

Introducing a Pictographic Language for Envisioning a Rich Variety of Enactive Systems with Different Degrees of Complexity

Invited Feature Article

Roger K. Moore^{1*}

¹ University of Sheffield, Sheffield, UK

*Corresponding author(s) E-mail: r.k.moore@sheffield.ac.uk

Received 17 June 2015; Accepted 11 January 2016

DOI: 10.5772/62244

© 2016 Author(s). Licensee InTech. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Notwithstanding the considerable amount of progress that has been made in recent years, the parallel fields of cognitive science and cognitive systems lack a unifying methodology for describing, understanding, simulating and implementing advanced cognitive behaviours. Growing interest in 'enactivism' - as pioneered by the Chilean biologists Humberto Maturana and Francisco Varela - may lead to new perspectives in these areas, but a common framework for expressing many of the key concepts is still missing. This paper attempts to lay a tentative foundation in that direction by extending Maturana and Varela's pictographic depictions of autopoietic unities to create a rich visual language for envisioning a wide range of enactive systems - natural or artificial - with different degrees of complexity. It is shown how such a diagrammatic taxonomy can help in the comprehension of important relationships between a variety of complex concepts from a pan-theoretic perspective. In conclusion, it is claimed that visual language is not only valuable for teaching and learning, but also offers important insights into the design and implementation of future advanced robotic systems.

Keywords Cognitive Science, Cognitive Systems, Robotics, Enaction, Visual Taxonomy, Pictographic Language

1. Introduction

Of the many difficulties faced by researchers in the parallel fields of cognitive science and cognitive systems, by far the most challenging is the sheer complexity of the entities being studied [1]. Even the simplest living organism or 'intelligent' autonomous agent is composed of a multitude of interconnected systems and subsystems, all of which are required to operate in such a way that appropriately organized (and effective) behaviours emerge. It is perhaps for these reasons that contemporary knowledge and solutions in this area inhabit a rich and varied conceptual space supported by a multitude of theoretical stances, empirical evidence, design architectures and technical implementations, which can be hard to relate to each other. This means that progress in our understanding, simulation and realization of advanced cognitive behaviours may be hampered by a lack of coherence at the meta-theoretical level [2].

These issues have been addressed by both fields, resulting in what seems to be broad agreement with respect to the capabilities that proposed models of natural and artificial cognitive systems should exhibit - capabilities such as categorization, understanding, action selection, problem-solving, knowledge representation, joint action, communication, emotion and adaptation [3, 4, 5, 6]. There is also common support for approaches that stress the importance of embodiment and situatedness [7, 8]; in particular, there is growing interest in the potential of *enactivism* [9, 10] to provide a new perspective for cognitive systems [11, 12, 13, 14].

1.1 Enactivism

The term 'enactive' was proposed by the Chilean biologist Francisco Varela (in the Afterword of his seminal book *The Tree of Knowledge* [9], co-authored with his close colleague Humberto Maturana) to designate a novel view of 'knowledge' in living systems. In Maturana and Varela's framework, living beings are defined as autonomous, continually self-producing ('autopoietic') unities, where behaviour is seen as a joint construct of an organism and its environment. Cognition is said to be based on the operational closure of an organism's nervous system and is viewed as effective action (as expressed in their well-known simile "*knowing is doing*" [9] p248).

Varela made it clear that the enactive approach was not anti-representational; rather, that an organism and its environment were two sides of the same coin. The enactive concept of knowledge - "*what is known is brought forth*" ([9] p255) - is, therefore, interestingly different from the classical cognitive or connectionist views.

1.2 A way forward?

While enactivism appears to provide a compelling framework for future progress across the fields of cognitive science and cognitive systems, what is still missing is a common language for expressing many of the key concepts involved. This paper attempts to lay a tentative foundation in that direction by taking inspiration from arguably one of the more incidental aspects of Maturana and Varela's work - their use of a series of simple pictographic devices to illustrate various stages of organizational complexity for autopoietic unities (see Section 2).

This paper shows how Maturana and Varela's original diagrammatic approach may be extended to create a rich visual taxonomy for envisioning a wide range of enactive systems with different degrees of complexity - even for systems that are *not* autopoietic. Originally developed for teaching purposes, it is shown herein how such a diagrammatic language can serve to both clarify and unify a diverse array of relevant topics from a pan-theoretic perspective, arguing that such an approach could pave the way towards a deeper cross-disciplinary understanding of some of the complex concepts involved.

2. Maturana and Varela's Pictograms

2.1 First-order autopoietic unities

As mentioned in Section 1.1, the key to Maturana and Varela's thinking was the concept of *autopoiesis* as a fundamental self-regulating mechanism for maintaining the organizational integrity of a living organism¹ - referred to as a 'unity'. For example, a self-maintaining boundary (such as a membrane) serves to create a persistency of structure (such as a single-celled animal), which Maturana and Varela termed a 'first-order autopoietic unity'. Their idea was that such a unity functions with 'operational closure' and, therefore, defines that which is internal and that which is external in the context of continuous interaction between the unity and its surrounding environment. These fundamental concepts were illustrated (on p74 of [9]) using a simple pictogram (as shown in Figure 1).

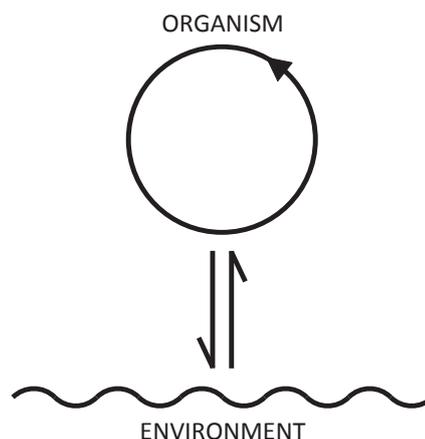


Figure 1. Maturana and Varela's pictogram for a self-maintaining first-order autopoietic unity and its coupling with its external environment (as illustrated on p74 of [9]). The circle represents an organism's boundary (for example, a membrane), the solid arrow on the circle represents the organism's dynamics (that is, its metabolism), the wavy line represents the organism's external environment and the double arrows represent the coupled interaction between the organism and its environment.

2.2 Second-order unities

Having defined (and illustrated) the essential properties of first-order autopoietic unities, Maturana and Varela went on to ask what would happen if two (or more) unities co-existed in the same neighbourhood. In this situation, non-destructive recurrent interaction between the unities would itself constitute a form of structural coupling, leading to two possible outcomes: the inclusion of one unity within the boundary of another (a *symbiotic* arrangement) or a mutual dependency between the individual unities (giving rise to a *metacellular* unity). Maturana and Varela termed these assemblies 'second-order autopoietic uni-

¹ Note that 'autopoiesis' (the maintenance of system organization) is related to, but not the same as, 'homeostasis' (the maintenance of system state).

ties² and again illustrated the concepts using simple pictograms (as shown in Figure 2).

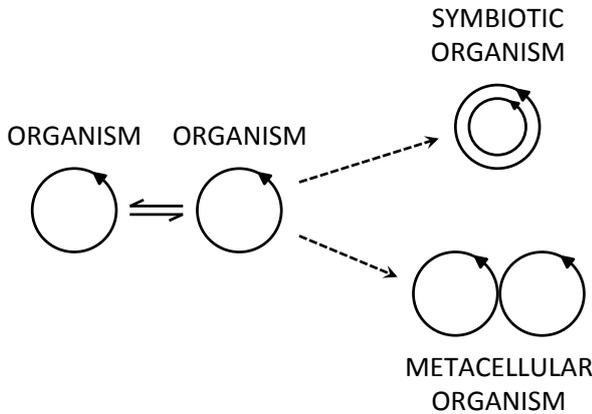


Figure 2. Maturana and Varela's depiction of second-order symbiotic and metacellular autopoietic unities formed from structural coupling between first-order unities

Maturana and Varela observed that the natural world contains a vast array of complex organisms whose organizational structures may be viewed as the result of structural coupling between first-order autopoietic unities. However, they noted that one particular multicellular arrangement - a *neural* assembly - was of special interest due to a neuron's ability to create long-range dependencies by coupling non-adjacent cells. As a result, they defined an organism that possessed a central nervous system as a 'second-order cognitive unity', illustrating it by using a simple modification to the original first-order pictogram: see Figure 3.

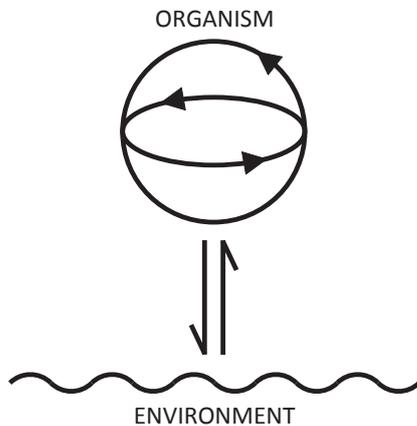


Figure 3. Maturana and Varela's depiction of a second-order cognitive unity. The outer circle represents an organism's boundary (as before) and the inner ellipse represents its central nervous system (with solid arrows to signify that Maturana and Varela considered that it also functioned with operational closure).

2.3 Third-order systems

Maturana and Varela employed their notion of second-order cognitive unities to discuss the behaviours of a wide

range of organisms, from the relatively simple (such as hydra) to the very complex (such as human beings). They noted that the possession of a nervous system opened up radical new forms of structural coupling between organisms, illustrating such 'third-order coupling' as shown in Figure 4.

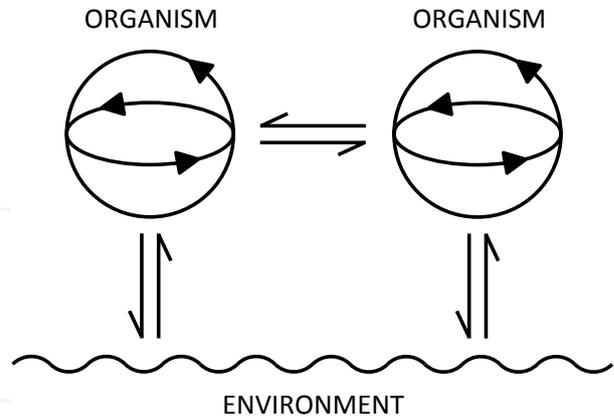


Figure 4. Maturana and Varela's depiction of third-order coupling between cognitive unities

Maturana and Varela observed that third-order coupling between cognitive unities gives rise to an impressive array of advanced interactive behaviours, such as coordination, communication, social systems and language. However, they did not present any further developments of their pictograms to illustrate such arrangements.

