Semantic Flexibility and Grounded Language Learning

Stephen McGregor [1] and **Thierry Poibeau** [1]

Abstract. We explore the way that the flexibility inherent in the lexicon might be incorporated into the process by which an environmentally grounded artificial agent – a robot – acquires language. We take *flexibility* to indicate not only many-to-many mappings between words and extensions, but also the way that word meaning is specified in the context of a particular situation in the world. Our hypothesis is that embodiment and embededness are necessary conditions for the development of semantic representations that exhibit this flexibility. We examine this hypothesis by first very briefly reviewing work to date in the domain of grounded language learning, and then proposing two research objectives: 1) the incorporation of high-dimensional semantic representations that permit context-specific projections, and 2) an exploration of ways in which non-humanoid robots might exhibit language-learning capacities. We suggest that the experimental programme implicated by this theoretical investigation could be situated broadly within the enactivist paradigm, which approaches cognition from the perspective of agents emerging in the course of dynamic entanglements within an environment.

1 Introduction

In the early 20th Century, Russell [57] performed a famous thought experiment involving a proposition about the baldness of the King of France. The object of this exercise was not actually regal coiffure; the point was that there was in fact no King of France at the time, and so the philosopher had attributed a measurable characteristic to a non-existent entity, rendering the truth-value of the statement ambiguous. The outcome of Russell's reflections was the inception of a programme designed to translate the vagaries of natural language as used by humans into a rigorously rule-bound system of logical expressions and a corresponding algebra of truth.

But several decades of cognitive linguistic and pragmatic theory and experimentation have given the lie to the idea that language is just a construct for conveying veridical propositions about the world [16, 23]. Psychological studies and mounting neurolinguistic research suggest that the interpretation of non-literal language is entangled with, and in some cases equivalent to, the processing of literal statements, and, moreover, that the context in which both figurative and literal language is encountered plays an important role in its processing [52]. It is evident that natural language is, in a very fundamental way, not simply about situations in the world; it is, rather, characterised by flexibility in terms of how it is applied in the course of agents attempting to achieve communicative goals, and as a consequence words very often do not explicitly denote in the way that Russell had hoped could be formalised [6].

The essentiality of non-literalness in natural language poses an interesting question for researchers interested in developing linguistically capable robots, or for that matter in using robots to study the way in which humans use language: how do lexical semantic representations grounded in an agent's experience of the world and interaction with other agents gain their fundamental flexibility? A general, and not unreasonable, assumption has been that robots learn first and foremost to ground words literally, in their experience of objects and actions in the world. But how do the representations learned in this way obtain the looseness and ambiguity that is ubiquitous to language in use [62]? This flexibility is, importantly, evident not only in more overt phenomena such as metaphor (which itself occupies a range from conventional to jarringly novel or even poetic), but more subtle phenomena such as image schemas [35, 33] and semantic type coercion [51, 15].

So, for instance, the way that human communicants quite naturally produce and interpret a sentence such as "I finished the book" actually imposes a mismatch between the argument type expected by the predicate *finished*, which invites an event in the objective position, and the type offered by the object *book*, which is a substantial thing. Similarly, the application of the verb *open* is itself remarkably open-ended: we can talk about "opening a package", "opening a door", "opening a shop", and so forth, all without in any obvious way transgressing the literal. How can we expect an agent that has acquired potentially quite specifically structured semantic representations for such actions and objects from basic encounters with them in the world to develop the architecture to seamlessly project from one categorical or conceptual domain to another? The way that language is actually observed in use threatens to perpetually confound any systematic attempt to map from structured symbolic representations of words to the type of actionable interpretations of sentences that we might hope to apply to the operations of an artificial agent.

Because of its role in the construction and transmission of concepts, language is often afforded a special status among cognitive phenomena, sometimes even presented as a kind of manifestation of thought itself [25]. In this paper, though, in line with a relevance theoretical account of the despecialisation of language [74], we seek to situate language on the same level as other behaviours exhibited by cognitive agents in the course of their interactions with environments and communities. In order to do this, we turn to Gibson's [30] notion of *affordance*, which we take to be the direct, non-representational perception of opportunities for action in an environment, and apply the same thinking that pertains to other objects to the lexical units of language. So, just as an agent can perceive a newspaper as something that affords the opportunity to swat a fly in a certain context, we argue that the lexicon should be perceived directly as an opportunity for communicating, and that words are picked up, used, and received by communicants in the same open-ended way.

Starting from this premise, we propose three desiderata for developing artificial language-using agents:

1. Lexical semantic units should be flexible from the ground up;

[1] Laboratoire Lattice, CNRS & École normale supérieure / PSL, Université Sorbonne nouvelle Paris 3 / USPC

2. Semantic flexibility should arise from environmental entanglement and be built into the structure of semantic representations themselves;
3. These representations should permit environmentally triggered projection into context-specific subspaces.

What follows is a position paper exploring the grounds for these stipulations, the ways in which they might be implemented, and some of the implications of such implementations. The next two sections offer a very cursory review of the current state of the art in terms of models of the emergence of semantics from multi-agent interactions and research investigating the way that embodied agents might acquire language in the real world. This review serves as the motivation for two ideas for the direction of future research in this area.

