2018-10

Representational fluidity in embodied (artificial) cognition

Windridge, D

http://hdl.handle.net/10026.1/12160

10.1016/j.biosystems.2018.07.007
BioSystems
Elsevier

Representational Fluidity in Embodied (Artificial) Cognition

David Windridge*
d.windridge@mdx.ac.uk

Serge Thill†
serge.thill@plymouth.ac.uk

Abstract

Theories of embodied cognition agree that the body plays some role in human cognition, but disagree on the precise nature of this role. While it is (together with the environment) fundamentally engrained in the so-called 4E (or multi-E) cognition stance, there also exists interpretations wherein the body is merely an input/output interface for cognitive processes that are entirely computational.

In the present paper, we show that even if one takes such a strong computationalist position, the role of the body must be more than an interface to the world. To achieve human cognition, the computational mechanisms of a cognitive agent must be capable not only of appropriate reasoning over a given set of symbolic representations; they must in addition be capable of updating the representational framework itself (leading to the titular representational fluidity). We demonstrate this by considering the necessary properties that an artificial agent with these abilities need to possess.

The core of the argument is that these updates must be falsifiable in the Popperian sense while simultaneously directing representational shifts in a direction that benefits the agent. We show that this is achieved by the progressive, bottom-up symbolic abstraction of low-level sensorimotor connections followed by top-down instantiation of testable perception-action hypotheses.

We then discuss the fundamental limits of this representational updating capacity, concluding that only fully embodied learners exhibiting such a priori perception-action linkages are able to sufficiently ground spontaneously-generated symbolic representations and exhibit the full range of human cognitive capabilities. The present paper therefore has consequences both for the theoretical understanding of human cognition, and for the design of autonomous artificial agents.

1 Introduction

In cognitive science, theories that cognition is, in some sense, embodied, can be traced back to two distinct origins (Chemero, 2009): a reaction to the perceived inadequacies of purely computationalist accounts, and a continuation of eliminitavist/anti-representationalist theories of mind. The latter aims to understand cognitive systems by characterising the dynamics of their behaviour and interactions within the world (often using the language of dynamical systems), and is usually explicit about positing fundamental roles for the body in cognition. The former, meanwhile, aims to characterise the computations taking place, and thereby often

*Department of Computer Science, School of Science and Technology, Middlesex University, The Burroughs, London NW4 4B, UK

†Centre for Robotics and Neural Systems, University of Plymouth, Plymouth PL4 8AA, UK and School of Informatics, University of Skövde, 54128 Skövde, Sweden.

ends up producing an account in which the role of the body is reduced to an interface with the world.

The observation that the precise role or details of the body are often unclear if not explicitly ignored is not new (see, *e.g.* Ziemke, 2003; Wilson, 2002). Robotic implementations of models of cognition contribute to this as they can lead to the false assumption that merely acquiring inputs from cameras and sending outputs to motors are sufficient to provide an embodied model. Such approaches "... reduce the body to a mere sensorimotor interface for internal processes that are still just as computational as they were 30 - 40 years ago" (Ziemke and Thill, 2014, p. 1). The main motivation often stated in robotics research is that such a minimal embodiment is necessary to ground symbols so that they acquire a meaning intrinsic to a cognitive agent as opposed to one that is given by an external observer (as per the symbol grounding problem, see Harnad, 1990). Once that intrinsic meaning has been established, cognition using such symbols can be entirely computational[1]. In the present paper, we argue that reducing the role of the body to such an interface is insufficient, even if it is accepted that human-level cognition can be adequately modelled in a computationalist framework assuming a body as a source and destination of data. This is because the body fundamentally shapes the computationalist framework itself.

As far as theories of embodiment go, we therefore intentionally take a very weak position. The contribution that follows from this is two-fold. First, as already stated, we demonstrate that the role of the body – even given this weak interpretation – must necessarily go beyond that of a sensorimotor interface. It must also go beyond what is required by symbol grounding considerations because it provides and shapes necessary computational mechanisms that cannot be disembodied. We note that this is not an anti-functionalist argument; we merely reject the claim that models that are implemented in a physical agent, but merely use the available sensors and actuators to collect and deliver information for otherwise computational approaches, are in any sense embodied (therein following Ziemke and Thill, 2014, albeit for different reasons).

Second, we demonstrate that if one were to instead approach this topic from a machine cognition angle (see, *e.g.*, efforts in so-called Artificial General Intelligence), then a body will again be necessary to enable the full range of human cognitive abilities. This remains the case even if one otherwise rejects stronger, possibly non-representationalist positions on embodiment (for example, so-called 4E cognition, which emphasises the embedded, embodied, enactive, and extended nature of cognition).

Of particular importance is that our argument does not rely on circumstantial evidence of the involvement of sensorimotor cortices in higher-level cognition. Although such an involvement is often put forward as support for embodied theories (see Chersi et al., 2010, for an example on language processing), it is just as often summarily rejected by critics (*e.g.* Mahon and Caramazza, 2008). We also avoid the traditional symbol grounding (Harnad, 1990) route as a motivation (we do, however, end up with an account in which symbols are also grounded by design). Rather, the argument is simply driven from identifying the computational abilities that human cognition requires, and demonstrating that those necessitate action and perception to be able to generate useful representational frameworks. While our argument therefore uses some terminology from ecological approaches, the focus on computation distinguishes the present argument from those approaches, which link cognition to both the environment and action/perception abilities through the concept of meaning (Gibson, 1979).

We build the argument by first showing that an autonomous adaptive agent needs both the ability to adapt to novel data *and* to update its representational capabilities in relation to that data. We then show that achieving the latter step requires the ability to generate falsifiable hypotheses about novel representational frameworks. This, in turn, requires the ability to act

[1]Note that such a minimal embodiment whereby a computational model is implemented in a robot is precisely the scenario that Searle (1980) considers – and then rejects – as the "robot reply" to the Chinese Room argument.

and perceive in the world (in particular, it requires an a priori notion of action states); thus the action possibilities and perceptive abilities of an agent are shown to be fundamental to shaping the representational frameworks used by the agent.

In the following sections, we begin with considering requirements on the representational frameworks of agents, and how it may be updated. We first presents insights from the literature on biological agents (including humans), and then discuss how one might approach this in an artificial agent.

2 Updating the representational framework

2.1 In biological agents

For biological agents generally speaking, the matter has been considered in epistemological terms in biosemantics, a sub-branch of biophilosophy proposed by Millikan (1987) as an attempt to subsume certain philosophical questions of representation and perception within the purview of biology. This includes, in particular, the contingencies that arise from consistency with respect to natural selection, where organisms are naturally-selected for efficiency of their representative capability in terms of either overall neuronal budget or total energy of processing. Thus, the biological organism's representative capability must, in addition to being maximally or near-maximally efficient, also *be of utility to the organism* in perpetuating its genetic code (*i.e.* it must be consistent with natural selection) if it is to be consistently propagated.