2.4 Impact and wider implications

Maturana and Varela's simple pictograms have enjoyed some modest popularity as illustrative devices. Maturana himself employed several variants of the pictogram shown in Figure 1 to represent various physiological and behavioural domains in a paper on structural drift in evolutionary development [15]. Likewise, a few authors have used the pictograms to illustrate the self-organized autonomous nature of a cognitive system [3]; they have even found their way into textbooks on management techniques [16]. However, the diagrams have essentially remained as presented in *The Tree of Knowledge* [9], with only limited development having subsequently taken place.

Nevertheless, by virtue of their continued usage, Maturana and Varela's pictograms have shown themselves to be compelling representational devices for capturing the essence of cognitive entities. Notwithstanding their extreme simplicity (almost trivializing the issues involved), the pictograms succeed in focusing attention on aspects of organization and behaviour that are relevant to both cognitive science and cognitive systems. However, there are many critical concepts that the original diagrams do not

² Maturana and Varela prevaricated as to whether metacellular systems themselves constitute first-order autopoietic systems. However, as they affirmed that such configurations do have operational closure, they proposed that the pictogram shown in Figure 1 could also be used to depict a cellular or a multicellular autopoietic unity.

capture. Maturana and Varela's pictograms, therefore, represent an inspiring starting point for the creation of a much richer visual language.

3. Towards an Enhanced Visual Language

3.1 Natural versus artificial systems

The concept of a 'unity' - an autonomous organizational system structure - is central to the ideas presented thus far, while Maturana and Varela's use of smooth circles and ellipses to represent *natural* systems may be appropriately complemented by the use of angular squares and rectangles to represent *artificial* systems - *machines* (see Figure 5). This apparently simple extension immediately offers an important insight: the circles depicting unities in Figures 1-4 incorporate solid arrows to signify that the natural system organization is autopoietic (self-maintaining). Whilst many artificial systems aim to achieve autonomy using homeostasis [17, 18], very few machines are autopoietic, i.e., capable of *self-repair* [19, 20], which is a potential requirement for a fully autonomous system that is often overlooked. Most existing machines are 'allopoietic' and thus do not warrant the inclusion of the arrow. Hence, the power of the pictographic representations to isolate and identify important system characteristics is immediately apparent.

Following the same line of argument, the embedding of a central nervous system within a living organism (as illustrated in Figure 3) inevitably lends itself to an analogous arrangement, whereby an artificial neural network (or some other artificial means of achieving cognition) might be embedded within a machine body (as shown in Figure 5c).

A 'cognitive robot' would, of course, be an example of such an artificial cognitive unity. However, it is important to appreciate that the ability to draw such an object does not imply that it exists. Rather, the pictograms clarify the characteristics that would need to be exhibited by an artificial system for it to be properly termed an embodied cognitive robot [21, 22].

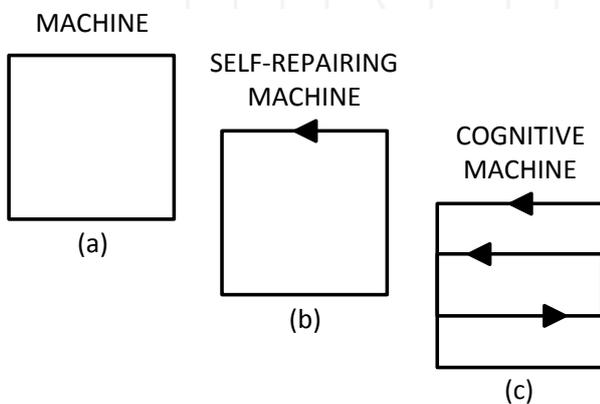


Figure 5. Proposed pictograms for artificial systems of differing complexity: (a) a machine, (b) a machine that is capable of self-repair and (c) a cognitive machine (also capable of self-repair)

3.2 Hybrid systems

Having established the basic visual vocabulary of using smooth circles and ellipses to depict natural systems, and angular squares and rectangles to depict artificial systems, it is immediately possible to conceive of *hybrid* configurations in which natural and artificial systems are combined in various symbiotic ways. Electromechanical prostheses for living organisms (such as a false limb or a replacement organ) are, of course, well-established artefacts, but the inverse idea of combining living tissue with a machine is much more futuristic [23]. Indeed, one only has to turn to the genre of science fiction to discover an astonishing array of symbiotic possibilities. For example, a cognitive machine embedded within a living host (such as the Borg in *Star Trek*) or a cognitive organism embedded in a machine body (such as a Cyberman from the BBC TV series *Doctor Who*) are both examples of a 'cyborg'. Likewise, it is possible to envisage the possibility of neural prostheses in either direction: artificial neural structures embedded in a living brain (to provide extra memory, for example) or natural neural structures embedded in an artificial brain. Some of these hybrid arrangements are illustrated in Figure 6.

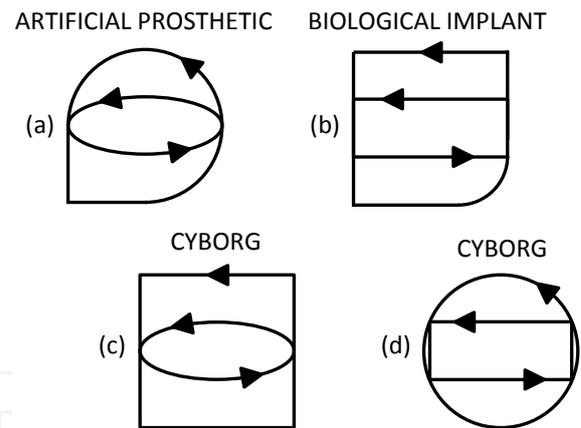


Figure 6. Proposed pictograms for a range of symbiotic natural and artificial hybrid systems: (a) a biological cognitive system with an artificial prosthetic, (b) a cognitive machine with a biological implant, (c) a biological nervous system embedded in a machine body and (d) an artificial cognitive system embedded in a biological body

Many of the configurations described in the foregoing are indeed in the realms of science fiction. Nevertheless, it is clear that even these first primitive steps towards the development of a visual language are successfully laying bare the conceptual space of possibilities in a clear and concise manner. Indeed, even stepping away from the futuristic scenarios posited by science fiction and back to the reality of contemporary media, the approach being taken here again points towards potentially important insights. For example, the reality of *Star Trek* is that an advanced humanoid robot, such as Mr Data, is, of course, portrayed by a human actor: *Brent Spiner*. In this case, the actor is passing himself off as a cognitive machine, which has the outward appearance of a human being but with

non-human skin tone and behaviours. This raises two related issues: (i) the notion that a 'costume' (or 'skin') can be used to change the appearance of an entity and (ii) the concept of an appearance that is a hybrid between a machine and a living system. Pictographically, the first can be represented by the addition of an outer layer with a gap signifying that it is wrapped around the entity as a covering, while the second can be represented by a shape that is part way between a circle and a square. These hybrid configurations are shown in Figure 7.

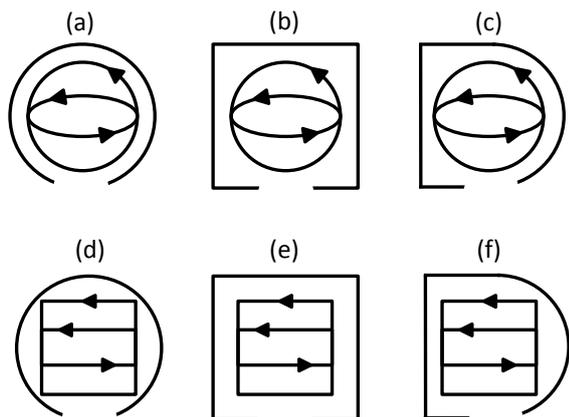


Figure 7. Proposed pictograms for a range of hybrid arrangements involving 'costumes' (or 'skins'): an actor portraying themselves as (a) another person, (b) a machine and (c) an android; and a cognitive machine configured to portray itself as (d) a living system, (e) a different machine and (f) an android

These extensions to the pictographic language reveal a combinatorial set of possible symbiotic configurations in which human beings portray themselves as other human beings (such as Dustin Hoffman's character in the film *Tootsie*), as a machine (such as the wizard in the film *The Wizard of Oz*) or as an android robot (such as Kryton in the TV series *Red Dwarf*), while machines portray themselves as facsimiles of living organisms (such as a dog in the case of Sony's *Aibo*), as androids (such as Honda's *Azimo* or the RobotCub Consortium's *iCub* [24]) or as an actual human being (such as *Geminoid F* [25]).

3.3 Two-way coupling

After the notion of an autopoietic unity, arguably the second most important aspect of Maturana and Varela's basic pictograms is the connectivity between a unity and its external environment (which includes other unities). In particular, Maturana and Varela point out the importance of continuous recurrent interactions in establishing a structural coupling between unities or between a unity and its environment. They illustrated such interactions by using arrows representing two-way reciprocal perturbations (as in Figures 1-4), but they did not tease out the topological implications of such connectivity.

³ Note that these principles are equally applicable to the internal dynamics of multicellular organisms, not just to the external dynamics of communities of cognitive agents.

Clearly, the collective behaviour of a community of unities depends on the connectivity between the individual unities. If all unities are coupled to one central unity (as in an insect colony), then the emergent behaviour is going to be very different from that which arises in the situation where unities are coupled to their nearest neighbours (as in a flock of birds). Such alternative configurations are relatively easily expressed in diagrammatic form; Figure 8 illustrates a selection of basic topological arrangements, each of which will have different consequences for the emergent behaviour.