2 Language Learning and Interaction

The idea of language as an act of meaning-making has served as the theoretical basis for an entire field of productive empirical projects investigating the emergence of semantics through interactions between agents communicating in an environment [66]. Work in this area has typically entailed experiments involving simulations of agents interacting in environments in which the emergence of communication provides a fitness advantage to either individuals or groups. As such, the theoretical grounding for this research has often incorporated the modelling of various components of the evolution of language [61].

An evolutionary approach to the emergence of language has lent itself to a consideration of how natural language, with its syntax[2] and corresponding open-endedness, might arise from the dynamics of more basic signalling phenomena [63]. Franke and Jäger [26] describe a model for coordinating expressions and interpretations between two communicants through language games grounded in the application of optimality theory [50] to signal interpretation, pushing multi-agent simulations into the realm of truth conditions and implicature. No assumptions are made here about the basis for constructing, broadcasting, and receiving signals, however, so this work does not provide, and is not intended to provide, an experimental basis for the semiotic processes by which units of language gain their compositional potential [28].

Lazaridou et al. [37] describe a model that involves pairs of agents playing language games in a simulated environment: the agents attempt to come to a consensus on semantic representations for both abstract objects represented as collections of categorical features (such as shape and colour) and real-world static images of objects. Compellingly, the authors illustrate how the representational systems that emerge in the course of their agents' interactions contain elements of compositionality. So, for instance, their agents learn to consistently apply a certain semantic component when converging on names for various green objects in the abstract version of the experiment. In a related experiment, through a clever interpolation of an independent image classification task with the joint naming task, the same authors show that their agents have a propensity for mapping symbols associated with one image class to the naming of a related task, using the same symbols in, for instance, identifying a picture of open water with a symbol used in the classification of a picture of a dolphin [38].

[2] There is an impressive body of work exploring, theoretically and empirically, emergentist models of grammar [65, 32]; here, we focus on the emergence of semantic flexibility.

It is important to note, however, that the basis of these semantic units are arbitrarily abstract. In the implementations described by Lazaridou et al., a representation for an object is composed of a sequence of integers, drawn from a vocabulary grounded simply in the pseudo-randomness of a computer simulation of a stochastic process. The semiotics of these representations are at best obscure; the open-ended compositionality of the symbols is an extrapolation of categorical structure that is inherent in the virtual environment, in the behaviour of the linguistic community, or in the pre-established cognitive architecture of an individual agent [67].

Because the interpretations associated with representations and the *way in which* those representations stand for things in the world are dissociated, there is little chance for a community of agents using this type of emergent but also abstract language to capture the lexical flexibility that is an essential component of natural language. As a case in point, linguists have noted the way that perception in general can be applied metaphorically to more abstract cognitive experiences, and moreover the way that certain modes of perception tend, at least within certain cultures and families of languages, to be reliably applied to certain metaphoric targets [56, 69]. Cognitive linguists have accordingly postulated a link between loose - in particular, metaphoric - lexical semantics and corresponding cognitive flexibility [36, 29]. It is unclear, however, and, we argue, unlikely that agents will through abstract interactions arrive at representational frameworks that facilitate mappings from, for instance, COLOUR to AFFECT if the representations themselves are not in some sense grounded in the specific mechanics of the way that the agent actually physically, bodily interacts with its environment.

It would be misleading to suggest that these experiments on the emergence of language through multi-agent interactions are performed in an entirely disembodied manner. There is often an explicit acknowledgement of the significance that the particular biology of an agent plays in language grounding. In practice, though, this environmental grounding is typically realised through the presentation of data that is in some very general sense in-the-world, so for instance the presentation of photographs of objects as part of a naming game. In these cases, the embodiment itself is presumed to be captured in the nature of the way that the agent passively processes the raw data, and the connection between the agent and the world becomes obscured by the opacity and abstractness of dense image processing networks.

We propose that physical embodiment and a corresponding model of semiotics that is grounded in the body and the environment of an agent is a necessary condition for generating the type of semantic representations that exhibit the flexibility of application that natural language exhibits universally and at the most basic level. In the next section, we will briefly review the state of the art of work with language learning robots and consider ways in which these machines might be used to capture the emergence of flexible semantic representations through interaction with an environment.

3 Language Learning and Embodiment

In a concise and informative survey of recent work in teaching robots abstract concepts and corresponding language, Cangelosi and Stramandinoli [10] postulate that lexical flexibility is achieved through a combination of embodied symbol grounding, transferal between grounded conceptual domains, and the interaction of language use and physical action. Early work in this area involved simulated robots learning to imitate actions and then combine these actions into novel routines in response to compound commands, using neural networks

to model the way that sensorimotor processes can map to representations that have very basic properties of, if not compositionality, at least concatenation [9]. Subsequent experiments have explored the way that the actual bodies of simulated robots can play a role in learning action-oriented linguistic activities such as counting: De La Cruz et al. [19] demonstrate that coupling sequences of spoken numbers with sequential finger counting during training greatly strengthens a simulated robot's ability to accurately count.

