level, this is perfectly congruent with our account of sensorimotor agency, according to which the sense of agency is constituted by different modes of adaptive regulation at the level of SM schemes. In fact, equating mastery with adaptive regulation, and intention-related SMCs with SM schemes, the two statements are identical at this level. However, we believe that the agency approach has the potential to develop this general idea in a way that is compatible with the relational nature of sensorimotor theory, according to which experience does not arise in the brain from representations or simulations, but rather from regularities in the domain of agent-environment interactions (at any scale). While at first it may seem as if such an approach cannot account for cases of covert action or anticipative behaviour, we argue in Deliverable 1.7 on virtual actions (and in Buhrmann and Di Paolo, 2014) that such doubts might be unwarranted.

**Future work, open challenges**

As we have mentioned in section 6, there exists so far no model successfully covering all three requirements for minimal agency, neither in the biological nor in the sensorimotor domain. In particular, there has been no attempt to model Piagetian equilibration of SM schemes as interpreted here. A first step has been taken with Egbert and Barandiaran’s habit model (2014), which, however, does not explicitly allow for the required hierarchical organization of complex SM organizations, nor for goal-directed adaptivity (equilibration). Also, though useful as a conceptual bridge, it lacks biological plausibility. It would hence be desirable to investigate the possibility of networks of SM organizations emerging in, say, neural net-like media. For example, it may be a worthwhile effort to study the possible appearance of dynamically synchronized neural ensembles resonating with specific internal demands and external contexts. If such ensembles could be reliably activated, local learning rules may be able to then tune their specific dynamics such that internal demands are matched with appropriate actions.

Regarding the sense of agency, the idea first sketched in this report of interpreting it in terms of different qualitative aspects of processes of equilibration of sensorimotor organizations needs to be worked out in more detail. In particular, the phenomenology of various pathologies, and existing neuroscientific results, should be matched with specific proposals for the different aspects of the sense of agency (pre-reflective feeling vs. judgement etc.).

**References**


53