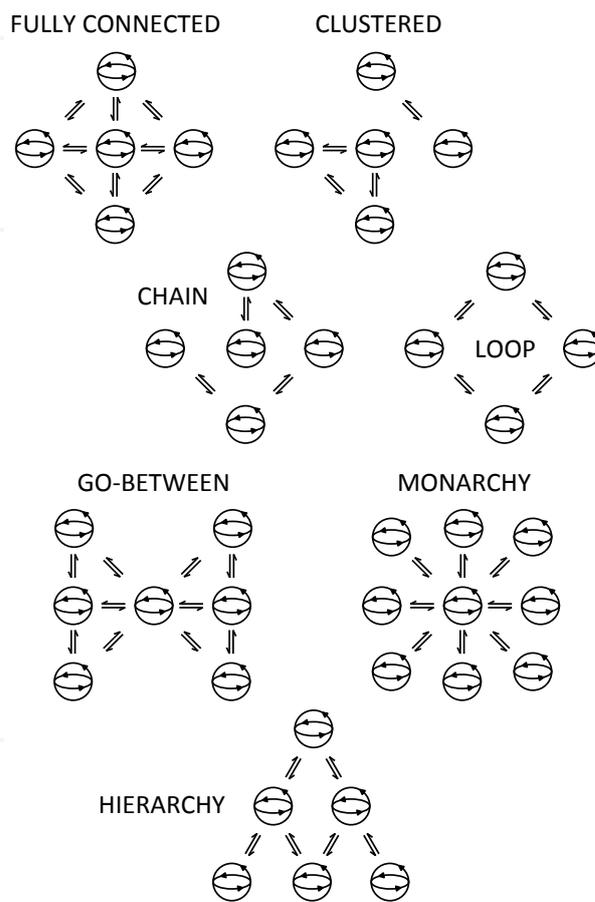


Figure 8. Illustration of alternative topologies for connectivity within a community of cognitive agents

The alternative network topologies illustrated in Figure 8 are clearly helpful in clarifying the differences that can emerge in the collective behaviours of swarms, flocks, groups, crowds and societies³. Of course, many other configurations are possible, including connectivity that changes dynamically. The key outcome here is that the proposed visual framework clearly provides a mechanism for focusing attention on the implications of alternative topological arrangements, while the pictograms help to clarify the issues involved, whether it is for modelling the

behaviour of natural systems or for instantiating required behaviour in artificial systems (for example, in the field of swarm robotics [26, 27]).

3.4 One-way information flow

The foregoing sections have treated interactivity as a two-way structural *coupling* between unities (or between unities and their environments); for a cognitive unity, these are essentially 'perception-action loops'. However, in reality, such a reciprocal arrangement is a particular case; a more general approach should accommodate the possibility of one-way connections (that is, potentially information-bearing *signals*), which are easily represented diagrammatically through the use of a single arrow in place of the double arrow used hitherto. As a consequence of this simple step, a distinction can now be made between interactive behaviours that are often the subject of much confusion - those which emerge from the one-way flow of information (which facilitate phenomena that may be characterized as the consequence of 'stimulus-response' or 'cause-and-effect' conditions), and those which emerge from two-way coupling (which exhibit phenomena such as 'synchrony', 'coordination' or 'co-action') - see Figure 9.

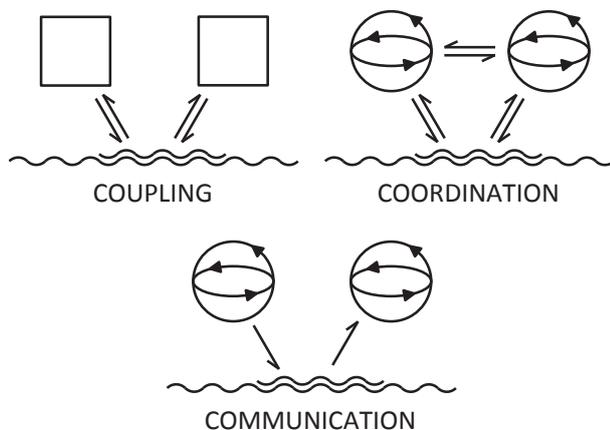


Figure 9. Illustration of the difference between the alignment consequences of two-way connectivity (such as the synchronization that emerges between metronomes, which are coupled through a shared physical environment, and coordinated joint action between two agents) and the communicative consequences of one-way connectivity (such as ants laying down pheromone trails in a shared physical environment)

Acknowledgement of the possibility of a one-way information flow leads to another valuable insight: the difference between the active/intentional sending or receiving of a signal and the passive/unintentional sending or receiving of a signal. For example, all unities radiate information by giving off heat, reflecting light, making sound, leaving deposits, disturbing the environment and so on. These are passive/unintentional signals that, nevertheless, may have some significance for another unity; for example, a predator may use such information to target its prey or a prey may use such information to avoid a predator. On the other

hand, a unity may actively generate such signals specifically in order to influence the behaviour of another unity or it may actively seek to receive such signals specifically in order to maintain awareness of salient activities. Both of these can be viewed as *intentional*, with the former corresponding to 'communicative' behaviour and the latter corresponding to 'attentional' behaviour. All of these important, yet subtle, distinctions may be readily captured in the visual language being developed here through the simple expedient of depicting active/intentional signals using an arrow drawn with a *solid* line, as well as depicting passive/unintentional signals using an arrow drawn with a *dotted* line - see Figure 10.

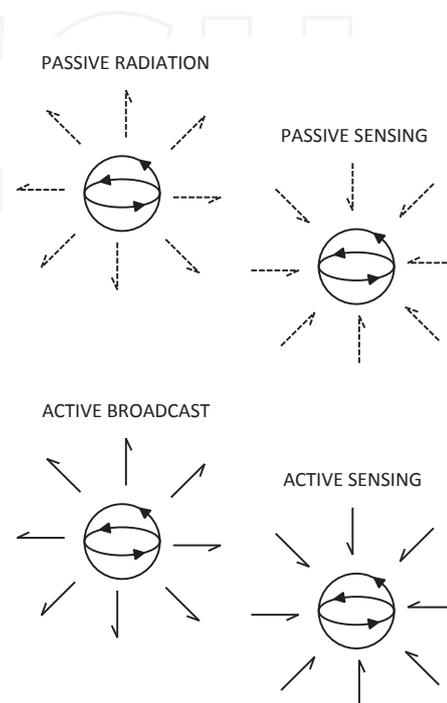


Figure 10. Examples of omnidirectional one-way signals: passive radiation (for example, a chemical pheromone being given off, which has the effect of attracting a mate), passive sensing (for example, a plant's sensitivity to ambient temperature), active broadcast (for example, a bird's alarm call) and active sensing (for example, a fish monitoring an electric field)

Furthermore, since an arrow naturally suggests directionality, it is straightforward to capture the difference between omnidirectional broadcasting/sensing and directional monocasting/sensing - see Figure 11. Of course, just because a unity employs directional signalling, there is no implication that there is an identified receiver. Likewise, just because a unity is employing directional receiving, there is no implication that there is an identified sender. Such intentional behaviours are indeed communicative or attentional, but the establishment of a connection between sender and receiver is not guaranteed. For example, a sender's behaviour might be intentional and directed, but the intended recipient may not be in range (and *vice versa*).

If the signalling behaviour of a sender is intended to draw the attention of a particular receiver (or *vice versa*), then the signals can be said to be *ostensive* and the arrows ought to

connect sender and receiver. This is a key point that will be returned to in Section 3.9 when discussing language-based interaction.

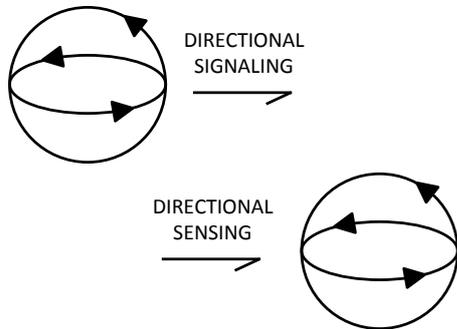


Figure 11. Examples of one-way signals: directional signalling (for example, a physical push) and directional sensing (for example, an animal’s active use of gaze)

3.5 Managing information flow

The foregoing takes the perspective of the sending or receiving unity. This means that, in a community of multiple unities, any particular signal may be active or passive depending on whether one takes the sender’s or the receiver’s point of view. For example, an active intentional signal sent by one unity may or may not be picked up by the passive sensing of another unity (as in the case where a mating call may be treated as mere background noise by receivers from a different species). Likewise, passive radiation by one unity may be actively sensed by another unity (as in the case of a predator seeking a prey). These alternative configurations may be captured by the appropriate use of solid and dotted arrows.

A number of other interesting situations emerge from this analysis. For example, a unity might actively minimize its radiated signals (such as a predator or prey avoiding being seen or minimizing the generation of sounds that would give away its position). In a similar vein, a unity might passively mask its radiated signals (as in the fixed camouflage patterns of many types of moth) or actively mask its radiated signals (as in the dynamic camouflage exhibited by cephalopods). Two of these possibilities are illustrated in Figure 12.

3.6 Intentional acts

One refinement that has potentially wide-ranging implications is to use the pictographic approach being developed here to tease apart different types of intentional behaviour [28]. In particular, it would seem to be important to be able to make a distinction between behaviour that is intended to change the state of a unity’s external environment (which could include other unities) from behaviour that is intended to change the state of the unity itself. For example, an organism might *probe* its environment specifically in order to determine the presence or absence of some desired feature (such as food). In this case, the intention is to gain information, rather than alter the world.

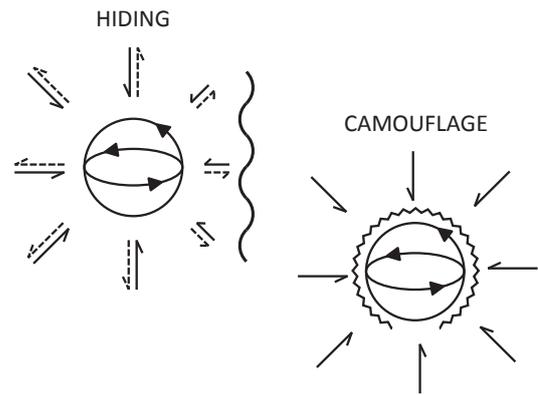


Figure 12. Illustration of individuals attempting to minimize radiated information while remaining alert to potential threats by (i) *hiding* behind part of the environment (with a consequent loss of sensing ability) or (ii) passing themselves off as part of their environment using *camouflage*

Likewise, it would seem to be important to be able to make a distinction between behaviour that is intended to change the state of the environment (or another unity) *directly* versus behaviour that is intended to change the state of the environment (or another unity) *indirectly*. For example, a unity could change the state of another unity by acting upon the environment or change the state of the environment by acting upon another unity. In both these cases, the intention is *mediated*.

These different types of behaviour may be described as ‘direct’, ‘reflective’ and ‘indirect’ intentional acts (which could be actions or signals); the proposed pictographic representations are illustrated in Figure 13.

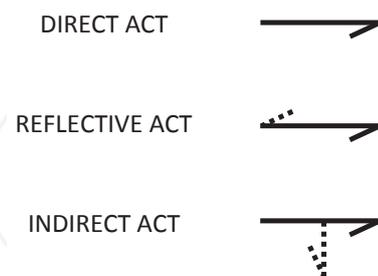


Figure 13. Illustration of three classes of intentional act: direct, reflective/probing and indirect/mediating

The consequence of creating these diagrammatic representations is that they enable distinctions to be made between activities, such as digging a hole or moving a rock (both examples of direct action on the physical environment), investigating/exploring an area (an example of reflective action on the physical environment intended to provide information to the exploring unity) or marking out territory (a communicative action on the physical environment intended to have an effect on another unity). The latter is particularly interesting, since this includes ‘stigmergy’ (for example, a method used by ants to influence each other’s behaviour by the laying down of pheromone trails in the

environment) [29]. All three configurations are illustrated in Figure 14.