A primary motivation for at least some linguistic experiments on robots has been to study the valid question of how noisy and unpredictable environments impact various models of symbol grounding [68]. An assortment of robots [27, 41] and corresponding platforms for running virtual simulations have proven valuable tools for exploring the way that an artificial agent interacts with its environment in the course of language learning, and the ways that environmental factors can both support and confound the learning process. An open question, though, is whether or not these simulations, or for that matter the robots themselves, achieve the level of abstraction necessary to capture the way in which an agent's physiognomy interacts with its environment in order to generate semantic representations with the flexibility inherent in natural language.[3] The idea of associating both sequences of actions and labels with underlying cognitive architectures seems well motivated and is an excellent starting point, but the actual way of being in the world that characterises the agent, the mechanisms of its action and the way that these mechanisms came to exist, are grounded out at a lower level of engineering decisions and corresponding hardware.

Hernández et al. [31], in their neuroanatomically inspired approach to grounding robotic language learning through visual stimuli, use both simulated and real robots but constrain their robots to maintain a static field of vision. This is a move to simplify input to an information processing architecture which is already, by design, complex; the purpose of the research described by those authors is in large part to explore ways that language-learning robots can be used to study the brain. That said, it is a move that is made at the expense of the idea that vision, and indeed all perception, depends on the activation of *sensorimotor contingencies* that make the experience of perceiving an active process of environmental engagement [44]. According to the sensorimotor account, it is through these contingencies that perceptions acquire their actual feel, and so, arguably, the basis for forming a complex representational framework for conceptualising the experience of being in the world.[4] The idea is that an agent perceives a situation in an environment based not merely on a passive analysis of the data available to its sensors, but from an unfolding engagement with the environment involving the coupling of sensors with actuators to such an extent that the distinction becomes blurred [43].

Working off of this account, and more generally from the enactivist standpoint on the emergence of representation as an outcome of the evolution of self-perpetuating, self-replicating processes in a complex environment [39, 70], we might conclude that an agent that is not participating in its environment is not really learning anything at all. Returning to the theme addressed in Section 2, we propose that, if linguistic flexibility is to be an essential component of a system of lexical semantic representations, then the representations must be in fact *products of an agent's actions*, rather than just abstractly mapped to them. Because we require our semantic representations to be structurally flexible, we propose it is necessary to consider before all else the way this flexibility will be instantiated in our agent's architecture, and the type of in-the-world agent we need in order to achieve this flexibility in the course of interaction with an environment.

[3] Wittgenstein [75] describes "the familiar physiognomy of a word, the feeling that it has taken up its meaning into itself," (p. 218).

[4] O'Regan [45] has argued that sensorimotor contingencies can actually account for phenomenal consciousness, a claim which is outside the scope of the current paper, though it is worth noting that metaphoric language is arguably a prerequisite for conceptualising the experience of having a mind, or, to put it differently, for constructing concepts about concepts [3, 40].

4 Modelling Lexical Flexibility

Practically speaking, it is important to note that the idea of flexibility in semantic representations has, broadly, been considered in some existing work. Wellens et al. [73], for instance, describe a set of grounded language learning experiments that explicitly target linguistic flexibility, albeit conducted with humanoid robots. The results of these experiments are compelling, but it is important to note that the concept of *flexibility* employed by those authors is related to but distinct from the one we are considering: where they are interested in the way that agents learn to map from potentially ambiguous intensional properties to potentially ambiguous extensional denotations, we are interested in the way that semantic representations can be contextually adapted in the process of constructing *ad hoc* concepts [11, 1]. Rather than presume that the agents that we would like to model have a developed conceptual framework ready for lexical semantic enhancement – in other words, have in some sense preformulated internal representations – we propose to explore the way in which representations themselves might come about as a result of having a physiognomy in an environmental situation.

In order to open up a consideration of the way in which grounded agents might acquire lexical semantic representations that are flexible from the ground up, we begin by adopting Rączaszek-Leonardi et al.'s [55] proposal to re-imagine the symbol grounding problem as a *symbol ungrounding problem*. This is effectively a call to take seriously the situated semiotics of the way that agents acquire their linguistic representations, and to consider language itself as a physically bounded *system of replicable constraints* [53]. Under this premise, linguistic development begins with the experience of very basic interpersonal communications in the environment of an early-stage language learner. These communications, as percepts in the learner's environment, are associated with *affordances*, or direct opportunities for action [30]. The things afforded by these primal communications are not necessarily associated with truth-values; they can instead be mapped to the sense of, for instance, mutual attention to one another that is an important element of early-life interactions, followed by a pushing-out into experiences of shared attention to objects in the world [54].

The process by which these early experiences of proto-linguistic communication solidify into a system of representational, compositional, and, importantly, flexible symbols lines up with the phenomenon of *ontogenetic ritualisation* as described by Spranger and Steels [64]. Those authors describe the way that the action of an infant reaching for but failing to obtain an object gradually develops into the ubiquitous communicative gesture of pointing to indicate the desire to have that object. This is an example of the way that the environment, its affordances, and the physical capacities and constraints of an agent become the semiotic basis for an emergent representation. Spranger and Steels perform an experiment involving a real robot in which they make the case that a learner robot develops a reaching gesture as a symbol for signalling a tutor robot to push a distant object towards them.

