An Ecological-Enactive Perspective on Language

Jasper C. van den Herik

Talking about Talking

An Ecological-Enactive Perspective on Language

Jasper C. van den Herik

This work was supported by The Netherlands Organisation for Scientific Research (NWO).

Cover design and printing: www.proefschriftenprinten.nl - Print Service Ede

ISBN: 978-94-92679-94-9

Talking about Talking

An Ecological-Enactive Perspective on Language

Praten over praten

Een ecologisch-enactief perspectief op taal

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
Rector Magnificus

Prof.dr. R.C.M.E. Engels

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on
Thursday the 6th of June 2019 at 13:30 hrs

by

Jasper Cornelis van den Herik
born in Utrecht, The Netherlands

Erasmus University Rotterdam

Doctoral Committee:

Promotors:

Prof.dr. F.A. Muller

Prof.dr. J.J. Vromen

Other members:

Prof.dr. J. de Mul

Prof.dr. E. Rietveld

Prof.dr. T.J. Taylor

Contents

List of figures

Acknowledgements

According to Edward Reed, how a child comes to enter her linguistic community is the central question in studying language development. I've always found it useful to view the process of doing a PhD in similar terms, as entering a meshwork of partially overlapping academic communities. While writing this thesis was a solitary activity, the tangible product of this process would have been impossible without all the wonderful people who have welcomed me to their communities over the past years.

First and foremost, I am referring to the philosophical community in Rotterdam. I would like to thank my promotor F.A. Muller for his supporting role, both in the process of writing the proposal, as well as for his sharp comments on the many drafts I produced over the past four and a half years. Without your analytic eye, Fred, this thesis would have surely taken a very different form. Thanks also to Jack Vromen, for being my second promotor. Without him, the process of finishing the PhD wouldn't have gone as smoothly as it did. My gratitude goes out to all my colleagues in Rotterdam, both academic and support staff, for providing a stimulating working environment and providing me with the possibility to share my work in a lunch lecture and faculty colloquium. In particular, I'm indebted to Daan Dronkers and James Grayot, for fruitful discussions and comments on drafts.

A second community I was part of was the project Antwerp-Rotterdam collaboration 'Getting Real about Words and Numbers'. I vividly remember Erik Myin contacting me on Thursday the 27th of March 2014. He came across a grant for a Flemish-Dutch collaboration and saw a possibility for combining Karim Zahidi's rejected application on Enactivism and mathematics with my rejected application on Enactivism and language. The deadline was April first.

What ensued was five days of frantic writing by Karim, Erik, Fred and myself. The fact that we managed to produce a proposal that got funded is a testament to our shared ways of thinking. I would like to thank Erik and Karim, for all the discussions we had in our regular meetings, the conferences and workshops we organised together, as well as their constructive criticisms on drafts and presentations. It has been a pleasure to work together. I would also like to thank Antwerp Enactivists Farid Zahnoun, Victor Loughlin, Zuzanna Rucinska, and Ludger van Dijk for providing a great research environment.

The Centre for Human Interactivity in Odense, Denmark has profoundly shaped the work in this thesis. I have fond memories of visiting the Centre on five occasions over the past years and meeting its members at a variety of other academic events. I thank Stephen Cowley, Sune Vork Steffensen, Theres Fester, Sarah Bro Trasmundi, Christian Mosbaek Johannessen, Thomas Wiben Jensen, Rasmus Gahrn-Andersen, and Alfonso Ramírez, for the welcoming atmosphere, offering me the possibility to share my work, many engaging discussions, and tasty beers in the pubs of Odense. Special thanks go out to CHI's Matthew Harvey, with whom I have had many interesting discussions and who has provided invaluable feedback on my writings.

One CHI event that was of particular importance for this thesis is the symposium in honour of Nigel Love organised in February 2016. It was at this symposium that I first met Talbot Taylor. His work introduced me to metalanguage and persuaded me of its importance in the study of language. I have learned a lot from our extensive correspondence and his careful and generous comments on drafts. I was very happy to receive an invitation from Talbot to write a contribution to Nigel Love's *festschrift*, which ended up being the second chapter of this thesis. It also sparked a thought-provoking correspondence with Nigel, Talbot, and myself, that led to a very interesting workshop on

integrational and ecological approaches to languaging in Rotterdam in June 2018.