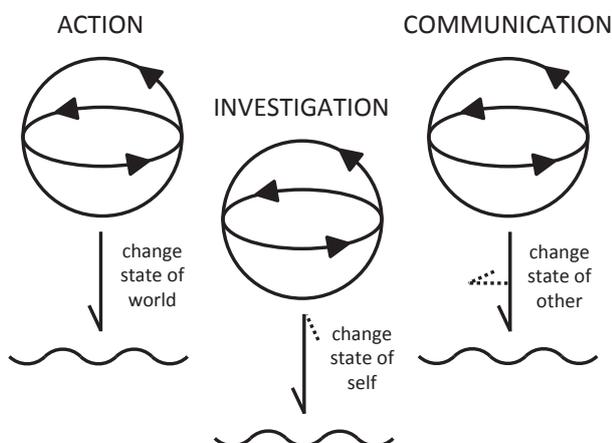


Figure 14. Illustration of the implications of the three types of intentional behaviour between a unity and its external environment: direct *action*, reflective *investigation* and indirect *communication*

3.7 Internal representations

One of the more controversial topics in cognitive science/systems is the issue of 'representation' - that is, whether or not information is retained inside a cognitive unity that, in some sense, mirrors the information outside the unity. For example, for an agent to navigate successfully in the world, the representationalist view would be that the agent must have access to an accurate *model* of its environment (such as a map). This is in stark contrast to the non-representationalist position that successful navigation (such as obstacle avoidance) is an emergent outcome of the interaction between an agent and its environment [30].

This debate came to a head in the field of Artificial Intelligence in the late 1980s with the appearance of explicitly non-representational 'behaviour-based robotics' [31, 32] as a practical⁴ alternative to the symbolic approach that was typical of 'GOFAI' (Good Old Fashioned AI) [33]. The issue is also central to enactivism, since enaction is founded on the key notion that organisms do not simply transform information received from their environment into internal representations (and *vice versa*). Rather, enactivism proposes a perspective in which organized behaviour - and, in turn, meaning - emerges from the dynamic coupling between organisms and their environments (including other organisms) [34].

It was already recognized in Section 3.3 that alternative coupling topologies give rise to different dynamical behaviours. So the issue here is not to question or undermine the enactive stance on representations, but simply to

adopt the position (i) that information correlating with properties of the external environment (sensorimotor traces, for example) may be retained within the structure of a unity, (ii) that such information can be regarded as a form of 'memory', (iii) that such information could act as a *prior* on the dynamics of interaction (that is, it might exert an influence on the attractor landscape) and (iv) that such information could be of some ecological value to the unity. As an example of the latter, sensorimotor information retained in memory may be exploited by a unity for various practical purposes, such as perception (the recognition of previously encountered sensorimotor events and scenarios), action (the efficient re-use of learnt sensorimotor behaviours) and prediction (the planning and optimization of future sensorimotor states prior to their enaction). These are all examples of *inference*.

Given this perspective, the degree to which perception, action and prediction take place, and the consequences for the coupled behaviour of a unity, is very much conditional on the fidelity and depth of the information that is retained: the larger and more detailed the memory, the higher the potential for successful outcomes. For example, even a simple spatiotemporal memory of some aspects of the external environment could, depending on its scope, facilitate a modest ability to recognize previously encountered locations and situations. Likewise, it would permit interaction with, and navigation of, the environment, as well as an ability to anticipate its dynamics (which could be crucial in situations where progress towards an intended goal cannot be readily evaluated).

In terms of the visual language being developed here, it is proposed that information that is retained within the cognitive unity is simply depicted as a pictogram contained within a unity's central nervous system. Hence, spatiotemporal memory can be illustrated by placing the pictogram for a unity's external environment (a wavy line) inside a unity's cognitive loop. This is illustrated in Figure 15a.

Beyond simple spatiotemporal memory, the next level of representational complexity would seem to be the decomposition (in memory) of the external environment into independent and semi-independent entities. This would facilitate an ability to distinguish between different objects in the environment, including other unities. It would also facilitate an ability to *emulate* the directly observable surface behaviours⁵ of other unities, thereby opening up the possibility of recognizing their identities, as well as anticipating their behaviours (albeit to a degree limited by the surface representation). The main advantage would be a dramatic increase in a unity's ability to generalize due to the combinatorial properties of the representational state-space arising from the decomposition.

⁴ Interestingly, the world's most successful commercial robot is iRobot's Roomba robot vacuum cleaner, which employs a non-representational behaviour-based 'subsumption' architecture.

⁵ A surface representation means that there is no explicit information relating to inferred hidden states or variables.

The ability to emulate the surface behaviour of another unity may be depicted by placing the pictogram for a second-order unity (see Section 2.2) inside a unity's cognitive loop. Furthermore, in order to make it clear that this is a representation of an external unity (that is, a representation of 'other' not 'self'), the inserted pictogram may be shown with dotted rather than solid lines. This is illustrated in Figure 15b.

Following on from emulating the surface behaviours of other unities, the third stage of representational complexity would seem to be the imputation of their internal states. In particular, this would relate to the attribution of motives and intentions to other unities - taking an 'intentional stance' [35]. It would also encompass the inference of affective states [36]. Again, such a step-up in representational complexity would confer significant operational advantages (for example, the ability to perform action understanding by inferring the intentions of another unity).

Successful inference of another unity's internal states could be regarded as the manifestation of 'empathy', which is illustrated by placing the pictogram for a third-order unity (see Section 2.3) inside a unity's cognitive loop. Again, since this is a representation of 'other' not 'self', the inserted pictogram may be depicted with dotted lines, as illustrated in Figure 15c.

Stepping up from imputing hidden variables, such as intentions and affective states, the fourth stage of representational complexity would seem to be the ability of one unity to model another unity with the potential effect of having a different perspective on the world from itself - a facility commonly referred to as 'Theory of Mind' (ToM) [37]. In order to do this, a unity would need to be able to infer the *beliefs* of another unity [38]. Armed with this information (represented in memory), a unity would then have the potential to exploit it in order to update its own beliefs and satisfy its own needs, desires and intentions.

The key difference between this arrangement and the three previous levels of representational complexity is that, in this configuration, the information embedded in memory not only incorporates an estimate of another unity's internal states, but also includes the *context* in which the other unity exists. In other words, such a unity would be in a position to model another unity as inhabiting a different environment from itself. This is illustrated by placing the pictogram for a third-order unity *and its environment* inside a unity's cognitive loop, as shown in Figure 15d.

As a final step, it is clear that the stages of representational fidelity proposed thus far lead to the possibility of a *recursive* arrangement, in which a unity's model of another unity could itself include internal representations (and so on *ad infinitum*). For example, a unity could represent another unity as having a model of the first unity (as illustrated in Figure 15e). Such a configuration - commonly

referred to as 'recursive mind reading' - would have a dramatic effect on the efficiency of the coupling between unities.

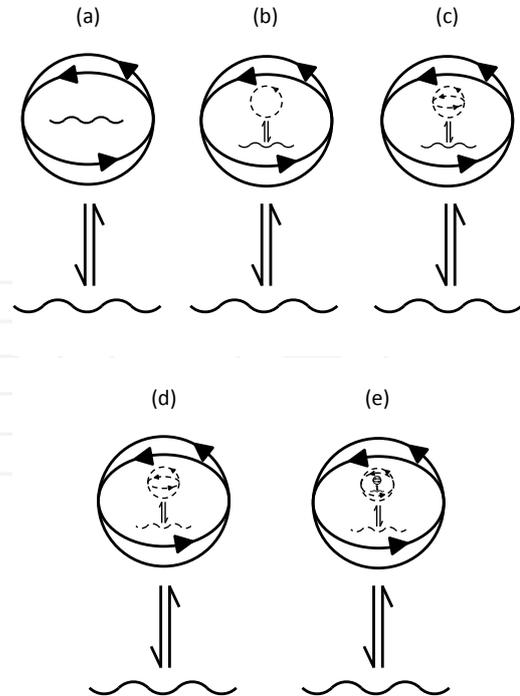


Figure 15. Illustration of internal representations of increasing fidelity: (a) *spatiotemporal memory*: representation of a unity's external environment; (b) *emulation*: representation of another unity's directly observable surface behaviour; (c) *empathy*: representation of another unity's internal intentions and/or affective states; (d) *ToM*: representation of another unity's different perspective; and (e) *recursive mind reading*: representation of another unity's model of the first unity. In each case, 'self' (and self's environment) is depicted using solid lines and 'other' (and other's environment) is depicted using broken lines.

It is interesting to appreciate that the series of pictograms shown in Figure 15 is not simply the outcome of the preceding arguments. Rather, the constraints imposed by the developing visual language help significantly in clarifying the logical progression of concepts that has been developed here. The rather satisfying outcome is that the pictograms illustrate, in a succinct form, how and why each step-up in representational complexity helps make a unity's external environment easier to understand, easier to navigate and easier to predict (where "easier" implies more accurate and/or more effective and/or faster). The pictographs, then, make it clear how the increasing representational complexity would confer significant adaptive benefit to natural living systems, as well as provide an effective blueprint for designing and implementing artificial cognitive agents, such as robots with different degrees of 'social intelligence' [39, 40].

3.8 Advanced interactivity

The possibility of different degrees of representational complexity (as discussed in the previous section) leads to

the realization that each level has implications, not only for the ability of one unity to understand and predict the behaviour of another, but also for the quality and type of coupling that could emerge between unities. For example, interacting unities, which possess information relating to each other's externally observable behaviour, would have the potential to exploit such information in order to optimize the coordination of their joint physical actions⁶. On the other hand, unities possessing representations of each other's internal intentions and/or affective states would be able to exploit such information in order to achieve more effective behavioural alignment through empathic coupling. Likewise, interacting unities possessing ToM would be able to exploit *deception* in their repertoire of behaviours, while those capable of recursive mind reading would be able to sustain high information rate interactions via much lower bandwidth sensorimotor channels (this is taken up in Section 3.9).

In general, shared representations facilitate coordinated behaviour⁷ [41, 42], so the quality and type of the shared information would have a direct impact on the resulting emergent behaviours. This, in turn, would depend on the fidelity and complexity of the internal representations: from physical interaction conditional on observable surface behaviours to empathic interaction conditional on inferred intentions and/or affective states, to ToM-based interaction conditional on inferred beliefs and to interaction conditional on recursive mind reading. Similarly, the degree to which behaviours such as *alignment* and *learning through imitation* are successful will be conditional on the level and fidelity of the shared representations [43, 44, 45]. Indeed, the notion that shared representations capture aspects of joint histories and mutual experiences has already been shown to be important in human-robot interaction [46, 47].