In practical terms, this means that the organism must be able to discriminate those entities (food, predators, mates, and more), that are key to its survival and reproduction Piaget (1970). However, the biological agent will also have simultaneously acquired, by natural selection, an *active capability* that is likewise evolved to maximize the organism's ability to propagate its genetic code; *i.e.* its ability to interact with the environment is adapted to maximize its survival and reproductive capability (a lobster's claws are evolved for opening shells; its eyes provide the appropriate visual capacity to achieve this end). The perceptual and the active capabilities of most organisms have thus evolved in lock-step; the organism perceives (since it must maximize efficiency of representation) only that which is relevant to its survival and reproduction *with respect to its active capacity to achieve these ends.*

While this primarily describes biological entities with a fixed post-natal representational framework Sipper (1995), humans have, to a larger degree than other animals, additionally acquired the capacity to reconfigure their neuronal and perceptual structure *in relation to* the environment in ways that go far beyond (whilst still incorporating) the immediate biological requirements (cf. e.g. Stevens and Neville (2006)). In other words, we have additionally evolved the capability of adapting our representational directly to the world on a life-time scale. Moreover, our perceptual re-configurations can be very abstract. Two principal operative criteria for this adaptive perceptual updating are apparent:

1. The need to obtain a maximally efficient representation of the environment
2. The need to ensure the discriminability of the active capabilities of the agent, as well as key entities related to survival/reproduction/nutrition.

By the "discriminability of the active capabilities" in the latter criterion, we mean the ability to perceive the outcomes of intended actions undertaken by the agent. That is, an intentional action (one initiated by the goal-setting aspect of the agent's cognition) should be susceptible to the sensory determination of its having taken place as intended. In straightforward terms we might say that an intentional action is one that has a specific percept as its success criterion.

2.2 In artificial agents

In an artificial learner, perceptual data generally exists on a manifold with an intrinsic coordinate set. Equally, however, perceptual data can consist of discrete entities within a common class (for example, specific physical cups within the class 'cup'). In first-order logical terms these may be equivalently represented by scoped variables; we shall refer to both as *perceptual parameters*.

An artificial agent that interacts with the physical world will typically need to learn online; this is a standard machine-learning setting in which data (as the above) is presented serially in time (*e.g.* in response to the agent's actions). It is necessarily forward-looking, predicting the label values of data not yet presented to the system, adjusting to any disparity with observed action outcomes. Such a system is thus inherently *adaptive*; although the degree of adaptation will vary from agent to agent (sophisticated variants may incorporate notions such as transfer learning (Pan and Yang, 2010; Taylor and Stone, 2009), anomaly detection (Chandola et al., 2009), and active learning (Settles, 2010; Koltchinskii, 2010)).

Interestingly, however – and despite this tendency towards increasing adaptivity – the majority of existing approaches typically assume an underlying consistency in the *representational characteristics* of the data; the data-stream presented to an agent is generally delineated in terms of a fixed set of classes, or a fixed set of features (for example, spatial interest points or texture-descriptors). Techniques exist that partially address these limitations, such as in learners incorporating Dirichlet processes or similar to spawn novel states in relation to the requirements of the data (Hoffman et al., 2010), which are thus capable of expanding their representational characteristics to a limited extent. However, such a learner would typically not be capable of spontaneously carrying out as fundamental a representational shift as that involved in the transition from, say, a low-level feature-based representation of the world (delineated e.g. in terms of CCD camera pixels) to an *object-based* representation of the world (delineated in terms of indexed entities with associated position, orientation, equivalency classes etc), unless a prior capacity for object representation had been incorporated into it.

Taking the notion of *autonomously adaptive* agency to its conceptual limit would thus require that both the *representational* capabilities of the learner as well its capacity to attain objective knowledge with respect to this capability should be included in the autonomous learning process. In other words, ideal artificial autonomous agents would be capable of spontaneously *reparametrizing* their representation of the world in relation to novel sensor data. They must not just be capable of updating their model of the world, generated in terms of some particular representational framework; they must also be able to find an appropriate transformation of the representational framework itself so that it most effectively[2] represents the totality of the temporal data.

Artificial agents, in the ideal case, therefore need to mimic the ability of biological agents of updating representational frameworks as sketched above. In the next section, we clarify the details of what this entails; in particular for an artificial agent designed to function, in some sense, in the real world.

[2]In general, this 'most effective representation' will be determined by an efficiency criterion: an agent would typically seek a transformation that minimises complexity (*e.g.* via a Minimum Description Length (Rissanen, 2010) or an Occam's Razor type criterion).

3 Requirements for representational updating in artificial agents

Assuming a world model W (which captures the agents current understanding of the world) and a representational framework R (which governs how this knowledge is represented), ideal agents must – to put the reasoning from the previous section in semi-formal terms – perform the *double* transformational mapping $R[W] \rightarrow R'[W']$, composed of the mappings $R \rightarrow R'$ and $W \rightarrow W'$, such that the W and W' are guaranteed to both represent the same intrinsic set of entities via some "noumenal equivalence" $Equiv(R[W], R'[W'])$.

Certain machine learning paradigms are inherently capable of such a reparametrisation (for example, manifold learning techniques (Zhang et al., 2012) and non-linear dimensionality reduction techniques Debruyne et al. (2010)), but for the present purposes, the specific technique adopted is not significant. The key point is that, at the termination of the process, the agent arrives at *both* a reparameterised perceptual framework R' (such as an orthonormal basis in manifold or sub-manifold coordinates) *and* a revised data set description W' with respect to the representational framework (e.g. following projection into the manifold coordinates). Since the initial, *pre*-exploratory representational framework will necessarily contain much redundancy with respect to an efficient *post*-exploratory representational framework, this reparameterisation will intrinsically involve a reduction in the number of parameters required to represent the data[3]. For example, the determination of some data-derived sub-manifold, M_s, necessarily implicates the existence of a projection operator such that the full range of data in the original domain, $W \subset M$, can be mapped into M_s – for instance, by collapsing data points along the orthogonal complement, $W'^{\perp}$ (M is the original sensory manifold, and M_s the re-mapped representational framework equipped with a suitable basis).

However, whilst there thus exists an *intrinsic* (though likely incomputable) parameterization of any given dataset when considered only in terms of the efficiency of representation, the ideal choice of representation will also necessarily – and similarly to what we previously discussed for biological agents – depend on the *purpose* to which the data set is put. When this purpose is interaction with the physical world, the notion of optimal reparameterization of observed data is not trivial. To give perhaps the simplest example of the resulting complications, we can consider Simultaneous Location and Mapping (SLAM, Engelhard et al., 2011; Strasdat et al., 2010). In this approach, the robotic agent's model of the world necessarily depends upon its calculation of its own position and orientation in the world (*i.e.* it must factor its own perspectival world-view into the world model). However, this positional calculation is *itself* dependent on (is relative to) the agent's model of the world (*i.e.* the agent describes its own position and orientation *in relation to* the world model). A SLAM agent will therefore position itself in the world (perhaps using active learning (Fairfield and Wettergreen, 2010) to minimize model ambiguity) by leveraging its own, uncertain model of the world. Interconnected ambiguities are thus always present in both the agent's self-model (of its location/orientation) and its model of the world, and the hope of SLAM robotics is that, following full exploration of the environment, these ambiguities converge to within some manageable threshold.