The International Society for the Study of Interactivity, Language and Cognition is a community that partially overlaps with the CHI. I thank Joanna Rączaszek-Leonardi and Fred Vallee-Tourangeau for encouraging me to submit an abstract for the special issue connected to the society's conference which I attended in June 2016. This abstract resulted in the third chapter of this thesis. I would like to thank Joanna for her thoughtful and precise comments as editor of this special issue, which helped me tremendously in dealing with the comments I received from the reviewers. Her enthusiasm for an ecological approach to language is infectious, and I have learned a lot from her work. I would also like to express thanks to Joanna and Catherine Read for introducing me to yet another community, the ecological psychologists, by organising and inviting me to symposia at the 2018 EWEP and the 2019 ICPA.

Another community that played a central role in my intellectual development over the past years is the research group of Erik Rietveld in Amsterdam. I want to thank Jelle Bruineberg for organising the regular reading group. I learned a lot from the discussions we had in these meetings. I also have fond memories of the many meetings I had with Julian Kiverstein in De Jaren, during which we discussed several books and articles we had read and which offered me a great opportunity to discuss my drafts in detail and shape my thinking. I'm looking forward tremendously to continue working with Erik and Julian.

Furthermore, I would like to thank the Enactivist Walkers, a loosely organised group consisting of Katja Abramova, Jelle Bruineberg, Carli Coenen, Ludger van Dijk, Sanneke de Haan, Zuzanna Rucinska, and Deva Waal. I hope we will be able to realise Carli's excellent idea of a walking conference some time in the future.

An acknowledgement section like this always fails to be exhaustive. There are many others who have contributed to this thesis. I would like to extend my thanks to the reviewers that took the time to provide me with invaluable comments, and the organisers of the conferences, workshops, and summer-schools I have attended over the past years. Particularly formative for me was the wonderful *Enriching Embodied Cognition* workshop in Istanbul organised by Lucas Thorpe, where I had the chance to discuss my ideas with Daniel Hutto for the first time.

Although this thesis is written, it must still be defended. I would like to thank Menno Lievers, Jos de Mul, Erik Myin, Erik Rietveld, Marc Slors, and Talbot Taylor for being members of the doctoral committee, and Karim Zahidi and Daan Dronkers for being my paranimfen.

I would like to thank my parents Leo van den Herik and Sophie Ouboter, and my sister Isi van den Herik. My interest in language can surely be traced to a house filled with books, an unrelenting propensity for discussion, and a constant correcting of linguistic mistakes – a metalinguistic activity par excellence!

Finally, this thesis could not have been written without Merel Segers. Thank you for your continuing love and support. I'm looking forward to our continuing time together and the initiation of our son Leo into our linguistic community.

1 Introduction

> Language cannot be explained ontologically or by referring to some kind of peculiar organ or device, as linguist Noam Chomsky assumes. Rather, it can only be explained ontogenetically. Language *is*n't, it happens.
>
> – Heinz von Foerster, *Understanding Systems*

Our son Leo is almost six months old. Although he shows a keen interest in producing vocalisations, he cannot speak yet. Over the next couple of years, his vocalisations will gradually turn into *words*, and he will start to do all kinds of things with these words. In doing so, he will become a participant in the linguistic activities of his community. This remarkable transformation raises two important questions: *how* is it possible for children to learn language? And *what* is it that they learn? How you answer these questions depends on what you think language is, *and* on what you think cognition is. Cognition roughly means thinking, but how cognition should be defined and explained is a matter of fierce debate, as we will see in this introduction.

This thesis proposes a perspective on language and its development by starting from two approaches. The first is the *ecological-enactive approach* to cognition. In opposition to the widespread idea that cognition is information-processing in the brain, the ecological-enactive approach explains human cognition in relational terms, as skilful interactions with a sociomaterial environment shaped by practices. A practice is a recurrent activity where participants are accountable to rules or criteria, e.g. making promises, playing chess, congratulating someone on their new job, and so on.[1] The second is the *metalinguistic*

1 I define a recurrent activity as a practice 'iff (1) there are public criteria or rules involved to which participants' behaviour is accountable, (2) there is a point to the activity, (3) and the participants understand the point of the activity and put the activity to use in furthering their own purposes.' (p. 112).

approach to language, which holds that reflexive or metalinguistic language use – talking about talking – is crucial for understanding language and its development. In particular, I defend two theses:

1. A child's initial communicative behaviour can be explained in terms of attentional actions: social actions that function by directing someone else's attention.
2. In order for the child's communicative behaviour to be sensitive to key properties of language, such as semantic content and normativity, she needs to learn metalinguistic skills.