We would like to suggest that an agent's lexicon could materialise in a similar way. What begin as basic patterns of signals deployed

simply to attract attention might evolve into more complex representations with the capacity to be interpreted in different ways based on the context in which they are encountered. In order to capture the flexibility of these semantic representations, however, we propose that they should take shape in a way that is deeply connected with the actual body of the agent. With this in mind, we suggest that the connections between the environment and the mechanical architecture of the agent should be as straightforward as possible, avoiding unnecessary networks of dense, complex hidden layers. Instead of having a multi-staged approach of first interpreting and then reacting to the environment, the representations of the environment can be incorporated with the responses to the environment. Then, rather than assigning a semantic label to the mapping from inputs to actions, we can cast perceiving, interpreting, and expressing as actions in themselves, associated with their own affordances.

So, for instance, if a robot is learning a representation for *open* in the context of a door, there could be information incorporated into the representation regarding not only the actions learned in order to open a door (which may consist of a sequence of learned or programmed subroutinnes), but also the various raw stimuli detected by the robots sensors as well as any action policies associated with adjusting sensors. The semantic representation for *open* is then cast as a vector $\overrightarrow{open}$, where the features of the vector are the combinations of states, actions, and environmental inputs associated with the door opening event. A subsequent identification of the utterance "open" should invoke this representation, but projected into the context in which the utterance is encountered by way of feature weights. In this way we hope to introduce flexibility to semantic representations through a mechanism for projecting the representation into a particular context, and some, but not all, of the components of the robot's representational framework for opening doors might be transferred to, for instance, opening packages or opening books.

The apparatus for building high-dimensional semantic representations is the subject of work in the *distributional semantic* paradigm, which seeks to generate representations for words based on observations taken across large-scale textual corpora [71, 14]. The problem with distributional semantics in the context of grounded language learning is, of course, that the representations are generated based on an analysis of a very large amount of textual data taken outside of any sort of real-world environment: this is very different than the way an embodied artificial agent encounters the world, and almost certainly results in a very different sort of representation than what we are after. An additional obstacle is the fact that the dimensions of distributional semantic representations tend to be quite abstract. These representational spaces are generally generated through either matrix factoring applied to a large and sparse matrix of co-occurrence statistics [21], or, more typically at present, through the opaque operations of neural networks trained on language modelling objectives [42, 47, 48].

By mapping representations to features of the actual environment encountered by the language-learning agent, though, the representations themselves are grounded in situations in the world. With this set-up, an agent receiving linguistic input might identify the input and then instantaneously project it into a subspace specified by the current environmental context. The world, as Brooks [7] has proclaimed, becomes its own model, and the environment itself provides the traction for projecting the representation into a contextually interpretable subspace. This way, something of the sense of *open* in the context of doors can be transferred to the literal but different contexts of books, or shops, and then onward along the graded slope of figurativeness towards the senses in which events, or communications, or indeed minds are things that can be opened. More generally this kind of representational architecture might provide the basis for satisfying the embodied and embedded requirements of semantic phenomena such as image schemas, where the embodied experience associated with an action maps onto a linguistic representation (especially of a preposition or phrasal verb).

There are still many decisions to be made regarding the level of abstraction of an environmentally grounded semantic representation: there will necessarily be mitigating layers of recurrent and convolutional neural networks processing noisy raw perceptual stimuli in time and space. By associating affordances with sensorimotor contingencies, we hope to lay the groundwork for building the actions associated with perceptions into representations themselves, rather than constructing representations as labels for mappings between perceptions and actions. But for now this remains a high-level theoretical commitment very much in need of a more thorough consideration of how it can actually be technically applied.

5 Building Dynamically Situated Linguistic Agents

The robots employed in the research surveyed in Section 3 share the particular property of being *humanoid*. This makes sense: language is an important part of human activities, and conversely the flexibility and compositionality evident in natural language are arguably uniquely human attributes [49, 22].[5] The environmental pressures experienced by humans and their ancestors on an evolutionary timescale have provided a set of physiological, cognitive, and social conditions particularly suited to linguistic communication. On top of that, the objective of much research involving language learning robots is, as previously mentioned, to use entities that are humanlike at a certain level of abstraction in order to study the relationship between features of human anatomy and language.

By the same token, though, the human body has come to fill the niche it does over the course of a very specific and immeasurably complex history of events spanning an immense period of time.[6] There are many components of being human that feed into the nature of language as observed in use, and, to return to a recurring theme, it seems difficult at best to specify the level of abstraction at which an artificial agent needs to be specified in order to emulate the grounding of natural language as experienced in human form.

With this in mind, it may be necessary to delve into a consideration of what Sloman has characterised as *possible minds* [60], a phrase intended to encompass the idea that what characterises the cognitive is not necessarily exclusively or indeed inherently human. The upshot of abandoning the doubly challenging objective of engineering humanlike language instantiated *in a human form* is, from an engineering perspective, that language can be treated as an objective, a teleological force guiding iterations of design decisions, rather than an emergent solution to an adaptational problem presented by a complex and changing environment. This opens the path for exploring the minimal requirements for *the kind of machine we would need* in order to implement an agent that develops, in the course of its entanglement with the world, a language.

[5] The extent to which natural language is either innately or uniquely human remains perhaps one of the most controversial topics in contemporary linguistics [12, 24]; we do not intend to take a stance on the subject, aside from to consider the real possibility of designing a language-using machine.