In general, the SLAM problem is not solvable unless certain *a priori* assumptions are made. A key such assumption is that the environment remains reasonably consistent over time. If an environment were to undergo some arbitrary spatial transformation at each iteration of the SLAM algorithm, then no convergence would be possible (and in fact there would be no

[3]The criteria for applying such a reductive reparameterisation are open; we could, for example, employ a model selection criterion such as the Akaike Information Criterion to arrive at a principled way to determine the allocation of manifold parameters in relation to the characterization of out-of-model data (the latter is related to minimum-description length (MDL) approaches, which in turn may be considered approximations of the 'intrinsic' (incomputable) Kolmogorov Complexity of the observed data set).

meaning to the concept of world model). However, even much milder perturbations of the spatial domain would be sufficient to ensure non-convergence of the algorithm.

A further key *a priori* assumption, one that shall be particularly important in the following, but which is often overlooked, relates to the robotic agent's *motor capabilities*. The robotic agent's motor capability may, in this case, be considered as *that which initiates the change of perspective/change of representation*. However, as such, it cannot in itself be subject to empirical uncertainty (unlike the world model), and must thus be assumed *a priori*. Colloquially, the agent might thus doubt it's location, or its world model, but it cannot, if it is to work at all, doubt the fact that a specific motor impulse has taken place (for instance, a *move forward* or *turn left* command). The agent cannot converge on a world model if, for instance, motor impulses to the actuators were to undergo arbitrary permutation. Even non-arbitrary permutation would not be distinguishable, even in principle, from a corresponding non-arbitrary permutation of the observed world data. (This non-distinguishability of perceptual manipulations from motor manipulations is absolutely fundamental, and has important consequences in our later argument).

Thus, both the world-model and the agent's (orientation/position-based) self-model are inherently posited relative to its motor impulses, which can be considered to represent the agent's intentions in the sense that the existence of a specific intention is necessarily not itself open to doubt to the agent, however uncertain its perceptual outcome might be. Model convergence on a complete world model occurs only when the outcome of all actions leads to predictable perceptual consequences (to within some given threshold). We can thus consider the world model as being mapped on to a grid of motor impulses such that, in a sense, the agent's active capabilities provide the metric for its perceptual data (see also Dewey, 1896; Glenberg, 1997; Lakoff and Johnson, 1999)).

To summarise, where there exists the capacity for updating the representational capacity of an agent in relation to perceptual data that it has sought-out *on the basis of its original representation*, there also needs to be some mechanism for guaranteeing that there is either sufficient *a priori* noumenal knowledge of the external world, or else that sufficient *a priori* assumptions are made regarding the mechanisms (specifically, agent actions) that initiate new data acquisition for the representation-updating procedure to converge. While this is already a concern in SLAM, the problem is much more acute in fully open-ended learning scenarios where whole new *categories* of perception can be generated. These *a priori* principles are fundamental, and have a long pedigree in philosophical terms; we discuss this next.

4 Fundamental epistemic restrictions on representational cognitive agency

4.1 Noumenal continuity across representational changes

We are essentially asking how, in an adaptive online learning context, it is possible to empirically validate a proposed change to an agent's representative capability (how is it, in a Popperian sense, possible to *falsify* a proposed representational update). Falsification of a world model is, by comparison, straightforward in a standard autonomous robotic system, in that a world model typically constitutes a set of proposed action possibilities (Gibson, 1979; McGrenere and Ho, 2000) gathered at-a-distance by a vision system. Thus, the visual model typically denotes a set of object hypotheses that may be verified via haptic contact (Saunders and Knill, 2004; Schlicht and Schrater, 2003).

In robotics, haptic contact is consequently often considered to be prior to vision, or at least *a priori* less prone to ambiguity than vision, something also evident in human terms.

However, in a hypothetical agent where there exists complete representational fluidity, such that a completely novel sensorium could be presented to an agent (for instance if sonar data were combined with visual data in some hybrid world description without any prior information as to the nature of the former), then it would not be possible to intrinsically favour one group of senses/sensors over another in order to delineate hypotheses about the world. Moreover, there would be no immediately obvious way to form hypotheses about the most appropriate representational framework to adopt.

To address this, we borrow a key insight from the philosopher Kant; namely that object concepts constitute *orderings* of sensory intuitions (Kant, 1999). Objects, as we understand them do not thus constitute singular percepts, but rather *synthetic unities* built upon an *a priori* linkage that must be assumed between sensory intuitions and the external noumenal world (these a priori links cannot be in doubt since they are a condition of empirical validation for synthetic unities). Implicit in this is the notion that actions can be deployed to test the validity of these synthetic unities (which being synthetic rather than analytic are only contingently true, and therefore falsifiable through experience). Actions are thus causally initiated by the agent and serve to bring aspects of the synthetic unities to attention (within the a priori strata of space and time) in a way that renders them falsifiable.

For Kant, assuming that spatiality and temporal causality are a priori means that they are assumed by the agent in order to have falsifiable perceptions at all. In principle, other ordering approaches to sensory data may be possible; however, it would be impossible for the agent to retain the continuity and falsifiability of object representation across such a fundamental transition of representation (it would also be impossible for a self-conscious agent to retain its identity – or "synthetic unity of apperception" – across such a fundamental representational chasm). This is the problem of "noumenal continuity": how can an agent that undergoes a change of representation framework at time t_0 ever be sure that the objects delineated at $t_0 - 1$ were the same objects as those delineated at $t_0 + 1$? Indeed, would the number of objects even be preserved?

In principle, there is thus a clear risk that an agent that undergoes a representational change would be severely limited in the extent to which it could use existing knowledge *across* these changes. One way to avoid this risk is when representational changes are built *hierarchically.* Such an approach will preserve an agent's ability to falsify both the representational changes as well as any object hypotheses (synthetic unities) formed in terms of these. Moreover, it does so while retaining online continuity of object identity when extended in perception-action terms.

We will demonstrate this for the example of hierarchical Perception-Action learning in the next section. By way of example though, consider how we, as humans, typically represent our environment when driving a vehicle. At one level, we internally represent the immediate environment in metric-related terms (*i.e.* we are concerned with our proximity to other road users, to the curb and so on, see Windridge et al., 2013b). At a higher level, however, we are concerned primarily with navigation-related entities (—e.g. how individual roads are connected). That the latter constitutes a higher hierarchical level, both mathematically and experientially, is guaranteed by the fact that the topological representation subsumes, or supervenes upon, the metric representation; that is to say, the metric-level provides additional "fine-grained" information to the road topology: the metric representation can be reduced to the topological representation, but not vice versa.