The development of this ecological-enactive perspective on language serves two functions. *First*, the ecological-enactive approach started by considering basic behaviour, such as locomotion and grasping. An approach in the cognitive sciences, however, should be able to account for the full gamut of human cognition. If the perspective developed in this thesis is viable, this is a contribution to extending the ecological-enactive approach to typically human forms of cognition. *Second*, this perspective throws new light on philosophical problems concerning language. In the different chapters, I deal with questions concerning the nature of linguistic knowledge, explanations of communicative behaviour, and the origins of semantic content and linguistic normativity.

This introduction serves to contextualise the perspective developed in this thesis. In the first part of this introduction, I introduce the *pragmatic turn* in the study of cognition. This turn marks the current shift from *classical cognitivism*, according to which all cognition consists essentially in brain-bound computations of mental representations, to an action-oriented paradigm, which redefines cognition as perceptually guided action. I first introduce classical cognitivism and some arguments against this approach, after which I introduce the

ecological-enactive approach on which this thesis builds. In the second part of the introduction, I introduce a widespread assumption in theoretical reflections on language, which is to take language to be an abstract system of codes. On this *code-view*, a language encodes thought and thereby allows for the transference of thoughts from one person to another. I show why this view of language goes hand in hand with a classical cognitivist account of cognition. I then introduce the metalinguistic approach, according to which language should be explained in terms of normative metalinguistic practices. In the third and final part of this introduction, I provide a short summary of each chapter.

1.1 Cognition: from representation to action

The cognitive sciences are currently making a *pragmatic turn* (Engel et al. 2013). In this paradigm shift (Stewart, Gapenne, Di Paolo 2010), the very concept of cognition is redefined. Classical cognitivism, the dominant approach since the middle of the twentieth century, defined cognition as computations over mental representations, a process that is tucked away inside the skull and is only instrumentally related to perceptual inputs and motor outputs. Gradually, this conception is making way for a radical embodied concept of cognition, in which action takes centre stage. Echoing Ryle, intelligence is not located in the causes of behaviour; instead, behaviour itself is performed intelligently. Human behaviour is explained in relational terms, as emerging from a history of skilful interactions with a sociomaterial environment.

In what follows, I briefly introduce classical cognitivism and some reasons for abandoning it. An alternative to classical cognitivism are E-approaches to cognition, according to which cognition is embodied, embedded, extended, enacted, and ecological. There are, however, many different E-approaches to cognition. Some of these are conservative, in that they

are best understood as variants of classical cognitivism. I introduce two of these conservative E-approaches. I end this section by introducing the ecological-enactive approach, the radical embodied approach on which this thesis builds, which is *not* a variant of classical cognitivism but opposed to it.

1.1.1 Classical cognitivism: the classical sandwich model of mind

Classical cognitivism was a response to the behaviourism of the first half of the twentieth century. The aim of classical cognitivism was to provide a realist understanding of mentalistic terms in the cognitive sciences. Without a doubt, one of the most ardent proponents of classical cognitivism was Fodor. He uses the term *representational theory of mind* (RTM) to designate this approach to the study of mind. He describes its commitments as follows:

> 1. The only psychological models of cognitive processes that seem even remotely plausible represent such processes as computational.
> 2. Computation presupposes a medium of computation: a representational system. (Fodor 1975, p. 27)

> RTM is a claim about the metaphysics of cognitive mental states and processes: Tokens of cognitive mental states are tokens of relations between creatures and their mental representations. Tokens of mental processes are 'computations'; that is, causal chains of (typically inferential) operations on mental representations. (Fodor 2008, p. 5)

This definition of cognitive states and processes exemplifies the two pillars of classical cognitivism: computation and representation. Cognition consists essentially in constructing mental representations of the world 'outside' the cognitive system and to perform computational operations on them. Traditionally, all computational systems were taken to be representational systems, as epitomised by the slogan 'no computation without representation' (Fodor 1981, p. 122). The guiding idea was that computations are, at least in part, individu-

ated by semantic properties of the representations they are computations on.[2] Recently, however, people have proposed non-semantic accounts of computation (e.g. Piccinini 2008; Villalobos & Dewhurst 2017). If these accounts are viable, a non-representationalist account of cognition could be computationalist. For the purposes of this thesis, I will not further discuss the notion of computation, but focus instead on the representational pillar.