[6] Davidson [18] has argued, by way of thought experimentation, that a word acquires its meaning *only* through a personal and communal history of use; deracinated from this history, it has no meaning at all, and is just another material thing that happens.

Authors such as Clark [13] and Zwaan [77] have made the case that an agent's environment provides an essential structural component to both cognition and language, and a natural continuation of these ideas is the enactivist approach to cognitive science. Enactivism embraces the idea that cognitive agents can be modelled in terms of self-perpetuating systems at the nexus of networks of environmental constraints: mindful individuals are, essentially, homeostats. By this account, representations are emergent properties of, rather than causal components within, systems calibrated for responding to specific contexts within complicated and unpredictable environments [72]. Enactivism has been explored in the context of self-organising systems and evolutionary robotics [76, 4], but linguistic applications remain elusive, perhaps because language is so naturally conceptualised in terms of representational intentionality [58]. Cuffari et al. [17] do, though, put forward a framework for an enactivist approach to language, built on the foundational concept of *participatory sense-making* as the basis for a process of *languaging* that is wrapped up in interactions between agents and within environments.

The first component of this two-pronged enactivist framework for language is a model of how language emerges through a series of resolutions of dynamic tensions that arise in the course of sense-making in a social, which is to say interactive, setting. Each resolution generates a new pair of tensions between the individual and the social components of the model, with this sequence of tensions and solutions ultimately resulting in the production of a linguistic mode of communication. The second component of the framework is a model designed to capture the dynamic couplings between sense-making, having a body, and being in the world which are the underlying components of the system that results in the emergence of language. This model is intended to reflect the ontogenetic component of language, by which the physical human interaction with the world serves as the basis for the construction of a language. Cuffari et al. describe this ontogeny as resulting in an *increasingly linguistic* mode of communication; for our purposes, we interpret this graded component of the model to correspond to the introduction of semantic flexibility to emergent linguistic representations.

Practically applied to the design of robots, this theoretical framework might begin to look something like a *subsumption architecture* [7]. The premise of these architectures is that low level systems of constraints, instantiated in the basic circuitry of a robot, become the operational units for higher level processes. In terms of language learning, the low level routines might involve basic objectives for social functionality such as avoiding collisions and conflicts (and these routines may themselves be contingent on lower level subroutines). This idea squares with Deacon's [20] hierarchical model of human cognition, which postulates that networks of constraints at a particular level of abstraction become the basis for emergent attractors which in turn become constraints for further emergent properties. Deacon, grounding his theory in the biosemiotic paradigm [34, 46], makes a compelling argument for the way that these levels of emergence eventually lead to the materialisation of representation, intentional, and even teleological components of a cognitive architecture.

But on what level of abstraction do semantic representations with the flexibility and compositionality characteristic of natural language emerge? We would be remiss to claim to have an answer to this question, but we propose that thinking about the design of language-learning robots that might take a very different form than ourselves is a good place to start. By stepping away from the constraints of the human body and its place in its evolutionary and social niches, we have the chance to explore the minimal conditions necessary for the emergence of a mode of language that exhibits open-ended semantic flexibility.

6 Conclusion

Based on a recognition of the fundamental flexibility of natural language, we have proposed two components for a new research project in language learning for artificial agents. The first involves the construction of an architecture in which semantic labels are emergent properties of direct connections between environment and physiognomy, motivated by the idea that these type of connections will facilitate the contextualisation that invites and grounds semantic flexibility. The second is a call to consider the ways in which non-humanoid agents might acquire language, a move which could allow us to focus on the basic conditions under which agent-environment entanglement results in the emergence of flexible semantic representations. Both of these positions are, as they stand, underspecified, and we present them simply as a starting point for future research. In the meantime, by way of a conclusion, we will explore two philosophical implications of the embedded and embodied framework that we have outlined.

The first implication regards the broad assumption that a robotic language facility would be in some way or another *portable*, either by taking the form of software that is installed on robotic hardware or else as an architecture that could be extrapolated from one machine and then transferred to any number of other machines [5]. We are pessimistic about this idea, though. If semantic representations arise out of a deep interconnection between the actual physical architecture of the robot and its environment, it is difficult to imagine how the representations learned by one robot in one set of circumstances would map to a different robot in a different situation. As mentioned throughout this paper, the level of abstraction at which the critical fusion between agent and environment occurs is ambiguous, perhaps inextricably so given that this remains a completely open problem in theoretical linguistics, as well; intuitively it seems likely that language happens at many different levels. With this in mind, it is unrealistic to expect that the combinations of basic materials, circuitry, mechanisms, and pre-programmed routines that constitute a *tabula rasa* robot, combined with the myriad environmental conditions encountered by a robot in the course of language-learning, will translate smoothly from one machine to another.

The second implication pertains to what we will coin the *Asimovian fallacy*, with reference to the three laws of robots famously laid out in Isaac Asimov's science fiction writing [2]. These tenets, which take the form of commandments about robots' actions, presume a high-level correspondence with the robot about its goal-directed behaviour. What is generally left to the reader's imagination in the original and subsequent literature is the representational form that goals will actually take within the cognitive architecture of a robot. We suggest, in line with Deacon [20], that goals only come about in the course of developing a system of intentional representations, and indeed as an emergent attractor of this process. If this is the case, then it must be impossible to separate the representation of goals from the semantic representations and underlying semiotic processes used to communicate about goals, and so, by the time a robot can be analysed as having goals, the behaviour of the machine will be inextricably entangled with its cognitive architecture, not something that can be isolated and modified.