When goals and sub-goals are explicitly delineated at each level, this is known in robotics as a *subsumption hierarchy* (Brooks, 1991). What we argue for in this paper is that perceptual subsumption and task subsumption need to be directly related to each other in an adaptive cognitive agent in order to achieve the maximal cognitive updating potential. In a fully adaptive online learner, it is then possible to allow representational induction by adopting a correspondingly hierarchical approach.

Thus, on the assumption of the existence of a priori means of validating low-level hypotheses (for example via haptic contact), it is possible to construct falsifiable higher-level representational hypothesis *provided that these subsume the latter*. For example, an autonomous agent might, following active experimentation, spontaneously conceive a high-level concept of action possibilities, or schema (Hintzman, 1986), such as that provided by a *container*. Clearly, in this case, the notion *container* subsumes the concept of *haptic contact*.

Continuity of noumenal identity must thus be guaranteed by the lowest level of the hierarchy, with the higher hierarchical levels then constituting progressive *abstractions*[4] and enrichments of the lower level representations. For example, an autonomous agent might initially represent the world in terms of (hypothetical) volume elements such as voxels or 3d meshes (the a priori bootstrap representation), but, following extensive experimentation, go on to generate an enriched representation of its world at a higher level in which "containers" and "non-containers" are the symbols in terms of which the world is delineated (note that the original, pre-symbolic representation of the world in volumetric terms remains subsumptively present).

4.2 Falsification of representational changes in terms of utility and compressibility

The falsifiability of aspects such as the symbolic representational notion "container" is guaranteed, just as it is possible to guarantee the falsifiability of the hypothesis of the existence of any specific container, by exploiting the fact that all such hypotheses can be linked to the lowest level of the hierarchy, at which they are rendered falsifiable by haptic contact: simply put, just as the agent can verify the *presence* of a container by testing whether the proposed container-entity is, in fact, capable of containing another object, the *high-level representational concept* "container" is rendered falsifiable by the fact that it is conceived along with a corresponding high-level action possibility e.g. "placing an object into a container" that necessarily subsumes lower-level concepts such as "haptic contact". Thus, the representational concept is rendered falsifiable on the basis of its *utility* and *compressibility*.

In other terms, the falsifiability of the representational notion "container" arises from actively addressing the question of whether this higher-level perception of the world (in terms of a series of objects in space that are either container-objects or non-container-objects) in fact constitutes a useful description of the world *i.e.* whether it yields a net compression in the agent's internal representation of its own possible interactions with the world. For example, if there were only a single container in the world, or if it were not possible to train an accurate classifier for containers in general, then it would be unlikely to constitute a useful description of the world; it would likely be more efficient simply to retain the existing concept of object without modification. However, when the world is in fact constituted of objects for which it is indeed an efficient compression of the agent's action-capabilities to modify the object concept in this way, then it *is* appropriate for a representationally-autonomous agent to spontaneously form a higher level of its representational hierarchy (for an example of this approach utilizing first-order logic induction, see Windridge and Kittler, 2010)).

Very often, compressibility will be predicated on the discovery of invariances in the current perceptual space with respect to randomized exploratory actions. Thus, for example, an agent might progress from a pixel-based representation of the world to an object-based representation of the world via the discovery that certain patches of pixels retain their (relative) identity under translation, such that it becomes far more efficient to represent the world in terms of indexed objects rather than pixel intensities (though the latter would, of course, still constitute the base

[4]These abstractions can be conceived of as symbolic. For example, Eliasmith (2013) proposes a cognitive architecture in which the symbolic entities manipulated in higher level cognition are built from successive (compressive) abstractions in the sensorimotor hierarchies.

of the representational hierarchy). This particular representational enhancement can represent an enormous compression (Wolff, 1987); a pixel-based representation has a parametric magnitude of P^n (with P and n being the intensity resolution and number of pixels, respectively), while an object-based representation typically has a parametric magnitude of $\sim n^o$, $o << n$, where o is the number of objects.

In positing this hierarchical approach to representational adaptation, we have thus outlined a framework in which complete representational-autonomy for an embodied machine learner becomes feasible, one in which representations are empirically falsifiable, and in which the noumenal continuity of identified entities can be assumed across representational transformations. A key aspect of this falsifiability is the requirement that the spontaneous generation of higher-level perceptions in the agent's representational hierarchy correlates directly with higher level actions. We now look more closely at this perception-action connection, and consider the low-level *a priori* guarantees of representative falsifiability.

5 Example: hierarchical Perception-Action learning

To conclude our main argument, we demonstrate in this section, how the above considerations can be implemented in practice in hierarchical Perception-Action learning architectures. Perception-Action (P-A) learning is a paradigm in robotics that aims to address significant deficits in traditional approaches to computer vision (Dreyfus, 1972). In particular, in the conventional approach to autonomous robotics, a computer vision system will typically be employed to construct a world model of the agent's environment *prior* to the act of planning the agent's actions within the domain. Visual data arising from these actions will then typically be used to further constrain the environment model, either actively or passively (in active learning the agent actions are driven by the imperative of reducing ambiguity in the environment model).

However, it is apparent that there exists in this approach, a very wide disparity between the visual parameterization of the agent's domain and its action capabilities within it (Magee et al., 2004; Nehaniv et al., 2002). For instance, the parametric freedom of a front-mounted camera will typically encompass the full intensity ranges of the Red, Green and Blue channels of each individual pixel of the camera CCD; thus the range of possible images that might be generated in each time-frame is of an extremely large order of magnitude, despite only a minuscule fraction of this representational space being ever likely to be experienced by the agent. By contrast, the agent's motor capability is likely to be very much more constrained parametrically (perhaps consisting only of the possible Euler angle settings of the various actuator motors). This disparity leads directly to the classical problems of framing (McCarthy and Hayes, 1969), an issue shared with alternative modalities to vision, such as LIDAR and SONAR.

5.1 P-A learning

P-A learning aims to overcome these issues by considering actions to be conceptually prior to perceptions (Granlund, 2003; Felsberg et al., 2009).. In other words, perceptual capabilities should depend on action capabilities and not vice versa. A P-A learning agent proceeds by randomly sampling its action space (so-called motor babbling). For each motor action that produces a discernible perceptual output in the bootstrap representation space S (consisting of *e.g.* camera pixels), a percept $p_i \in S$ is greedily allocated. The agent thereby progressively arrives at a set of novel percepts that relate directly to the agent's action capabilities in relation to the constraints of the environment (*i.e.* the action possibilities that exist in the environment): the agent learns to perceive only that which it hypothesises that it can change. Thus, the set of

experimental data points $\cup_i p_i \subset S$ can, in theory, be generalized over so as to create a percept-*manifold* that can be mapped onto the action space via the injective relation $\{actions\} \rightarrow \{percept_{\text{initial}}\} \times \{percept_{\text{final}}\}$ (Windridge and Kittler, 2010, 2008; Windridge et al., 2013a).