In classical cognitivism, mental representations are defined as amodal symbols: symbols that have a non-perceptual representational format, and that therefore bear no systematic similarity to what is represented. Mental representations were taken to be sentences in a language of thought called *mentalese,* which means that mental representations are propositionally structured and descriptive in nature (Fodor 1975, 2008; Devitt 2006; Pinker 1994, 2007).

The argument for representationalism is given by Marr (1982/2010, p. 3) as follows:

> if we are capable of knowing what is where in the world, our brains must somehow be capable of representing this information – in all its profusion of color and form, beauty, motion, and detail. The study of vision must therefore include not only the study of how to extract from images the various aspects of the world that are useful to us, but also an inquiry into the nature of the internal representations by which we capture this information and thus make it available as a basis for decisions about our thoughts and actions.

One important implication of cognitivism's representationalism is that a clear distinction can be made between those processes that are cognitive and those that are not. Perception and action are merely the inputs and outputs to a cognitive system, but are not part of it. In Marr's quote, this can be gleaned

2 Fodor (*personal correspondence* cited in Piccinini 2008, p. 235), explains why computation must be individuated by semantic properties: 'What distinguishes […] causal sequences that *constitute computations* from those that don't? Answer, the former preserve *semantic properties* of the strings (paradigmatically, they take one from true inputs to true outputs). This requires that the tokened states have semantic interpretations (since, of course, only what is semantically interpreted can be evaluated for truth). So, in that sense, *the representations in question are individuated by their semantic properties inter alia*.'

from the idea that vision consists essentially in *extracting* from *images* those aspects of the world that are useful for planning a course of action. The idea that what is available to vision is images, not a world, comes from starting from the light projected on the retina and calling this projection the retinal image. Importantly, the information that reaches the retina is thought to be ambiguous and impoverished. For example, the retinal image is two-dimensional, yet we perceive a three-dimensional world. Cognitive processes are therefore thought 'to infer' representations of the world, based on this ambiguous information.

Hurley (2001) describes this view of cognition as the *classical sandwich model of cognition.* Cognitive processes are sandwiched between peripheral processes of perception and action (see figure 1). Sensory input is processed and transformed into mental representations, which are then sent to the central cognitive core in the brain. Based on these representations, the central cognitive core decides on a course of action and sends instructions for performing these actions to the motor system. This view implies a one-way causal flow: information enters the system through the senses, gets processed, and this in turn results in certain actions. Moreover, the process is modular: sensory, cognitive, and action processes are distinct from one another. The sandwich is *classical* when cognitive processes are thought to consist in performing computations over mental representations.

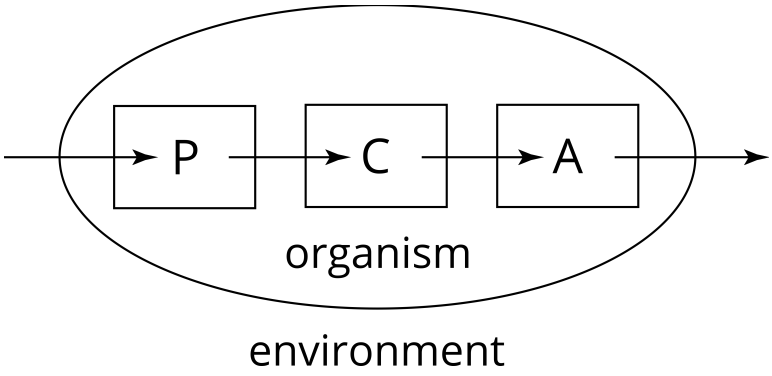

Figure 1: The classical sandwich model of cognition.