Each of these scenarios may be captured by suitable pairings of the relevant pictograms shown in Figure 15 (and their machine counterparts).

Interestingly, the developing visual language not only illuminates the possibility of different degrees of coupling between unities, but it also leads to an understanding that the most advanced forms of interactivity involve recursive mind reading between matched conspecifics⁸. Indeed, it can be seen that each level represents a form of *predictive coding* [48, 49], with recursive mind reading providing the highest quality predictors and, in turn, facilitating the highest information rate coupling, given a particular set of low information rate sensorimotor channels. Also, the pictograms are able to clarify that senders and receivers may employ directed intentional/attentional behaviour (see Section 3.4), which means that the highest level of

coupling would seem to be ostensive-inferential communicative interaction incorporating recursive mind reading - this is a definition of *language* [50]!

3.9 Linguaging

The foregoing section makes a compelling argument (based on the developing pictographic taxonomy) that ostensive-inferential coupling, founded on recursive mind reading, facilitates coordinated behaviour between matched unities, which can be reasonably termed 'linguaging' [51, 52]. Indeed, the discussion thus far ties together a variety of contemporary perspectives on language as an interaction system [53] (especially in human-robot interaction [54]), the importance of ToM and predictive models for language behaviour [55, 56, 57], and the power of a shared narrative [58].

This general idea can now be linked with the concepts developed in Section 3.3 with regard to direct, indirect and reflective action to create an interesting connection with speech act theory [59]. *Direct* linguaging is equivalent to a 'declarative' speech act (for example, "*Here is some food.*": an informative linguistic action), *reflective* linguaging corresponds to an 'interrogative' speech act (for example, "*Where is the food?*": a linguistic action that is intended to elicit an informative linguistic or non-linguistic response) and *indirect* linguaging corresponds to an 'imperative' speech act (for example, "*Get me some food!*": a linguistic action that is intended to elicit an appropriate physical action).

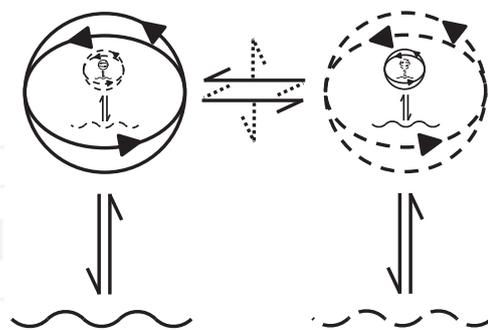


Figure 16. Illustration of conversational interaction (dialogue) using a combination of all three forms of ostensive behaviour (illustrated in Figure 14) to establish language-based coupling between two interlocutors. One unity (and its environment) is depicted using solid lines and the other unity (and its environment) is depicted using broken lines. As can be seen, communicative interaction is founded on two-way recursive mind reading.

In practice, all three forms of speech act may be employed in a mutual two-way communicative coupling between linguistically-enabled unities. The resulting behaviour is,

⁶ Note that joint behaviour could be cooperative or competitive.

⁷ This does not imply that such optimization can only be achieved through explicit cognitive processes. It could also arise from the conditioning effect of such information on the location of the attractors in the state-space of the dynamically coupled system.

⁸ It is assumed here that participants in an interaction are matched in their abilities and that shared representations are balanced in some sense. In general, this is not necessarily true - especially when living systems interact with artificial systems. The implications of mismatched abilities are discussed in Section 3.10.

of course, commonly referred to as 'dialogue' (conversational interaction) and it is possible to invoke a combined pictogram to reflect the richness of such coupling between interlocutors - see Figure 16. Furthermore, because of its enactive roots, the pictographic representation emphasizes dialogue as distributed sense-making [60, 61] across a dynamical system [62] (in contrast to the more traditional view of dialogue as message-passing within a strict turn-taking framework).

3.10. Human-animal-robot interaction

Much of the latter discussion has referred to unities without regard to whether they are natural or artificial (or hybrids), while the majority of the illustrations have used pictograms that suggest living organisms⁹. Moreover, in the sections on interactivity, there has been an implicit assumption that interacting unities are matched in their structures and capabilities. The next step is to break away from this premise and consider situations where interactivity occurs between *mismatched* unities, such as between humans and animals, or between natural unities and artificial agents.

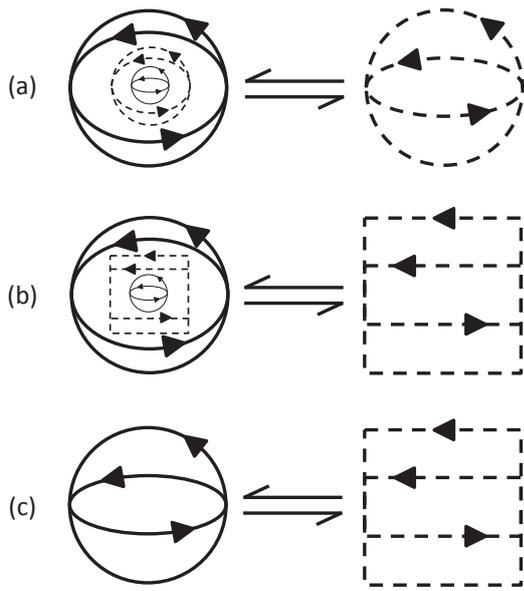


Figure 17. Illustration of the coupling between (a) humans and animals, (b) humans and robots, and (c) animals and robots. In each case, the pictograms make it clear that the mismatch in structures and capabilities will have important implications for the nature and type of interactions that can take place. In these particular examples, the animals and robots are depicted as *not* possessing a ToM (although, for animals, this is a controversial topic [63, 64]).

Several of these mismatched conditions are illustrated in Figure 17. As can be seen, the pictographic language developed thus far is easily extended to cover the interactions between humans and animals, between humans and robots, and between animals and robots. What is interesting is that the implications of such mismatched scenarios

are immediately made clear; for example, as discussed in Section 3.7, a human being naturally ascribes intentions to an animal or a robot, even though they may in reality possess no such abilities. Likewise, animal-robot interaction is, by necessity, conditional on surface physical behaviours alone. Such mismatched situations go some way to explain the difficulties that are often encountered in human-robot interaction, while the pictograms make explicit some of the challenges that need to be addressed [65].

Of particular interest is whether the emerging visual taxonomy offers any interesting insights into language-based interaction between a human being and a cognitive robot. Based on the pictographic representations developed so far, Figure 18 illustrates the required configuration and, as can be seen, the implication is that such an arrangement will only function properly if *both* partners (the human and the robot) exploit ostensive-inferential recursive mind reading. Since such abilities are beyond the capabilities of state-of-the-art robots and autonomous agents, this might serve to explain the somewhat primitive nature of contemporary language-based human-robot interaction [66]. It also explains why it is not appropriate to simply take off-the-shelf speech technology components (automatic speech recognizers, speech synthesizers and dialogue systems) and merely interface them with a mobile robotic platform; the integration must be much deeper if it is to be effective [67].

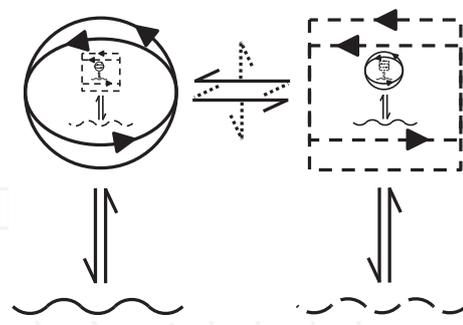


Figure 18. Illustration of language-based interaction between a human being and a cognitive robot. As in Figure 16, communicative interaction is founded on two-way recursive mind reading.

Indeed, the immense mismatch between the cognitive, behavioural and linguistic capabilities of human beings and those possessed by even the most advanced artificial cognitive systems, coupled with human beings' propensity to deploy a complex ToM perspective when dealing with interactive agents, underpins the importance of providing artificial systems with some degree of visual, vocal and behavioural coherence, if the interaction is to be at all effective [68, 69, 70]. Indeed, if such coherence is not the subject of careful and balanced design, then there is a real

⁹ It is hoped that the reader has been mentally mapping the pictograms onto the corresponding artificial forms and considering the implications for advanced robotic systems.

danger of inadvertently creating the conditions under which the human user will fall into the 'uncanny valley' [71] and reject the system, due to an inability to accommodate its anomalous behaviours [72, 73, 74].

3.11. Self-awareness

Finally (and with the risk of entering into difficult and controversial territory), it is possible to use the visual language being developed here to address the enigmatic issue of self-awareness: the sense of agency and consciousness [75, 76, 77]. For example, it has already been argued (in Section 3.7) that memory may be used to store representations of increasing degrees of complexity, which can contribute to increasing levels of efficiency in a unity's ability to interact with its external environment (including other unities). However, the representations thus far have been concerned with the fidelity of stored information pertaining to *other*, whereas the same arguments may be made with respect to the fidelity of stored information pertaining to *self*.

This raises an obvious philosophical question: if, by virtue of its existence, a unity already has access to information about itself, why (as is argued by enactive and radical embodied theories of cognition [78]) would it need to invoke an additional representation of itself? The answer, emerging from the insights created by the visual language being developed here (and tying in with contemporary views on the role of predictive processing in linking exteroceptive and interoceptive perceptions [79, 80]) is that such internal representations have *predictive* power. That is, a representation may be used to explore "What if?" scenarios prior to committing the organism to some, potentially catastrophic, action. Hence, a unity that has an ability to simulate itself would be in a position to *imagine* itself in different spatiotemporal contexts (that is, different times or places)¹⁰. Such a situation is illustrated in Figure 19a.

In addition, by following the line of argument developed in Section 3.7 with respect to recursive mind reading, it is possible to hypothesize a configuration in which the representation of self could itself contain a representation of self (and so on). Such an arrangement would facilitate an ability for a unity to think about itself thinking about itself thinking about... (and so on). This situation is illustrated in Figure 19b.

Turning now to the concepts developed in Section 3.9, the visual language suggests that it would be possible to invoke internal representations of self based on ostensive-inferential recursive mind reading. In other words, the pictograms make it clear that it would be possible to invoke language as a linearized abstracted internal communications channel with a high information rate between self and self. If such

an arrangement involves dialogue between self and simulated self, then this represents (literally) talking to oneself. If the configuration involves communication between simulated self and simulated self, then this would correspond to an internal dialogue¹¹. The latter configuration is illustrated in Figure 19c.