When such a perceptual manifold is created (representing a *generalization* over the tested space of action possibilities), this then permits an *active* sampling of the perceptual domain - the agent can propose actions with perceptual outcomes that have not yet been experienced by the agent, but which are consistent with its current representational model (which guarantees falsifiability of the perceptual model). It is in this way that P-A-learning constitutes a form of active learning: randomized selection of perceptual goals within the hypothesized perception-action manifold leads more rapidly to the capture of data that might falsify the hypothesis than would otherwise be the case (for example, if the agent were performing randomly-selected actions within in the original motor domain). Thus, while the system is always "motor babbling" in a manner analogous to the learning process of infant humans, the fact of carrying out this motor babbling in a higher-level P-A manifold ensures that the learning system as a whole more rapidly converges on the correct model of the world.

5.2 Hierarchical P-A learning

In principle, this P-A motor-babbling activity can take place in *any* P-A manifold, of whatever level of abstraction; we may thus, by combining the idea of P-A learning with the notion of hierarchical representation presented above, conceive of the notion of a *hierarchically subsumptive perception-action learner*, in other words combining Brooke's notion of task subsumption with the P-A notion of action preceding perception. Such a system, employing iterative top-down motor-babbling and bottom-up parametric reparameterization to generate a PA subsumption hierarchy, was practically demonstrated, for example, by Windridge and Kittler (2007) and Shevchenko et al. (2009).

In these systems, a vertical representation hierarchy is progressively constructed for which randomized exploratory motor activity at the highest level of the corresponding motor hierarchy rapidly converges on an ideal representation of the agent's world in terms of its symbolic affordance potentialities. These systems thus converge upon both a model of the world, and an ideal strategy for its representation in terms of the learning agent's action capabilities within it (the generalization and parametric compression mechanisms in these systems, however, were extremely different; employing string concatenation with redundancy elimination, and first-order rule induction with reverse instantiation, respectively).

Perceptual goals thus exist at all levels of the perception-action hierarchy, and the subsumptive nature of the hierarchy means that goals and sub-goals are scheduled with increasingly specific content as high-level symbolic goals (such as "place ball in cup") are progressively grounded through the hierarchy. To pick up an earlier example, as humans, we may conceive the high-level intention "drive to work", which in order to be enacted, involves the execution of a large range of sub-goals with correspondingly lower-level perceptual goals such as "stay in the center of the lane", and so on.

5.2.1 Bijectivity constraints between action possibilities and percepts

To ensure that these hierarchical goals are most efficiently represented, it is necessary to impose a *bijectivity* between actions and possible percept transitions in order to induce the correct relationship between representational subsumption for percepts P and Brooke's task subsumption in relation to actions A. In particular, in order to retain falsifiability of a proposed new action possibility, it is necessary to impose up on any hypothesized new hierarchical perception-action level the bijective constraint $\{\text{actions}\} \leftrightarrow \{P_{\text{initial}}\} \times \{P_{\text{final}}\}$, where the initial percepts are the

necessry observed state of the world to initiate the new high-level action and the final percepts are the target observed state of the world expected at the end of the action.

Critically, this construction permits top-down instantiation of goal parameters: a high level action naturally schedules a series of subtasks at different levels of the task hierarchy (the actual sequencing of actions will depend on the optimisation mechanism). Also importantly, each subtask has its own perceptual goal of the appropriate level of hierarchical complexity (*i.e.* the appropriate 'depth'). Each target object must thus have a representation at *every* level of the hierarchy (e.g. a "container" is also an "object" and is also "voxel cloud"). This representational subsumption is what allows the agent to falsify a spontaneously-hypothesized P-A notion such as *container*[5].

5.2.2 Representational shifts in hierarchical P-A learning

The learning mechanism required in a bootstrap hierarchical P-A learning agent is dictated by the supervised classification problem intrinsic in generalising the outcome of exploratory actions driven at the highest level of the hierarchy. In particular, the outcomes of exploratory actions (predicated for instance on the proposed notion "container") result either in the successful achievement of the final perceptual state or its failure. Each exploratory action can thus be accompanied by a binary label $\{\text{achieved}, \text{not\_achieved}\}$. The set of exploratory actions then form a training set that a supervised classification system can generalize over. The generalized set of actions (with appropriate perceptual goals) classified as achievable thus represents a set of *testable action hypotheses*.

However, this generalization is not in itself sufficient to give rise to a new (hypothesized) level of the hierarchy (and thus a new representational framework); for this, we require that the set of percepts corresponding to the goal states of the generalised action can undergo parametric reduction. In particular, they must be capable of parametric reduction such that the bijectivity constraint $\{\text{Action Possibilities}\} \leftrightarrow \{P_{\text{initial}}\} \times \{P_{\text{final}}\}$ holds (this perceptual parametric reduction naturally implies a novel higher-level action hypothesis). Only in this way can high-level symbolic propositions such "place the ball in the cup" be formulated *ab initio*.

This cycle (exploration, induction, perceptual reparameterization/high-level action generation) can be iterated over until convergence is achieved (when all action goals hypothesized to be achievable are in fact achievable). It was shown by Windridge and Kittler (2007) and Shevchenko et al. (2009) that this is a form of active learning that can speed up world-model learning by several orders of magnitude. Further, motor babbling within such an iterative bootstrap P-A learning system can be shown to necessarily become increasingly *intentional* as time progresses; an initial low level exploratory impulse generated randomly results in apparently random movement similar to that of a new-born child, while a randomly-generated high-level exploratory impulse instantiating the perceptual parameters of, for example, the conjectural action possibility "put into" would result in the apparently-purposive action, such as placing a ball into a cup.

In terms of the previously discussed P-A bijectivity, the high-level action "put into" is parametrised by the symbolic notion of "container". The parametrisation can be treated in terms of first-order predicate logic, with the action predicate '$Put\_Into_n(O_{n-1}, C_n)$' being scoped over the the object variable O_{n-1} and the container variable C_n. The symbolic per-

[5]Interestingly, representational subsumption occurs spontaneous in convolutional neural networks and is a key factor in their state-of-the-art performance (Ranzato et al., 2007; Girshick et al., 2014; Liu et al., 2017), not least because of the ready transferability of representations (Oquab et al., 2014). The bijectivity constraint can thus in principle be incorporated into the deep-learning objective function of an embodied agent to produce an agent capable of open-ended learning (deep visuo-motor learning having already been demonstrated Porzi et al., 2017; Levine et al., 2016).

ceptual notion "container" hence subsumes the symbolic perception notion "object". This is a strict form of conceptual subsumption that, when translated via the bijectivity principle into the action space, becomes equivalent to Brooke's notion of task subsumption: the perceptual goal implicit in the action specifies a target state that a lower-level task must be scheduled to achieve – for example, tasks of the form "place hand around object", "move hand" , and so on. The P-A bijectivity principle does not specify *how* the task is to be optimally; this is a free (and potentially hybrid) mechanical choice within the framework. For example, in an artificial agent, optimal control may be used at some particular hierarchical level, while simulated annealing might be employed on another level to optimize task scheduling.