1.1.2 Against sandwiches

Two main lines of criticism have been levelled against the modularity that follows from a classical sandwich model of cognition. The first stems from a wealth of empirical evidence that shows that the brain cannot be neatly compartmentalised into distinct perception, cognition, and action modules. Perhaps the most well-known example is given by *mirror neurons*. These neurons, as first described in monkeys by Di Pellegrino et al. (1992), are part of the 'premotor' cortex. They are activated in the performance of particular hand movements but are also activated when monkeys observe the same hand movement without doing anything. What were initially thought to be regions dedicated to motor processes thus turn out to play a role in perception as well. Moreover, there is now a wealth of evidence that the actions one is currently engaged in modulate and bias sensory activations (for an overview, see Engel et al. 2013). Acquiring new behaviours and performing certain actions can have modulatory effects on sensory areas in the brain. For example, in monkeys that learned to use a rake to gather food that would have otherwise been out of

reach, bimodal neurons that initially responded to somatosensory stimulation of the hand and visual stimulation near the hand extended their receptive fields to include the tool (Maravita & Iriki 2004). The action of using a tool thus changes perceptual processes. This has been argued to show that perceptual processes cannot be described without reference to the action context, and therefore cannot be thought to be isolated from action in the way suggested by the classical sandwich model.

The second line of criticism is more fundamental. Marr describes the explanatory problem posed by vision in terms of the extraction of useful aspects of the world from images, where the image is understood as arising from the projection of light on the retina. Here the perceiver is characterised as a passive recipient of sensory information from the environment. This passivity of the perceiver is not only assumed, but also enforced in experiments, in which subjects typically look at two-dimensional computer monitors (not three-dimensional objects)[3] while their heads are restrained. In the real world, however, we are not typically passive in this way. Instead, we actively explore our environments *in order to* be able to see what we need to see. Noë (2012, p. 40) describes this as follows:

> perceptual experience is an activity whereby we bring things near us into focus for perceptual consciousness. We peer, and squint, and move, and adjust ourselves, nearly continuously, in order to come near to, achieve access to and stabilize our contact with the world around us.

Another example is given for haptic perception of a sponge by Myin (2003, p. 42):

3 Snow *et al.* (2011) show that a 'repetition suppression', i.e., reduced firing rates in certain brain areas when an object is repeatedly shown to a person, only holds for watching 2D images. They could not find this effect when people are repeatedly shown the same 3D object. What had been assumed to be a property of perception *simpliciter*, following the idea that we perceive *images* anyway, turns out to be a property of 2D image perception.

> the experience of softness comes about through a specific pattern encountered in a sensorimotor exploration, including such facts as that if one pushes on a soft object, it yields.

Perceiving the softness of the sponge relies constitutively on acting on the sponge by pressing or squeezing it. If you take away the exploratory action in the perception of softness you are not left with 'pure perception'; rather, you would not be able to perceive the softness of the sponge at all.

What these examples show, is that action is a constitutive part of perception, and not merely a consequence of it. Action and perception are inextricably intertwined, not only at the level of the brain but also at the level of the organism in its environment. However, what this conclusion entails for the study of cognition is a matter of debate.

1.1.3 Conservative E-approaches

Over the last thirty years, a group of approaches to the study of cognition has gained in popularity according to which cognition is embodied, embedded, extended, enactive, ecological. These E-approaches have very different theoretical commitments. Some E-approaches, for example, are best understood as variants of the cognitivist framework, whereas others propose a radical break.[4]

On the conservative end of the spectrum, we find a concept of 'embodied cognition' that is nevertheless brain-bound. On these accounts, an 'embodied concept [i.e., a mental representation] is a neural structure that is actually part of, or makes use of, the sensorimotor system of our brains' (Lakoff & Johnson, 1999, p. 104). The idea is that cognition does not consist in the manipulation of amodal symbols, as proposed by classical cognitivism, but rather that modality specific systems underlie cognition. For example, some cognitive scientists argue that besides descriptive amodal symbols, mental representations can also

4 In a trivial sense, all cognition is embodied. Nobody in the cognitive sciences proposes a substance dualism in which cognition exists independent of *any* physical realisation.

be depictive and have a pictorial format (e.g. Pearson & Kosslyn 2015). Barsalou's (1999, p. 578) *Perceptual Symbol System* is a good example of this idea: 'Subsets of perceptual states in sensory-motor systems are extracted and stored in long-term memory to function as symbols.' In other words, on these views 'cognition is embodied in the sense that the mechanisms for perception and action are the same as the mechanisms for concept manipulation and reasoning' (Aizawa 2015, p. 758).