These perspectives on 'representation' align well with the latest thinking in cognitive science, where it is now being argued that "trying to suppress the notion of representation... is seriously misguided" [81].

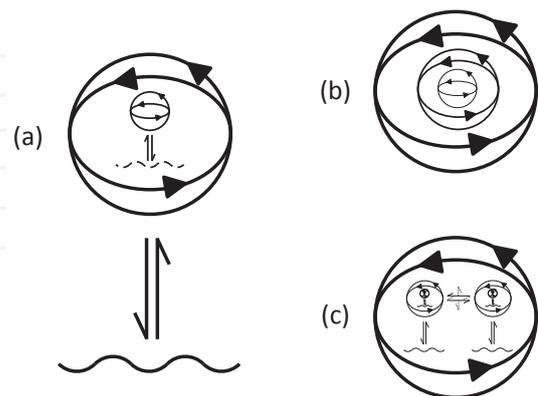


Figure 19. Illustration of different types of self-awareness: (a) a unity imagining itself in a different context, (b) a unity thinking about itself thinking about itself and (c) a unity holding an internal dialogue with itself

Finally, it is clear that all these configurations call for some kind of *workspace* in order for a unity to be able to run the 'What if?' simulations; this would seem to require a consolidated approach across different sensorimotor capabilities, if optimal solutions are to be discovered. Such a perspective appears to tie in closely with 'Global Workspace Theory' (GWT) [82] and, in turn, with contemporary theories concerning consciousness [83, 84].

4. Discussion

The invention of external aids, such as *visualization* to amplify our natural cognitive capacity to comprehend obscure and complex information, is a ubiquitous feature of human endeavour [85, 86]. In the same spirit, the pictographic language developed in this paper does appear to facilitate tantalizing glimpses into some of the most difficult challenges facing the fields of cognitive science and cognitive systems. In line with Maturana and Varela's original pictograms (Figures 1-4), the extensions presented here are remarkably simple in design. Nevertheless, potentially valuable insights emerge, precisely because of the clarity injected by the use of such straightforward graphical devices.

¹⁰ Indeed, a unity could even imagine itself as a different self.

¹¹ For example, mentally telling oneself to do something or creating an internal narrative account of self (a configuration that Maturana hypothesized as the seat of consciousness).

4.1 Related work

The approach presented here is reminiscent of von Neumann's attempt to introduce a logic of self-reproduction [87], as well as more recent attempts to taxonomize the structure of self-sustaining systems in diagrammatic form, such as the 'operator hierarchy' developed by Jagers op Akkerhuis [88]. Although superficially similar to the scheme presented here, Jagers op Akkerhuis' pictographic representations are primarily concerned with capturing the *evolution* of living systems, particularly at the cellular level. His pictograms do not offer any special insights into interactive behaviour, especially at the cognitive level or with respect to empathy, ToM or language. However his concept of a 'memon' (characterized by an organism's possession of a neural network) as the most complex organizational 'closure'¹² is interesting, as he uses his scheme to predict that technical memons (as opposed to animal memons) are likely to provide the next step in the evolution of life on our planet: "*a conclusion that tears down the walls between biology and robotics and places artificial intelligence at the heart of evolution on earth*" [89].

Furthermore, the informational perspective developed here appears to be highly compatible with the notion of emergent representation as introduced in the 'Interactivist' framework [90]. In particular, Bickhard [91] proposed a number of (unordered) levels of representation, several of which map directly onto the scheme proposed here. However, the present analysis reveals a number of interesting issues not addressed by Bickhard. For example, Bickhard nominates intentionality as the first level of representation, but does not take into consideration the position adopted here that the intentional stance only makes sense in relation to some aspect of the external environment. Moreover, somewhat surprisingly, Bickhard does not address ToM.

In the field of social robotics, Dautenhahn has employed pictographic illustrations in order to convey some of the core concepts in social intelligence, problem-solving and empathy [39, 40]. Likewise, Ziemke [92] and Vernon [93] provide somewhat visual approaches to modelling 'higher level' cognition. However, none of these visualization schemes aspires to provide a comprehensive pictographic language of the form described in this paper.

4.2 Use in teaching and learning

As well as being an exercise in attempting to understand complex issues in the organization and behaviour of living and non-living systems, it is important to note that the pictographic language presented in this paper was originally developed in order to clarify such issues in the context of teaching and learning. The pictograms were initially

constructed for a university second-year undergraduate course on *bio-inspired computing and robotics* in a high-ranking computer science department, specifically in order to help the students assimilate key concepts introduced in four hours of lectures on 'cognition' and 'interaction'. The course has been running in this form for several years and feedback from the students has been very positive. In particular, the course attendees are appreciative of the integrative nature of the visual language, allowing a wide range of topics to be linked together within a single conceptual framework.

This positive experience in teaching and learning underlines the value of providing a systematic classification of cognitive systems. Even though, by its simplifying nature, such a taxonomy inevitably has some limitations (addressed below), it can nevertheless provide students and researchers with a clear frame of reference for understanding such complex systems.

4.3 Limitations

Of course, it is only possible to go so far in capturing complex conceptual issues in simple pictographic representations. Such an approach is, by its very nature, bound to result in crude approximations, high-level abstractions and the potential omission of important details. There are real dangers that the intuitions implicit in the pictograms may be ill-founded or plain wrong, thereby causing confusion and misunderstanding rather than clarification and insight¹³.

It is also not clear to what extent such an approach might help in designing the specifics of any particular system. For example, the internal representations depicted in Section 3.7 onwards sidestep precise details about how the different representations impact on perception and action. Likewise, the pictographs illustrated in Figures 16 and 18 say little about the special 'particulate' (combinatorial) structure of speech and language [94].

Taking a more general perspective, these potential difficulties are well-established problems in the philosophy of science [95]; there are always open questions regarding (i) the explanatory power of a selected scientific model, (ii) the risks associated with taking a reductionist stance and (iii) the assumptions that have to be made in order to facilitate progress. Aware of these issues, the approach taken here has attempted to be pan-theoretic, drawing on a wide cross-section of relevant literature and addressing head-on some of the more controversial issues (such as cognitive representations, which are discussed in Section 3.7). Care has also been taken to position the visual language, such that it is sufficiently general in order to avoid any major conceptual pitfalls whilst being sufficiently precise that it captures

¹² A 'closure' in Jagers op Akkerhuis' operator hierarchy is somewhat analogous to a 'unity' in Maturana and Varela's scheme.

¹³ There are even potential problems with Maturana and Varela's original pictograms; for example, it has been debated whether a nervous system is actually autopoietic and, therefore, deserving of the two arrows depicted on the inner ellipse in Figure 3.

specific instances of complex configurations. Indeed, the fact that a number of important insights has emerged, as a direct consequence of designing the various pictograms, provides evidence that the scheme does appear to provide a parsimonious account of a wide range of important issues without committing too many sins.

4.4 Implications for advanced robotic systems

Although relatively few examples have been given of non-living systems (due to the development of the pictographic taxonomy taking primary inspiration from the behaviours of living organisms), the concepts portrayed in Sections 3.3 to 3.9 are of equal relevance to robotic systems. For example, the different arrangements for connectivity within a community of cognitive agents portrayed in Figure 8 would map directly onto equivalent topologies in swarm robotics (and illustrated with the appropriate machine-like pictograms).

In a similar way, the different types of active/passive signalling illustrated in Figure 10 and Figure 11, if redrawn for non-living systems, would serve to clarify some of the issues that need to be considered when designing, implementing and operating a robot's sensors and actuators. Likewise, the different classes of intentional behaviour portrayed in Figure 13 (and shown in action in Figure 14) also apply directly to robotic systems and focus awareness on the degree of autonomy that such systems actually have. Finally, the depictions in Figure 15 and Figure 19 of the value and power of internal representations of increasing fidelity (including self-awareness) have important implications for advanced robotic systems. For example, Winfield [96] has argued that self-awareness will lead to enhanced safety in physical human-robot interaction and, ultimately, toward ethical behaviour in autonomous robots.

Overall, the pictographic language developed here facilitates the visualization of a huge array of complex arrangements involving both living and non-living systems (and their environments). Its value lies in the ease with which key conceptual issues are brought to the fore: a feature that should prove immensely useful in the design and implementation of future advanced robotic systems. In particular, it should serve to raise pertinent issues in the mind of a robot system designer (for example, with respect to different aspects of biological agency/autonomy), as well as facilitate clearer thinking about the broad conceptual issues involved, notwithstanding the limitations discussed above.

4.5 Possible extensions

The previous sections have touched on a wide variety of contemporary topics in cognitive science and cognitive

systems; however, there are many aspects of the structure and behaviour of natural/artificial systems that have yet to be addressed. For example, there would be great value in teasing apart the different dimensions associated with the familiar, yet surprisingly ill-defined, notion of 'intelligence'. Useful starting points in this area could be the evolutionary perspective proposed by Mithin [97] or decomposition as outlined by Winfield [98]. However, it is unclear how such concepts might be depicted pictographically.

Likewise, the visual language developed thus far does not capture the morphological complexity of different unities. Clearly the number of degrees of freedom associated with the physical 'body' of a unity heavily determines its ability to interact with its external environment (including other unities), but a satisfactory way of handling this pictographically has yet to be worked out¹⁴. In addition, the current visual language does not distinguish between different sensorimotor modalities for interaction and communication.

Other topics worthy of investigation (in the context of designing suitable pictographic representations) are:

- *teleoperated robots* (that is, a person-machine metacellular organization in which one is in direct control of the other);
- *intelligent virtual agents* (that is, non-embodied agents such as an avatar or Apple's Siri);
- *emotion* (a representation of internal states, such as illustrating a location in the classic 3-D space comprising valence, arousal and dominance [99] and their relevance to human-robot interaction [100, 101]);
- *learning* (either ontogenetically during its own developmental life cycle [102] or phylogenetically as a community evolves over time [103]);
- *centralized versus distributed unities* (unlike natural systems, artificial systems and hybrids are not confined to local connectivity);
- *intrinsic motivations* (the representation of internal drives, needs, beliefs, desires and intentions [38, 104, 105]);
- *homeostatic control* (the regulatory mechanisms required to maintain preferred internal/external states [106, 107]); and
- *affordances* (the notion that the coupling between a unity and its environment is fundamentally conditional on the characteristics of both [108, 109]).

Finally, it has been suggested¹⁵ that there may be value in animating the pictograms to illustrate some of the dynam-

¹⁴ One possibility is to use the number of sides on a unity to indicate the number of degrees of freedom (on the basis that a living system has a huge number and hence warrants a circle), but this would not adequately capture the decomposition of a unity into coupled (symbiotic) elements.