This is the stuff of science fiction. More immediately, though, an important assumption that is more or less ubiquitous to simulations of the emergence of language through multi-agent interaction is that the agents participating in language games are endowed with goals

that pertain to persisting in their environments. Moreover, these goals are generally explicitly communicative: there is a presumption that effective communication gives an agent a fitness advantage in the environment. But there is a subtle speciousness lurking in the presumption that communicative goals could precede communication itself. Having goals is an emergent property of having intentional representations, so the idea of the goal of communication can only arise in the course of the analysis of the individual dynamic situations in which communication happens, and of the general history of situations over which communication evolves.

In fact, a significant consequence of the ontogenetic model of the emergence of language presented by Cuffari et al. [17] is that the individual appears in tandem with the emergence of language through social interaction.[7] The implication here is that the very idea of an *agent* is wrapped up in the same processes that lead to the emergence of a system of language characterised by an openly flexible lexicon. The conceptualisation of goals requires a process of representing things that are not present by way of a systems of interacting constraints. The language of *objectives* is a way of talking about emergent properties of these systems, rather than an integral causal component of them.

There is an additional ethical ramification to the idea that robots cannot have goals without having semantics. Bryson [8] has made a case that humans should not build machines to which they have ethical obligation: we should draw a line at developing robots with a cognitive architecture that would compel us to consider, for instance, the psychological comportment of the machine. This seems reasonable enough, but in the end it is a stipulation that might be entirely anathema to building language-using robots, or indeed other semi-sophisticated forms of linguistic AI, in the first place. If we accept that language proper entails a conceptual structure that necessarily corresponds to some sort of cognitive architecture, then, by the time we are actually talking with our devices, we have already entered into a situation where we have no choice but to consider, at least in a superficial or performative sense, that those mechanical systems are cognitive entities that in turn entail ethical commitments beyond what we would assign to a mere lump of matter. If we fail to make this commitment, then what is presented as language use is immediately relegated back to the realm of mere signalling posturing as language.

REFERENCES

[1] Nicholas Allott and Mark Textor, 'Lexical pragmatic adjustment and the nature of ad hoc concepts', *International Review of Pragmatics*, **4**, 185–208, (2012).

[2] Isaac Asimov, *I, Robot*, Bantam Books, 1950.

[3] John A. Barnden, 'Metaphor, self-reflection, and the nature of mind', in *Cognition and Affect*, ed., D. N. Davis, 45–65, Information Science Publishing, Idea Group Inc., (2005).

[4] Randall D. Beer, 'The dynamics of active categorical perception in an evolved model agent', *Adaptive Behavior*, **11**(4), 209–243, (2003).

[5] Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, Oxford, UK, 1st edn., 2014.

[6] Robert B. Brandom, *Making It Explicit: Reasoning, Representing, and Discursive Commitment*, Harvard University Press, Cambridge, MA, 1994.

[7] Rodney Brooks, 'Intelligence without representation', *Artificial Intelligence*, **47**, 139–159, (1991).

[7] Simondon [59] maintains that indviduals come about as a part of an ongoing process of individuation, and, significantly for our purposes, that individuation itself is mediated by technology: technology in general serves the purpose of permeating the threshold between the subjective and the objective, between the cognitive agent and the environment that serves as the object of cognition.

[8] Joanna J. Bryson, 'Robots should be slaves', in *Close engagements with artificial companions: key social, psychological, ethical and design issues*, ed., Yorick Wilks, 63–74, John Benjamins Publishing Company, Netherlands, (2010).

[9] Angelo Cangelosi and Thomas Riga, 'An embodied model for sensorimotor grounding and grounding transfer: Experiments with epigenetic robots', *Cognitive Science*, **30**(4), 673–689, (2006).

[10] Angelo Cangelosi and Francesca Stramandinoli, 'A review of abstract concept learning in embodied agents and robots', *Philosophical Transactions of the Royal Society B: Biological Sciences*, **373**(1752), 1–6, (2018).

[11] Robyn Carston, 'Metaphor: Ad hoc concepts, literal meaning and mental images', *Proceedings of the Aristotelian Society*, **110**(3), 297–323, (2010).

[12] Noam Chomsky, *Aspects of the Theory of Syntax*, The MIT Press, Cambridge, 1965.

[13] Andy Clark, 'Language, embodiment, and the cognitive niche', *Trends in Cognitive Sciences*, **10**(8), 370–374, (2006).

[14] Stephen Clark, 'Vector space models of lexical meaning', in *The Handbook of Contemporary Semantic Theory*, eds., Shalom Lappin and Chris Fox, 493–522, Wiley-Blackwell, (2015).

[15] Ann Copestake and Ted Briscoe, 'Semi-productive polysemy and sense extension', *Journal of Semantics*, **12**, 15–67, (1995).

[16] William Croft and D. Alan Cruse, *Cognitive Linguistics*, Cambridge University Press, 2004.