These approaches deviate from classical cognitivism in that they deny the fact that perception and action are peripheral processes that can be understood in separation of the central cognitive core. In this way, they aim to accommodate the empirical evidence that shows that, on the level of the brain, no clear distinction between perception, cognition, and action can be made. However, they do not break with the cognitivist assumption that cognition should be understood as involving mental representations. These approaches are therefore best understood as slight modifications of classical cognitivism, in that they allow for other representational formats besides the amodal symbols originally envisaged.

Another E-approach that is best understood as a slight modification on the classical cognitivist framework is the idea that cognition is sometimes extended beyond the brain. This idea, known as the extended mind hypothesis, was first defended by Clark and Chalmers (1998), who introduced the scenario of Otto and Inga who both wanted to visit a museum. Where Inga uses her biological memory to remember where the museum is, Otto suffers from Alzheimer's and therefore cannot remember the location. In order to counteract this, he carries a notebook with him in which he records information he would otherwise forget. In order to 'remember' the location of the museum, he must consult his notebook. Clark and Chalmers (1998, p. 13) think that in this

example, we have good reason to assume that Otto's notebook is a genuine part of his cognitive system:

> For in relevant respects the cases are entirely analogous: the notebook plays for Otto the same role that memory plays for Inga. The information in the notebook functions just like the information constituting an ordinary non-occurrent belief; it just happens that this information lies beyond the skin.

Besides the idea that cognitive processes sometimes extend beyond the brain, the extended mind hypothesis is perfectly in line with cognitivist assumptions. This does not mean that all cognitivists agree to the conclusions of Clark and Chalmers. One important argument against the extended mind is to argue that whereas it is indeed the case that cognition is causally dependent on all kinds of things, this does not show that cognition is *constitutively* dependent on it. Adams and Aizawa (2010) call this the *coupling-constitution fallacy*: not everything that is causally coupled to a cognitive system thereby becomes part that cognitive system. In their view, the sandwich model of mind can be saved because of its classical fillings:

> a principled basis for thinking that today, cognitive processes typically occur only within the brain, or central nervous system, is in plain view. It lies in the familiar cognitivist view that cognition involves certain sorts of manipulations of non-derived representations. (Adams & Aizawa 2010, p. 579).

1.1.4 Against classical fillings: the hard problem of content

Whether mental representations exist is a hotly debated issue with a long pedigree in the philosophy of mind and cognition (see e.g. Brooks 1991; Varela, Thompson, & Rosch 1991; Keijzer 2001; Ramsey 2007; Chemero 2009; Rowlands 2015). Here I want to present a recent argument against representationalism put forward by Hutto & Myin (2013), which they call *the hard problem of content*.

In order to understand this problem, it is important to realise that a philosophically robust notion of representation requires that representations have *semantic properties* (Pitt 2018).[5] The idea is that a mental representation represents something as being a particular way, where that thing need not be that particular way. This idea is captured in the concept of *content* (Hutto & Myin 2017; Neander 2017; Rescorla 2016). Having content means having correctness conditions of some kind. Imagine for example that an object is mentally represented as *red*. This mental representation is correct when the represented thing is indeed red, and is incorrect otherwise. Traditionally, following the propositional nature of mental representations, these correctness conditions were taken to be *truth* conditions, or, in the case of desires or goals, *satisfaction* conditions. In the case of other representational formats, such as the sensorimotor representations of conservative embodied cognition, these correctness conditions can also be *accuracy* conditions.

A representational theory of cognition is only viable if an account can be given of how mental representations get their content. Content is essentially a normative concept (Cash 2008). Rowlands (2017, p. 4217) explains:

> cognitive states—thoughts, beliefs, etc.—make a *normative* claim on the world. If I have a belief with the content that *p*, then the world *should* be *p*. If the world is not *p* then something has gone wrong.

To give a naturalistic account of content is thus to give an account of the place of semantic normativity in the natural world.

According to Hutto and Myin (2013), proponents of a representational theory of cognition must face up to the *hard problem of content*. They start from the observation that the prime candidate for naturalising content in the contem-

5 It is a matter of some debate whether cognitive scientist's use of the word *representation* requires this philosophically robust notion of representation. In particular, it is unclear whether the explanatory success of cognitive science theories that use the word *representation* hinges on a robust notion of representation (e.g. Hutto & Myin 2013, pp. 113ff.; Ramsey 2007).

porary debate is the concept of information. However, the only scientifically credible form of information depends on the notion of covariance.[6] This is also the notion of information that philosophers rely on: when a perceived state of affairs, or signal, correlates to a distal state of affairs, it is thought that the signal carries information about the state of affairs.