Note that for this subsumptive task-scheduling to be possible at all, the perceptual target must have a corresponding perceptual subsumption. To place and object in a container, the agent must move its object-containing end effector toward a container *object* - on this level, the target is an object, and only with respect to the higher-level action '$Put\_Into_n(O_{n-1}, C_n)'$ is the object *also* recognised as a container; i.e. it has an additional, higher-level action possibility characterised by the variable C_n. In other words, a *specific* instantiation of a container in C_n also necessarily implies a corresponding instantiation of an object in O_{n-1}, representing the targeted entity considered only as an object. A cup is thus both an entity for containing coffee and a solid object, with, for example, mass and a certain geometrical configuration).

The induced PA concept "container" is thus instantiated by specific container-objects, however the notion itself is a generalization of the action possibilities of containers in general. The induced notion "container" is thus symbolic in that it can be employed (since it is capable of entering into discrete, relational juxtapositions with other symbolic entities) to pursue potentially counter-factual possibilities via instantiation (for example, by attempting to treat a random object as a container).

6 Discussion and Conclusions

In this paper, we explored what necessary mechanisms an artificial agent needs to possess to achieve certain aspects of human cognition. In particular, we highlighted the ability to update representational frameworks in a manner that is useful to agent in question. We highlighted the importance of noumenal continuity in relation to the question of empirical validity and concluded that this can be achieved through an appropriate subsumptive architecture embodying a perception-action bijectivity criterion. We went on to demonstrate that hierarchical P-A learning is a framework in which such an architecture can be realised. Overall, the proposed framework for spontaneous symbol abstraction in open-ended cognitive learning is thus intended to be of maximal generality, being learning-mechanism agnostic, on the proviso that the bijectivity constraint is observed. The contributions of this framework are two-fold, having implications for both theories of cognition and the design of artificial agents, that we discuss here.

Throughout the paper, we have retained the assumption that cognition can generally be thought of as symbolic computation. As far as theories of embodiment go (Chemero, 2009), this is a weak position: the body exists initially only because we are interested in agents that operate in the physical world. Nonetheless, we quickly found that the body must play a role that goes far beyond being a mere sensorimotor interface to the world. An agent that proposes a representational update must also propose a way to falsify this representational update in a Popperian sense. We have shown, following Kant, that this must be achieved through a priori action possibilities. This is the crucial aspect that fundamentally involves the body in even otherwise computational processes: the representational frameworks that are used are entirely dependent on the embodiment of the agent.

The role of the embodiment of an agent therefore goes beyond an interface between a computational model and the physical world; it also goes beyond merely providing a mechanism by which to ground symbols (Harnad, 1990); rather it determines the representational framework used for computation itself. The theoretical contribution to the study of cognition therefore is that, even on the computational end of the spectrum, the body has to be more than a mere interface. It follows from this that the precise nature of the body must be considered even if one is otherwise interested in constructing a purely computational account of cognition. First of all, this is because perception is not purely external; living beings also integrate interoceptive information (Stapleton, 2011) in cognitive mechanisms. Second, if an agent's perceptive and motor abilities shape the framework in which its cognitive processes take place, then characterising this fully might require a detailed understanding of the precise nature of these abilities. For example, a human cannot pick up a red-hot piece of iron while a robot might, even if both have appropriately shaped grippers. That said, it remains to be explored what precisely the consequences of differences in body are for higher-level cognition in particular (see also Thill, 2018).

We note of course that in 4E approaches to cognition, or even non-representationalist interpretations (Chemero, 2009), the role of the body is arguably embedded significantly further. Ziemke (2016), for example, reviews a number of frameworks that focus on the role of internal bodily mechanisms – for example, homeostatic mechanisms – in grounding sensorimotor interaction itself, concluding, "[a]t least in the case of natural cognition, that sensorimotor interaction with the environment is itself deeply rooted in the underlying biological mechanisms, and more specifically layered/nested networks of bodily self-regulation mechanisms". P-A architectures such as the one considered here have also been considered in more deeply embodied terms than we do here; Vernon et al. (2015), for example, discuss the relevance of the internal body (again, including mechanisms such as homeostatic regulation) in achieving such a P-A coupling in natural agents.

Here, we have therefore merely demonstrated a minimally necessary role. This argument also demonstrates that using a robot merely to ground symbols (*e.g.* Stramandinoli et al., 2011) is not enough to claim an embodied model of cognition in a meaningful sense (though it is of course valid, as in the cited study, to explore specific aspects of embodiment, such as symbol grounding, in this manner).

For agents that have no particular embodiment, for example systems as sometimes envisioned by some proponents of so-called Artificial General Intelligence, the lack of means to falsify representational updates through action linkages in the Kantian sense means that there is no way of meaningfully generating such updates since there is no principled way of falsifying them. Any proposed representational framework would be feasible in principle, but without a means to evaluate its utility in the world the sole remaining *intrinsic* criterion for favouring one framework over another would be its compressive capability (natural selection as a means of establishing framework utility would obviously also be inapplicable to a disembodied artificial agent). The only fully model-independent (which is to say representation-independent) criterion for compressibility that could be applied would involve a determination of the underlying Kolmogorov complexity of the input stream; however, this is incomputable even in principle. There would also be no principled *a priori* reason as to why compression could not be allowed to be lossy, and therefore nothing to prevent the system from collapsing all inputs into a single bit. To avoid this, some additional metric might be artificially imposed to assess whether the framework update is useful to the functioning of the system; however this would necessarily only implicate a criterion of *success* or *completion* with regard to the updating procedure rather than a criterion of *falsifiability*. If the ability for representational updating is crucial for the general nature of human intelligence, then the inability to achieve this without embodiment fundamentally limits the utility of disembodied models of cognition, at least in terms of achieving such

a generality.