¹⁵ by interested colleagues

ics. This could be a very valuable extension; for example, expanding upon Heider and Simmel's famous hand-crafted stop-motion cartoon in order to illustrate apparent intentionality between simple agents [110].

5. Conclusion

This paper has introduced a set of putative extensions to Maturana and Varela's original pictographic depictions of autopoietic unities to create a rich visual language for envisioning a wide range of enactive systems - natural or artificial - with different degrees of complexity. Originally developed for teaching purposes, it has been shown how such a diagrammatic taxonomy may be used to illustrate a remarkable range of advanced configurations (such as hybrid systems, swarm topologies, intentionality, communicative behaviour, empathy, ToM, languaging, human-animal-robot interaction and self-awareness).

As well as drawing on a large cross-section of relevant literature, the research has attempted to take a *pan-theoretic* perspective. As a consequence, a number of interesting insights has emerged; not only does the visual taxonomy reveal the paucity of many current approaches, but it also appears to provide a parsimonious depiction of a wide range of important research questions, many of which have implications for our understanding of cognition as well as the effective design and implementation of future cognitive systems (such as robots).

As a *language*, the pictographic taxonomy has expressive power, so there are many combinations of symbols whose implications are yet to be explored. It is hoped, therefore, that other researchers will not only find value in the existing framework, but will also explore its potential and contribute appropriate extensions and improvements.

The author has found the pictographic language very useful for teaching and learning within his own institution. However, the test of the utility of such a taxonomy is whether it is sufficiently intuitive, consistent and helpful in order for it to be embraced by the wider research community. To that end, a glossary of pictographic components is included as an addendum¹⁶. It is intended that this may be used as a 'quick reference' guide, in exactly the same way as a sheet of circuit symbols aids electronic systems design.

6. Acknowledgements

The author would like to thank Dr. Fred Cummins (University College Dublin), Prof. Alan Winfield (Bristol Robotics Laboratory), Dr. Peter Wallis (Manchester Metropolitan University), Dr. Etienne Roeche (University of Reading), Prof. Vincent Mueller (Anatolia College, Thessaloniki), Prof. Aaron Sloman (University of Birmingham) and Dr. Gerard Jagers op Akkerhuis (Wageningen University) for stimulating discussions relating to the topic

of this paper. This work was supported by the European Commission [grant numbers EU-FP6-507422, EU-FP6-034434, EU-FP7-231868 and EU-FP7-611971] and the UK Engineering and Physical Sciences Research Council [grant number EP/I013512/1].

7. References

- [1] Schmid, U., Ragni, M., Gonzalez, C. & Funke, J. (2011). The challenge of complexity for cognitive systems. *Cognitive Systems Research*, 12(3-4), 211-218.
- [2] de Kamps, M. (2012). Towards truly human-level intelligence in artificial applications. *Cognitive Systems Research*, 14(1), 1-9.
- [3] Vernon, D., Metta, G. & Sandini, G. (2007). A survey of artificial cognitive systems: implications for the autonomous development of mental capabilities in computational agents. *IEEE Transactions on Evolutionary Computation*, 11(2), 151-180.
- [4] Girard, B., Tabareau, N., Pham, Q. C., Berthoz, A. & Slotine, J.-J. (2008). Where neuroscience and dynamic system theory meet autonomous robotics: a contracting basal ganglia model for action selection. *Neural Networks*, 21(4), 628-641.
- [5] Bekkering, H., de Bruijn, E. R. A., Cuijpers, R. H., Newman-Norlund, R., van Schie, H. T. & Meulenbroek, R. (2009). Joint action: neurocognitive mechanisms supporting human interaction. *Topics in Cognitive Science*, 1, 340-352.
- [6] Langley, P., Laird, J. E. & Rogers, S. (2009). Cognitive architectures: research issues and challenges. *Cognitive Systems Research*, 10(2), 141-160.
- [7] Anderson, M. L. (2003). Embodied cognition: a field guide. *Artificial Intelligence*, 149, 91-130.
- [8] Vernon, D. & Furlong, D. (2007). Philosophical foundations of enactive AI. In M. Lungarella et al. (Eds.), *50 Years of AI Festschrift, LNAI 4850* (pp. 53-62). Heidelberg: Springer-Verlag.
- [9] Maturana, H. R. & Varela, F. J. (1987). *The Tree of Knowledge: The Biological Roots of Human Understanding*. Boston, MA: New Science Library/Shambhala Publications.
- [10] Varela, F. J., Thompson, E. & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.
- [11] Vernon, D., Metta, G., & Sandini, G. (2010). Embodiment in cognitive systems: on the mutual dependence of cognition and robotics. In J. Gray & S. Nefti-Meziani (Eds.), *Embodied Cognitive Systems*. London: Institution of Engineering and Technology (IET).
- [12] Froese, T. & Ziemke, T. (2009). Enactive artificial intelligence: investigating the systemic organiza-

¹⁶ The latest version of the glossary can be found at <http://www.dcs.shef.ac.uk/~roger/pictograms.html>.

- tion of life and mind. *Artificial Intelligence*, 173(3-4), 466-500.
- [13] Vernon, D. (2010). Enaction as a conceptual framework for developmental cognitive robotics. *Journal of Behavioral Robotics*, 1(1), 89-98.
- [14] Georgeon, O. L. & Ritter, F. E. (2012). An intrinsically-motivated schema mechanism to model and simulate emergent cognition. *Cognitive Systems Research*, 15-16, 73-92.
- [15] Maturana-Romesin, H., & Mpodozis, J. (2000). The origin of species by means of natural drift. *Revista Chilena de Historia Natural*, 73(2), 261-310.
- [16] Goldspink, C., & Kay, R. (2010). Autopoiesis and organizations: a biological view of social system change and methods for their study. In M. R. Olivas-Lujan & T. Bondarouk (Eds.), *Advanced Series in Management* (Vol. 6, pp. 89-110). Bingley, UK: Emerald Group Publishing Limited.
- [17] Grand, S. (2004). *Growing Up With Lucy: How to Build an Android in Twenty Easy Steps*. London: Weidenfeld & Nicolson.
- [18] Franchi, S. (2013). Homeostats for the 21st century? Simulating Ashby simulating the brain. *Constructivist Foundations*, 9(1), 93-101.
- [19] Yim, M., Shen, W., Salemi, B., Rus, D., Moll, M., Lipson, H., Klavins, E. & Chirikjian, G. (2007). Modular self-reconfigurable robot systems [Grand challenges of robotics]. *IEEE Robotics & Automation Magazine*, 14(1), 43-52.
- [20] Arunalatha, G., & Jhony, C. (2010). Self-healing robots. In *INTERACT-2010* (pp. 283-287). New York: IEEE.
- [21] Sharkey, N. E. & Ziemke, T. (2001). Mechanistic versus phenomenal embodiment: can robot embodiment lead to strong AI? *Cognitive Systems Research*, 2(4), 251-262.
- [22] Haselager, W. F. G. (2005). Robotics, philosophy and the problems of autonomy. *Pragmatics & Cognition*, 13(3), 515-532.
- [23] Chan, V., Park, K., Collens, M. B., Kong, H., Saif, T. A. & Bashir, R. (2012). Development of miniaturized walking biological machines. *Nature Scientific Reports*, 2(857).
- [24] Metta, G., Vernon, D., Natale, L., Nori, F., & Sandini, G. (2008). The iCub humanoid robot: an open platform for research in embodied cognition. In R. Madhavan & E. Messina (Eds.), *Proceedings of the Performance Metrics for Intelligent Systems Workshop (PerMIS)* (Vol. 1090, pp. 50-56). Gaithersburg, MD: NIST.
- [25] Nishio, S., Ishiguro, H., & Hagita, N. (2007). Geminoid: teleoperated android of an existing person. In A. C. de P. Filho (Ed.), *Humanoid Robots, New Developments* (pp. 343-352). Vienna, Austria: I-Tech.
- [26] Sahin, E., & Spears, W. M. (Eds.). (2005). *Swarm Robotics* (Vol. 3342). Berlin, Heidelberg: Springer.
- [27] Hinchey, M. G., Sterritt, R. & Rouff, C. (2007). Swarms and swarm intelligence. *Computer*, 40(4), 111-113.
- [28] Cummins, F. (2014). Agency is distinct from autonomy. *Avant: Trends in Interdisciplinary Studies*, 5(2), 98-112.
- [29] Marsh, L. & Onof, C. (2008). Stigmergic epistemology, stigmergic cognition. *Cognitive Systems Research*, 9(1-2), 136-149.
- [30] Braitenberg, V. (1984). *Vehicles: Experiments in Synthetic Psychology*. Cambridge, MA: MIT Press.
- [31] Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, 47, 139-159.
- [32] Arkin, R. C. (1998). *Behavior-Based Robotics*. Cambridge, MA: MIT Press.
- [33] Haugeland, J. (1985). *Artificial Intelligence: The Very Idea*. Cambridge, MA: The MIT Press.
- [34] Di Paolo, E., Rohde, M., & De Jaegher, H. (2010). Horizons for the enactive mind: Values, social interaction, and play. In *Enaction: Towards a New Paradigm for Cognitive Science*. Cambridge, MA: The MIT Press.
- [35] Dennett, D. (1989). *The Intentional Stance*. Cambridge, MA: The MIT Press.
- [36] Davidson, R., Scherer, K. R., & Goldsmith, H. (2003). *Handbook of Affective Sciences*. Oxford, New York: Oxford University Press.
- [37] Baron-Cohen, S. (1997). *Mindblindness: Essay on Autism and the Theory of Mind*. Cambridge, MA: The MIT Press.
- [38] Bratman, M. E. (1987). *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.
- [39] Dautenhahn, K. (1995). Getting to know each other - artificial social intelligence for autonomous robots. *Robotics and Autonomous Systems*, 16(2-4), 333-356.
- [40] Dautenhahn, K. (1997). I could be you - the phenomenological dimension of social understanding. *Cybernetics and Systems*, 28(5), 417-453.
- [41] Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, 10(2), 70-76.
- [42] Pezzulo, G. (2011). Shared representations as coordination tools for interaction. *Review of Philosophy and Psychology*, 2, 303-333.
- [43] Dautenhahn, K., & Nehaniv, C. L. (2002). *Imitation in Animals and Artifacts*. Cambridge, MA: The MIT Press.
- [44] Nehaniv, C. L., & Dautenhahn, K. (2007). *Imitation and Social Learning in Robots, Humans and Animals*. Cambridge, UK: Cambridge University Press.
- [45] Cangelosi, A., Metta, G., Sagerer, G., Nolfi, S., Nehaniv, C., Fischer, K., Tani, J., Belpaeme, T.,