A concept of information-as-covariance can explain what actually caused a particular behaviour in an organism. But, by itself, this covariance relation is not the right kind of relation to constitute content. For one, whereas covariance relations are symmetric, contentful relations are not (Hutto 2008, p. 48). A map represents the territory, but the territory does not represent the map. Moreover, the failure of covariation does not imply inaccuracy, falsity, or unsatisfied conditions of satisfaction. Although smoke usually co-varies with fire, the smoke that emanates from a smoke machine doesn't falsely indicate the presence of fire. This means that the concept of information, by itself, is not going to be able to do the heavy lifting in giving a naturalistic account of content (cf. Van den Herik 2013). Crucially, for the constitution of content we need something more: we do not need a descriptive account of what actually caused the tokening of a mental representation, we need a normative account of what *should* cause the tokening such that in some cases a representation is caused by something it *should not* have been caused by. In other words, to explain representational properties is to explain the possibility of *mis*representation.

Over the years, many explanations of mental content have been proposed (e.g. Harman 1973; Dretske 1981; Block 1986; Fodor 1990; Papineau 1984). The most promising of these is generally taken to be *teleosemantics*, as initially proposed by Millikan (1984). Teleosemantics tries to extract the normativity of representations from evolved biological functions. It goes beyond covariance because it assumes that 'the content of a representation is determined, in a very

6 See Miłkowski (2015) for a critique of the assumption that covariance is the only naturalistically respectable notion of information.

important part, by the systems that interpret it' (Millikan 2005, p. 100). The general idea is that sensory mechanisms produce mental representations that are 'consumed', as Millikan calls it, by other parts of the cognitive system. The content is fixed by the evolved biological function that a representation fulfils in enabling the consumer mechanism to direct the behaviour of the organism. For example, a mental representation of a fox will cause a rabbit to flee. In normal conditions, the mental representation of the fox will be caused by the presence of a fox. However, it can also be caused by other things in the environment, for example, something that looks remarkably like a fox, and will then also cause the rabbit to flee. In these cases, the teleosemanticist argues, there is a *mis*representation, because the mental representation and consumer mechanism's evolved function is fleeing from foxes, not fleeing from things-that-look-like-foxes. In other words, 'the content is that condition under which the resulting behaviour *would* be appropriate, whether or not the actual circumstances that caused the representation are of that type' (MacDonald & Papineau 2006, p. 6).

As Hutto and Myin (2013) point out, it is commonly assumed that these teleosemantic theories fail. The root problem, Fodor (1990) argues, is that selectionist explanations, i.e., explanations in terms of natural selection, are extensional. In the example of the rabbit fleeing from a fox, the fleeing behaviour's function is to escape from the fox. And so a selectionist explanation can show that the function of the mental representation causing the fleeing behaviour is to flee from this fox. But for explaining representational content, this is not enough. The reason is that the object of the evolved biological function can be described in many different ways: as a fox, as a predator, as dangerous, and so on. For the purposes of fulfilling the biological function, determining the description of the fox is irrelevant. As long as the rabbit flees from foxes, the function is fulfilled. However, the description is very relevant to determining

representational content, for each of the different descriptions has different correctness conditions.

In other words, in order to be able to talk of representational content, the biological function must not only determine in response to *what*, extensionally speaking, the mental representation evolved, it must also determine an intensional description of that thing. Hutto and Myin (2013, p. 80) summarise: 'Even if we can specify what is meant to be targeted that would give us exactly no reason to think that the targeted item is represented in a truth-conditional, referential, or otherwise semantic way'.

Biological functions can fail to fulfil their function. The rabbit might flee from something that is not a fox, or fail to flee from a fox. In these cases, some of the internal states of the rabbit might be the same as when it was fleeing from a fox. However, none of this shows that the rabbit represents anything, correctly or incorrectly. In the words of Burge (2010, p. 301), there simply is 'a root mismatch between representational error and failure of biological function.'