There are a few points to note about the framework proposed here. First, at no stage is there any requirement for global hierarchical consistency of representation (for example, as humans, we do not embody a set of Cartesian coordinates or similar to describe the geography of our locale; rather, we retain a series of motor imperatives that are triggered in relation to key percepts: e.g. we thus 'turn left at the town-hall' rather than head to a particular set of coordinates when navigating). For a Perception-Action learning agent, the environment "becomes its own representation" (Newell and Simon, 1976), which naturally represents a very significant compression of the information that an agent needs to retain. This further relates to the issue of symbol grounding (Harnad, 1990) in that symbolic representations are abstracted from the bottom-up here (Marr, 1982; Gärdenfors, 1994; Modayil, 2005; Granlund, 2003). In principle, the present framework is thus a variation on the notion of perceptual symbol systems (Barsalou, 1999), and a symbolic description is arrived at similarly than in related approaches (for instance the so-called Semantic Pointer Architecture, see Eliasmith, 2013). As in those approaches, symbols are thus always intrinsically grounded by nature of their construction. The distinguishing feature is that in the current context, this grounding is also the guarantor of their falsifiability, as required for representational upgrading[6].

We also observed, in passing, that motor-babbling at the top of the representation hierarchy would necessarily involve the spontaneous scheduling of perceptual goals and sub-goals at the lower level of the hierarchy in a way that (as the hierarchy becomes deeper) looks increasingly "intentional" (a phenomenon that is readily apparent in the development of motor movement in human infants). This has implications for social robotics; in particular, it becomes possible to envisage *communicative* actions. Here, the same bijectivity considerations apply to perceptions and actions as before, however the induction and grounding of symbols can in principle now be conducted through linguistic exchange. The utility constraint on generated symbols remains: they have to relate to a compact and useful set of action possibilities. In addition, however, these action possibilities must be common to the communicating agents since meaningful linguistic exchange can only occur between agents with similar sensorimotor capabilities, a notion that relates directly Wittgenstein's concept of the *language game* for which the idea of a private language is meaningless.

In principle, such an approach would implicitly be a hierarchical generalisation of Steels's (1997) famous talking heads experiment in artificial language formation. We can thus envisage the coeval generation of perceptual symbols and their corresponding actions within a community of agents employing P-A subsumption. Typically, the most efficient form of communication between individuals is in terms of the highest levels of the P-A hierarchy, such that recipients of a linguistic token ground its meaning via their internal P-A hierarchy (thus we tell someone to "watch out for the car", rather than instructing them on which specific muscles to activate in order to accomplish this task). In related work, Thill and Twomey (2016) explore these issues in more detail, and, in particular, discuss how a framework such as that of Eliasmith (2013) – which also employs hierarchical structures to derive symbol-like entities – can be used to investigate how exactly differences in the sensorimotor experiences of two agents (living or artificial) might impact their ability to communicate about concepts.

To conclude, we have demonstrated how representational fluidity, a necessary component of cognition, can be achieved in a computationalist framework. In addition to the specific contributions discussed above, this is of general relevance for studying agents, living or artificial, for whom such a fluidity might play a role in cognition. We note of course that this is to some degree a philosophical question first: if one rejects representationalist frameworks outright as

[6]It can also be noted that, in humans at least, internal bodily mechanisms as reviewed by Ziemke (2016), are likely relevant for this functionality. We do not explore it further here since we are primarily concerned with a more computationalist stance.

an adequate way of modelling cognition, then there is also little point in considering representational fluidity in the way we have here. There is, however, a point in considering the role of hierarchical structures, and P-A coupling in cognition. In particular, the idea that agents might guide the exploration of their abilities and environments using hypotheses that are falsifiable through action does not depend on a computationalist account. As such, much of what we have discussed in more theoretical terms in this paper still applies, even if a model to demonstrate this will rather look rather different.

Acknowledgment

We acknowledge financial support from the EC H2020 research project Dreams4Cars (no. 731593).

References

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4):577–660.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, 47:139–159.

Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3):15.

Chemero, A. (2009). *Radical Embodied Cognitive Science*. MIT Press, Cambridge, MA.

Chersi, F., Thill, S., Ziemke, T., and Borghi, A. M. (2010). Sentence processing: linking language to motor chains. *Frontiers in Neurorobotics*, 4(4).

Debruyne, M., Hubert, M., and Van Horebeek, J. (2010). Detecting influential observations in kernel pca. *Computational Statistics & Data Analysis*, 54(12):3007–3019.

Dewey, J. (1896). The reflex arc concept in psychology. *The Psychological Review*, (3):356–370.

Dreyfus, H. (1972). *What Computers Can't Do*. New York: Harper and Row.

Eliasmith, C. (2013). *How to build a brain: a neural architecture for biological cognition*. Oxford University Press, Oxford.

Engelhard, N., Endres, F., Hess, J., Sturm, J., and Burgard, W. (2011). Real-time 3d visual slam with a hand-held rgb-d camera. In *Proc. of the RGB-D Workshop on 3D Perception in Robotics at the European Robotics Forum, Vasteras, Sweden*, volume 2011.

Fairfield, N. and Wettergreen, D. (2010). Active slam and loop prediction with the segmented map using simplified models. In *Field and Service Robotics: Results of the 7th International Conference*, volume 62, page 173. Springer.

Felsberg, M., Wiklund, J., and Granlund, G. (2009). Exploratory learning structures in artificial cognitive systems. *Image and Vision Computing*, 27(11):1671–1687.

Gärdenfors, P. (1994). How logic emerges from the dynamics of information. *Logic and Information Flow*, pages 49–77.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton-Mifflin, Boston.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 580–587.

Glenberg, A. (1997). What memory is for. *Behavioral and Brain Sciences*, 20(1):1–55.

Granlund, G. (2003). Organization of architectures for cognitive vision systems. In *Proceedings of Workshop on Cognitive Vision*, Schloss Dagstuhl, Germany.

Harnad, S. (1990). The symbol grounding problem. *Physica D*, (42):335–346.

Hintzman, D. L. (1986). Schema abstraction in a multiple-trace memory model. *Psychological review*, 93(4):411–428.

Hoffman, M., Blei, D. M., and Bach, F. (2010). Online learning for latent dirichlet allocation. *Advances in Neural Information Processing Systems*, 23:856–864.

Kant, I. (1999). *Critique of Pure Reason*. Cambridge University Press.

Koltchinskii, V. (2010). Rademacher complexities and bounding the excess risk in active learning. *The Journal of Machine Learning Research*, 11:2457–2485.

Lakoff, G. and Johnson, M. (1999). *Philosophy in the Flesh : The Embodied Mind and Its Challenge to Western Thought*. Harper Collins Publishers.

Levine, S., Finn, C., Darrell, T., and Abbeel, P. (2016). End-to-end training of deep visuomotor policies. *Journal of Machine Learning Research*, 17(39):1–40.

Liu, M., Shi, J., Li, Z., Li, C., Zhu, J., and Liu, S. (2017). Towards better analysis of deep convolutional neural networks. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):91–100.

Magee, D., Needham, C. J., Santos, P., Cohn, A. G., and Hogg, D. C. (2004). Autonomous learning for a cognitive agent using continuous models and inductive logic programming from audio-visual input. In *Proc. of the AAAI Workshop on Anchoring Symbols to Sensor Data*.