[17] Elena Clare Cuffari, Ezequiel Di Paolo, and Hanne De Jaegher, 'From participatory sense-making to language: there and back again', *Phenomenology and the Cognitive Sciences*, **14**(4), 1089–1125, (2015).

[18] Donald Davidson, 'Knowing one's own mind', *Proceedings and Addresses of the American Philosophical Association*, **60**(3), 441–458, (1987).

[19] Vivian De La Cruz, Alessandro Di Nuovo, Santo Di Nuovo, and Angelo Cangelosi, 'Making fingers and words count in a cognitive robot', *Frontiers in Behavioral Neuroscience*, **8**, 13, (2014).

[20] Terrence W. Deacon, *Incomplete Nature: How Mind Emerged from Matter*, W. W. Norton & Company, New York, NY, 2011.

[21] Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman, 'Indexing by latent semantic analysis', *Journal for the American Society for Information Science*, **41**(6), 391–407, (1990).

[22] Daniel Dennett, *Kind of Minds*, Weidenfeld & Nicolson, London, 1996.

[23] Vyvyan Evans, *How Words Mean: Lexical Concepts, Cognitive Models, and Meaning Construction*, Oxford University Press, 2009.

[24] Daniel L. Everett, 'Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language', *Current Anthropology*, **46**(4), 621–646, (2005).

[25] Jerry A. Fodor, *The Language of Thought*, Harvard University Press, Cambridge, MA, 1975.

[26] Michael Franke and Gerhard Jäger, 'Bidirectional optimization from reasoning and learning in games', *Journal of Logic, Language and Information*, **21**(1), 117–139, (2012).

[27] M. Fujita, Y. Kuroki, T. Ishida, and T. T. Doi, 'Autonomous behavior control architecture of entertainment humanoid robot sdr-4x', in *International Conference on Intelligent Robots and Systems*, volume 1, pp. 960–967, (2003).

[28] Bruno Galantucci and Simon Garrod, 'Experimental semiotics: A review', *Frontiers in Human Neuroscience*, **5**, 1–15, (2011).

[29] Raymond W. Gibbs Jr., *The Poetics of Mind*, Cambridge University Press, 1994.

[30] James J. Gibson, *The Ecological Approach to Visual Perception*, Houghton Miffline, Boston, 1979.

[31] Daniel Hernández García, Samantha Adams, Alex Rast, Thomas Wennekers, Steve Furber, and Angelo Cangelosi, 'Visual attention and object naming in humanoid robots using a bio-inspired spiking neural network', *Robotics and Autonomous Systems*, **104**, 56–71, (2018).

[32] Kris Jack, Chris Reed, and Annalu Waller, 'A computational model of emergent simple syntax: Supporting the natural transition from the one-word stage to the two-word stage', in *Proceedings of the Workshop on Psycho-Computational Models of Human Language Acquisition*, (2004).

[33] Mark Johnson, *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*, University of Chicago Press, 1990.

[34] Stuart A. Kauffman, *At Home in the Universe: The Search for the Laws of Self-Organization and Complexity*, Oxford University Press, 1995.

[35] George Lakoff, *Women, Fire, and Dangerous Things*, University of Chicago Press, 1987.

[36] George Lakoff and Mark Johnson, *Metaphors We Live By*, University of Chicago Press, 1980.

[37] Angeliki Lazaridou, Karl Moritz Hermann, Karl Tuyls, and Stephen Clark, 'Emergence of linguistic communication from referential games with symbolic and pixel input', in *International Conference on Learning Representations*, (2018).

[38] Angeliki Lazaridou, Alexander Peysakhovich, and Marco Baroni, 'Multi-agent cooperation and the emergence of (natural) language', in *International Conference on Learning Representations*, (2017).

[39] Humberto Maturana and Francisco Varela, *The Tree of Knowledge*, Shambhala, Boston, MA, 1987. Translated by Robert Paolucci.

[40] Stephen McGregor, Matthew Purver, and Geraint Wiggins, 'Metaphor, meaning, computers and consciousness', in *8th AISB Symposium on Computing and Philosophy*, (2015).

[41] Giorgio Metta, Giulio Sandini, David Vernon, Lorenzo Natale, and Francesco Nori, 'The icub humanoid robot: An open platform for research in embodied cognition', in *Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems*, pp. 50–56, (2008).

[42] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean, 'Efficient estimation of word representations in vector space', in *Proceedings of ICLR Workshop*, (2013).

[43] Alva Noë, *Action in Perception*, The MIT Press, Cambridge, MA, 2004.

[44] J. Kevin O'Regan and Alva Noë, 'A sensorimotor account of vision and visual consciousness', *Behavioral and Brain Sciences*, **24**, 939–1031, (2001).

[45] J.K. O'Regan, *Why Red Doesn't Sound Like a Bell: Understanding the Feel of Consciousness*, Oxford University Press, USA, 2011.

[46] Howard H. Pattee, 'The physics of symbols: Bridging the epistemic cut', *Biosystems*, 5–21, (2001).

[47] Jeffrey Pennington, Richard Socher, and Christopher D. Manning, 'GloVe: Global vectors for word representation', in *Conference on Empirical Methods in Natural Language Processing*, (2014).