- Sandini, G., Nori, F., Fadiga, L., Wrede, B., Rohlfing, K., Tuci, E., Dautenhahn, K., Saunders, J. & Zeschel, A. (2010). Integration of action and language knowledge: a roadmap for developmental robotics. *IEEE Transactions on Autonomous Mental Development*, 2(3), 167-195.
- [46] Mirza, N. A., Nehaniv, C. L., Dautenhahn, K. & te Boekhorst, R. (2006). Interaction histories: from experience to action and back again. In *Fifth International Conference on Development and Learning*. Bloomington, Indiana: IEEE Computer Science Press.
- [47] Broz, F., Nehaniv, C. L., Kose-Bagci, H. & Dautenhahn, K. (2012). Interaction histories and short term memory: enactive development of turn-taking behaviors in a childlike humanoid robot. <http://arxiv.org/abs/1202.5600>.
- [48] Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the IRE*, 37(1), 10-21.
- [49] Friston, K. & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B*, 364(1521), 1211-1221.
- [50] Scott-Phillips, T. (2015). *Speaking Our Minds: Why human communication is different, and how language evolved to make it special*. London, New York: Palgrave MacMillan.
- [51] Cowley, S. J. (2009). Distributed language and dynamics. *Pragmatics & Cognition*, 17(3), 495-508.
- [52] Worgan, S. F. (2011). Towards an artificial model of 'linguaging': reviewing the distributed language hypothesis. *Language Sciences*, 33(1), 229-236.
- [53] Bickhard, M. H. (2007). Language as an interaction system. *New Ideas in Psychology*, 25(2), 171-187.
- [54] Nehaniv, C. L., Forster, F., Saunders, J., Broz, F., Antonova, E., Kose, H., Lyon, C., Lehmann, H., Sato, Y. & Dautenhahn, K. (2013). Interaction and experience in enactive intelligence and humanoid robotics. In *2013 IEEE Symposium on Artificial Life (ALife)* (pp. 148-155). IEEE.
- [55] Malle, B. F. (2002). The relation between language and theory of mind in development and evolution. In T. Givón & B. F. Malle (Eds.), *The Evolution of Language out of Pre-Language* (pp. 265-284). Amsterdam: Benjamins.
- [56] Pickering, M. J. & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(04), 329-347.
- [57] Friston, K. J. & Frith, C. D. (2015). Active inference, communication and hermeneutics. *Cortex*, 68, 129-143.
- [58] Friston, K. & Frith, C. (2015). A Duet for one. *Consciousness and Cognition*, 36, 390-405.
- [59] Austin, J. L. (1980). *How to Do Things with Words*. Oxford: Oxford University Press.
- [60] Cowley, S. J. (2004). Contextualizing bodies: human infants and distributed cognition. *Language Sciences*, 26(6), 565-591.
- [61] Linell, P. (2013). Distributed language theory, with or without dialogue. *Language Sciences*, 40(0), 168-173.
- [62] Fusaroli, R. & Raczaszek-Leonardi, J., Tylen, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, 32, 147-157.
- [63] Premack, D. & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1(04), 515-526.
- [64] Bugnyar, T., Reber, S. A. & Buckner, C. (2016). Ravens attribute visual access to unseen competitors. *Nature Communications*, 7, 10506.
- [65] Winfield, A. F. T. (2010). You really need to know what your bot(s) are thinking about you. In Y. Wilks (Ed.), *Close Engagements with Artificial Companions* (pp. 201-208). Amsterdam: John Benjamins Publishing Company.
- [66] Moore, R. K. (2016). Is spoken language all-or-nothing? Implications for future speech-based human-machine interaction. In *International Workshop on Spoken Dialogue Systems (IWSDS)*. Saariselkä, Finland.
- [67] Moore, R. K. (2015). From talking and listening robots to intelligent communicative machines. In J. Markowitz (Ed.), *Robots that Talk and Listen* (pp. 317-335). Boston, MA: De Gruyter.
- [68] Goetz, J., Kiesler, S. & Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human-robot cooperation. *12th IEEE International Workshop on Robot and Human Interactive Communication*. Berkeley, CA.
- [69] Bartneck, C., Kanda, T., Mubin, O. & Al Mahmud, A. (2009). Does the design of a robot influence its animacy and perceived intelligence? *International Journal of Social Robotics*, 1(2), 195-204.
- [70] Moore, R. K. & Maier, V. (2012). Visual, vocal and behavioural affordances: some effects of consistency, *5th International Conference on Cognitive Systems - CogSys 2012*. Vienna, Austria.
- [71] Mori, M. (1970). Bukimi no tani (The uncanny valley). *Energy*, 7, 33-35.
- [72] Mitchell, W. J., Szerszen Sr., K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M. & MacDorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *I-Perception*, 2(1), 10-12.
- [73] Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J. & Frith, C. (2011). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience*, 7(4), 413-422.

- [74] Moore, R. K. (2012). A Bayesian explanation of the 'Uncanny Valley' effect and related psychological phenomena. *Nature Scientific Reports*, 2, 864.
- [75] Synofzik, M., Vosgerau, G. & Newen, A. (2008). I move, therefore I am: a new theoretical framework to investigate agency and ownership. *Consciousness and Cognition*, 17(2), 411-424.
- [76] Barandiaran, X., Rohde, M. & Di Paolo, E. A. (2009). Defining agency: individuality, normativity, asymmetry and spatiotemporality in action. *Adaptive Behavior*, 17, 367-386.
- [77] Tannenbaum, E. (2009). Speculations on the emergence of self-awareness in big-brained organisms: the roles of associative memory and learning, existential and religious questions, and the emergence of tautologies. *Consciousness and Cognition*, 18(2), 414-427.
- [78] Thompson, E. & Varela, F. J. (2001). Radical embodiment: neural dynamics and consciousness. *Trends in Cognitive Sciences*, 5(10), 418-425.
- [79] Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17(11), 565-573.
- [80] Seth, A. K. (2014). The cybernetic Bayesian brain. *Open Mind*. Frankfurt am Main: MIND Group.
- [81] Clowes, R. W. & Mendonça, D. (2016). Representation redux: is there still a useful role for representation to play in the context of embodied, dynamicist and situated theories of mind? *New Ideas in Psychology*, 40(A), 26-47.
- [82] Baars, B. J. (2005). Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Progress in Brain Research*, 150, 45-53.
- [83] Shanahan, M. (2010). *Embodiment and the Inner Life*. Oxford, New York: Oxford University Press.
- [84] Dehaene, S., Charles, L., King, J.-R. & Marti, S. (2014). Toward a computational theory of conscious processing. *Current Opinion in Neurobiology*, 25, 76-84.
- [85] Norman, D. A. (1993). *Things that Make Us Smart: Defending Human Attributes in the Age of the Machine*. Boston, MA: Addison Wesley Longman Publishing Co. Inc.
- [86] Card, S. K., Mackinlay, J. & Schneiderman, B. (1999). *Readings in Information Visualization: Using Vision to Think*. San Diego, CA: Academic Press.
- [87] von Neumann, J. (1966). *Theory of Self-Reproducing Automata*. (A. W. Burks, Ed.). Urbana & London: University of Illinois Press.
- [88] Jagers op Akkerhuis, G. (2010). *The Operator Hierarchy: A Chain of Closures Linking Matter, Life and Artificial Intelligence*. Ph.D. thesis, Radboud University, Nijmegen.
- [89] Jagers op Akkerhuis, G. (2009). The operator hierarchy: evolution beyond genes. In *ECAL-09*. Budapest, Hungary.
- [90] Bickhard, M. H. (2009). Interactivism: a manifesto. *New Ideas in Psychology*, 27, 85-89.
- [91] Bickhard, M. H. (1998). Levels of representationality. *Journal of Experimental & Theoretical Artificial Intelligence*, 10(2), 179-215.
- [92] Ziemke, T. & Lowe, R. (2009). On the role of emotion in embodied cognitive architectures: from organisms to robots. *Cognitive Computation*, 1(1), 104-117.
- [93] Vernon, D., Lowe, R., Thill, S. & 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*, 6, 1660.
- [94] Abler, W. L. (1989). On the particulate principle of self-diversifying systems. *Social Biological Structures*, 12(1), 1-13.
- [95] Kuhn, T. S. (1975). *The Structure of Scientific Revolutions* (2nd ed.). Chicago, IL: Univ. of Chicago Press.
- [96] Winfield, A. F. (2014). Robots with internal models: A route to self-aware and hence safer robots. In J. Pitt (Ed.), *The Computer After Me: Awareness And Self-Awareness In Autonomic Systems* (pp. 237-252). London: Imperial College Press.
- [97] Mithen, S. (1996). *The Prehistory of the Mind*. London: Phoenix.
- [98] Winfield, A. (2011). Directions in artificial cognition: new thoughts on thinking, *EU-FP7-UK Cognitive Systems and Robotics Workshop*. London.
- [99] Mehrabian, A. (1996). Pleasure-arousal-dominance: a general framework for describing and measuring individual differences in temperament. *Current Psychology: Developmental, Learning, Personality, Social*, 14, 261-292.
- [100] Breazeal, C. (2003). Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, 59(1-2), 119-155.
- [101] Fellous, J. & Arbib, M. (2005). *Who needs emotions?: The brain meets the robot*. Oxford, New York: Oxford.
- [102] Cangelosi, A. & Schlesinger, M. (2015). *Developmental Robotics: From Babies to Robots*. Cambridge MA, London: MIT Press.
- [103] Nolfi, S. & Floreano, D. (2000). *Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines*. Cambridge, MA: MIT Press.
- [104] Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, 50, 370-396.
- [105] Baldassarre, G., Stafford, T., Mirolli, M., Redgrave, P., Ryan, R. M. & Barto, A. (2014). Intrinsic motivations and open-ended development in animals,

humans, and robots: an overview. *Frontiers in Psychology*, 5, 985.

- [106] Powers, W. T. (1974). *Behavior: The Control of Perception*. Hawthorne, NY: Aldine.
- [107] Powers, W. T. (2008). *Living Control Systems III: The Fact of Control*. Montclair, NJ: Benchmark Publications.
- [108] von Uexkull, J. (1957). A stroll through the worlds of animals and men. In C. Schiller (Ed.), *Instinctive Behavior*. New York: International Universities Press.
- [109] Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, Acting, and Knowing: Toward an Ecological Psychology* (pp. 67-82). Hillsdale, NJ: Lawrence Erlbaum.
- [110] Heider, F. & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, 57(2), 243-259.

INTECH

INTECH