Mahon, B. Z. and Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris*, 102(1):59 – 70. Links and Interactions Between Language and Motor Systems in the Brain.

Marr, D. (1982). *Vision: A Computational Approach*. Freeman & Co., San Fr.

McCarthy, J. and Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence*, (4):463–502.

McGrenere, J. and Ho, W. (2000). Affordances: Clarifying and evolving a concept. In *Proceedings of Graphics Interface 2000*, pages 179–186, Montreal, Canada.

Millikan, R. G. (1987). *Language, Thought, and Other Biological Categories: New Foundations for Realism*. The MIT Press; Reprint edition.

Modayil, J. (2005). Bootstrap learning a perceptually grounded object ontology. Retr. 9/5/2005 http://www.cs.utexas.edu/users/modayil/modayil-proposal.pdf.

Nehaniv, C. L., Polani, D., Dautenhahn, K., te Boekhorst, R., and Canamero, L. (2002). Meaningful information, sensor evolution, and the temporal horizon of embodied organisms. In Standish, Abbass, B., editor, *Artificial Life VIII*, pages 345–349. MIT Press.

Newell, A. and Simon, H. (1976). The theory of human problem solving; reprinted in collins & smith (eds.). In *Readings in Cognitive Science, section 1.3.*

Oquab, M., Bottou, L., Laptev, I., and Sivic, J. (2014). Learning and transferring mid-level image representations using convolutional neural networks. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1717–1724.

Pan, S. J. and Yang, Q. (2010). A survey on transfer learning. *Knowledge and Data Engineering, IEEE Transactions on*, 22(10):1345–1359.

Piaget, J. (1970). *Genetic Epistemology.* Columbia University Press, New York.

Porzi, L., Buló, S. R., Penate-Sanchez, A., Ricci, E., and Moreno-Noguer, F. (2017). Learning depth-aware deep representations for robotic perception. *IEEE Robotics and Automation Letters*, 2(2):468–475.

Ranzato, M., Huang, F. J., Boureau, Y. L., and LeCun, Y. (2007). Unsupervised learning of invariant feature hierarchies with applications to object recognition. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8.

Rissanen, J. (2010). *Minimum description length principle.* Springer.

Saunders, J. and Knill, D. C. (2004). Visual feedback control of hand movements. *J. of Neuroscience*, 24(13):3223–3234.

Schlicht, E. J. and Schrater, P. R. (2003). Bayesian model for reaching and grasping peripheral and occluded targets. *Journal of Vision*, 3(9):261.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3:417–424.

Settles, B. (2010). Active learning literature survey. *University of Wisconsin, Madison.*

Shevchenko, M., Windridge, D., and Kittler, J. (2009). A linear-complexity reparameterisation strategy for the hierarchical bootstrapping of capabilities within perception–action architectures. *Image and Vision Computing*, 27(11):1702–1714.

Sipper, M. (1995). An introduction to artificial life. *Explorations in Artificial Life (special issue of AI Expert)*, pages 4–8.

Stapleton, M. (2011). *Proper embodiment: the role of the body in affect and cognition.* PhD thesis, The University of Edinburgh.

Steels, L. (1997). The origins of syntax in visually grounded robotic agents. In Pollack, M., editor, *Proceedings of the 10th IJCAI, Nagoya*, pages 1632–1641. AAAI Press, Menlo-Park Ca.

Stevens, C. and Neville, H. (2006). Neuroplasticity as a double-edged sword: Deaf enhancements and dyslexic deficits in motion processing. *Journal of Cognitive Neuroscience*, 18(5):701–714.

Stramandinoli, F., Cangelosi, A., and Marocco, D. (2011). Towards the grounding of abstract words: A neural network model for cognitive robots. In *Neural Networks (IJCNN), The 2011 International Joint Conference on*, pages 467–474.

Strasdat, H., Montiel, J., and Davison, A. (2010). Scale drift-aware large scale monocular slam. In *Proceedings of Robotics: Science and Systems (RSS)*, volume 2, page 5.

Taylor, M. E. and Stone, P. (2009). Transfer learning for reinforcement learning domains: A survey. *The Journal of Machine Learning Research*, 10:1633–1685.

Thill, S. (2018). What we need from an embodied cognitive architecture. In Sequeira, J., Ventura, R., and Ferreira, I., editors, *Cognitive Architectures*, Intelligent Systems, Control and Automation. Springer. In press.

Thill, S. and Twomey, K. (2016). What's on the inside counts: A grounded account of concept acquisition and development. *Frontiers in Psychology: Cognition*, 7(402).

Vernon, D., Lowe, R., Thill, S., and Ziemke, T. (2015). Embodied cognition and circular causality: on the role of constitutive autonomy in the reciprocal coupling of perception and action. *Frontiers in Psychology: Cognitive Science*, 6(1660).

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4):625–636.

Windridge, D., Felsberg, M., and Shaukat, A. (2013a). A framework for hierarchical perception-action learning utilizing fuzzy reasoning. *Cybernetics, IEEE Transactions on*, 43(1):155–169.

Windridge, D. and Kittler, J. (2007). Open-Ended Inference of Relational Representations in the COSPAL Perception-Action Architecture. In *Proc. of International Conf. on Machine Vision Applications (ICVS 2007)*, Germany.

Windridge, D. and Kittler, J. (2008). Epistemic constraints on autonomous symbolic representation in natural and artificial agents. In *Studies in Computational Intelligence: Applications of Computational Intelligence in Biology*, volume 122, pages 395–422. Springer Berlin Heidelberg.

Windridge, D. and Kittler, J. (2010). Perception-action learning as an epistemologically-consistent model for self-updating cognitive representation. In *Brain Inspired Cognitive Systems 2008*, pages 95–134. Springer.

Windridge, D., Shaukat, A., and Hollnagel, E. (2013b). Characterizing driver intention via hierarchical perception-action modeling. *Human-Machine Systems, IEEE Transactions on*, 43(1):17–31.

Wolff, J. G. (1987). Cognitive development as optimisation. In Bolc, L., editor, *Computational Models of Learning*, pages 161–205, Heidelberg. Springer-Verlag.

Zhang, Z., Wang, J., and Zha, H. (2012). Adaptive manifold learning. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(2):253–265.

Ziemke, T. (2003). What's that thing called embodiment? In *Proceedings of the 25th Annual meeting of the Cognitive Science Society*, pages 1305–1310.

Ziemke, T. (2016). The body of knowledge: On the role of the living body in grounding embodied cognition. *Biosystems*, 148:4 – 11.

Ziemke, T. and Thill, S. (2014). Robots are not embodied! conceptions of embodiment and their implications for social human-robot interaction. In *Proceedings of Robo-Philosophy 2014: Sociable robots and the future of social relations*, pages 49– 53, Amsterdam, NL. IOS Press BV.