[48] Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer, 'Deep contextualized word representations', in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 2227–2237, (2018).

[49] Steven Pinker, *The Language Instinct: How the Mind Creates Language*, William Morrow, 1994.

[50] Alan Prince and Paul Smolensky, *Optimality Theory: Constraint Interaction in Generative Grammar*, Wiley, 2008.

[51] James Pustejovsky, *The Generative Lexicon*, MIT Press, Cambridge, MA, 1995.

[52] Karolina Rataj, Anna Przekoracka-Krawczyk, and Rob H J Van der Lubbe, 'On understanding creative language: The late positive complex and novel metaphor comprehension', *Brain Research*, **1678**, 231–244, (2018).

[53] Joanna Rączaszek-Leonardi, 'Language as a system of replicable constraints', in *Laws, Lanuage and Life*, eds., Howar Hunt Pattee and Joanna Rączaszek-Leonardi, 295–333, Springer, (2012).

[54] Joanna Rączaszek-Leonardi and Iris Nomikou, 'Beyond mechanistic interaction: Value-based constraints on meaning in language', *Frontiers in Psychology*, **6**(1579), (2015).

[55] Joanna Rączaszek-Leonardi, Iris Nomikou, Katharina J. Rohlfing, and Terrence W. Deacon, 'Language development from an ecological perspective: Ecologically valid ways to abstract symbols', *Ecological Psychology*, **30**(1), 39–73, (2018).

[56] Richard Rorty, *Philosophy and the Mirror of Nature*, Princeton University Press, 1979.

[57] Bertrand Russell, 'On denoting', *Mind*, **14**(56), 479–493, (1905).

[58] John R. Searle, *Intentionality: An Essay in the Philosophy of Mind*, Cambridge University Press, 1983.

[59] Gilbert Simondon, *On the Mode of Existence of Technical Objects*, Univocal Publishing, 1958/2017. Trans. by Cécile Malaspina and John Rogove.

[60] Aaron Sloman, 'The structure of the space of possible minds', in *The Mind and the Machine: Philosohpical Aspects of Artificial Intelligence*, eds., Steve Torrance and Ellis Horwood, 35–42, (1984).

[61] Kenny Smith, Henry Brighton, and Simon Kirby, 'Complex systems in language evolution: The cultural emergence of compositional structure', *Advances in Complex Systems*, **6**(4), 537–558, (2003).

[62] Dan Sperber and Deirdre Wilson, *Relevance: Communication and Cognition*, Blackwell, 2nd edn., 1995.

[63] Matthew Spike, Kevin Stadler, Simon Kirby, and Kenny Smith, 'Minimal requirements for the emergence of learned signaling', *Cognitive Science*, **41**(3), 623–658, (2017).

[64] Michael Spranger and Luc Steels, 'Discovering communication through ontogenetic ritualisation', in *4th International Conference on Development and Learning and on Epigenetic Robotics (ICDL-EPIROB)*, pp. 14–19, (2014).

[65] Luc Steels, 'The emergence of grammar in communicating autonomous robotic agents', in *ECAI 2000, Proceedings of the 14th European Conference on Artificial Intelligence, Berlin, Germany, August 20-25, 2000*, pp. 764–769, (2000).

[66] Luc Steels, *The Talking Heads experiment: Origins of words and meanings*, Computational Models of Language Evolution, Language Science Press, 2015.

[67] Luc Steels and Tony Belpaeme, 'Coordinating perceptually grounded categories through language: A case study for colour', *Behavioral and Brain Sciences*, **28**, 469–529, (2005).

[68] Karla Stépánová, Frederico B. Klein, Angelo Cangelosi, and Michal Vavrecka, 'Mapping language to vision in a real-world robotic scenario', *IEEE Trans. Cognitive and Developmental Systems*, **10**(3), 784–794, (2018).

[69] Eve Sweetser, *From Etymology to Pragmatics: Metaphor and Cultural Aspects of Semantic Structure*, Cambridge University Press, 1990.

[70] Evan Thompson, *Mind in Life*, Harvard University Press, Cambridge, MA, 2007.

[71] Peter D. Turney and Patrick Patel, 'From frequency to meaning: Vector space models of semantics', *Journal of Artificial Intelligence Research*, **37**, 141–188, (2010).

[72] Francisco J. Varela, Evan Thompson, and Eleanor Rosch, *The Embodied Mind*, The MIT Press, Cambridge, MA, 1991.

[73] Peter Wellens, Martin Loetzsch, and Luc Steels, 'Flexible word meaning in embodied agents', *Connection Science*, **20**(2–3), 173–191, (2008).

[74] Deirdre Wilson and Dan Sperber, *Meaning and Relevance*, Cambridge University Press, 2012.

[75] Ludwig Wittgenstein, *Philosophical Investigations*, Basil Blackwell, Oxford, 3rd edn., 1953/1967. trans. G. E. M. Anscombe.

[76] Mario Zarco and Tom Froese, 'Self-optimization in continuous-time recurrent neural networks', *Frontiers in Robotics and AI*, **5**, 96, (2018).

[77] Rolf A. Zwaan, 'Embodiment and language comprehension: Reframing the discussion', *Trends in Cognitive Sciences*, **18**(5), 229–234, (2014).