The hard problem of content is particularly hard when a cognitivist has to solve it. The reason for this is that the cognitivist holds that all cognition has to be explained in terms of content-bearing mental states. This means that the normativity of content cannot be explained in terms of cognitive processes, and therefore must be explained in purely biological or physical terms. The hard problem of content, however, looks very different when approached from a non-representational approach, e.g. the ecological enactive-approach. By providing a contentless account of some forms of cognition, including those forms underlying basic social interaction, the theoretical resources available for solving the hard problem are much more suited to the task at hand. In chapter 4, I call this the *ecological-enactive explanatory reversal*: on the ecological-enactive approach, content is not needed to explain cognition, but instead, content can be explained by cognition, including social forms of cognition.

Others, including an earlier time-slice of myself (Van den Herik 2014), assumed that Hutto and Myin's hard problem of content entails that the concept of content should be eliminated from our theoretical reflections on human behaviour altogether (e.g. Harvey 2015, Rosenberg 2015, Alksnis 2015). This conclusion is warranted given the assumption that the only viable understanding of content is a realism where contents are understood as abstract objects such as propositions.[7] On such a realist view, allowing for content in one's explanations would amount to adopting a non-naturalist position. Rosenberg (2015) explains: 'Naturalism cannot help itself to causal contact with abstract objects, and that's the only kind of contact there is.' However, a semantic nihilism of this kind does not seem to be able to do justice to our use of language. As Lance (2017, p. 163) says, 'Even the most semantically nihilistic philosopher must admit that there is something pragmatically significant to the question of whether a given pair of sentences "share content."' For example, in everyday situations, we can judge two different utterances to be 'saying the same thing'. While the word 'content' might not be used in these interactions, the everyday concepts of 'what someone said', or the 'what someone means' express the relevant notion. Viewed in this way, content is a phenomenon that needs to be explained.

Finally, it should be noted that not everybody shares Hutto & Myin's (2013; 2017) pessimism with respect to solving the hard problem of content (see for example Shea 2013; Miłkowski 2015; Mollo 2015). For the purposes of this thesis, I will not further engage this debate. Instead, I will focus on building on the ecological-enactive approach.

7 Pitt (2018, §6) describes this standard position as follows: 'the issue is not how to naturalize content (abstract objects can't be naturalized), but, rather, how to specify naturalistic content-determining relations between mental representations and the abstract objects they express.'

1.1.5 Radical embodied cognition

Radical embodied approaches to cognition propose a root and branch reconceptualisation of cognition. In contradistinction to the conservative approaches discussed earlier, that modify the cognitivist notion of mental representations, radical embodied approaches aim to explain cognition *without* invoking mental representations. Clark (1997, p. 147), who coined the term *radical embodied cognition*, describes it as follows: 'Structured, symbolic, representational, and computational views of cognition are mistaken. Embodied cognition is best studied by means of non-computational and non-representational ideas and explanatory schemes'. Wilson and Golonka (2013, p. 1) explain:

> Embodiment is the surprisingly radical hypothesis that the brain is not the sole cognitive resource we have available to us to solve problems. Our bodies and their perceptually guided motions through the world do much of the work required to achieve our goals, replacing the need for complex internal mental representations.

In this thesis, I build on one radical embodied approach, namely the ecological-enactive approach. This approach combines insights from enactivism and ecological psychology into a unified post-cognitivist framework. In what follows, I first introduce enactivism and ecological psychology. I then sketch the contours of the ecological-enactive approach.

1.1.6 Enactivism

Enactivism is an approach to cognition first introduced by Varela, Thompson, and Rosch (1991, p. 173), who characterise the approach as follows:

> In a nutshell, the enactive approach consists of two points:
>
> (1) perception consists in perceptually guided action and
>
> (2) cognitive structures emerge from the recurrent sensorimotor patterns that enable action to be perceptually guided.

Instead of the classical sandwich model of mind, the enactive approach conceives of cognition in terms of sensorimotor loops: perception and action form two sides of the same process. This focus on sensorimotor loops has another implication to which enactivism owes its name. Perceptual processes are constitutively dependent on the possibility to act. This implies, as Stewart (2010, p. 3) puts it:

> what the world 'is' *for* the organism amounts to neither more nor less than the consequences of its actions for its sensory inputs; this in turn clearly depends on the repertoire of possible actions. Without action, there is no 'world' and no perception.

The world of the organism is thus not pre-given but *enacted* (co-constructed, or *brought forth* (Maturana & Varela 1987)) by its own activity. There are currently three main approaches that fly under the banner of enactivism (for a recent overview see Ward, Silverman, & Villalobos 2017).

Firstly, there is *autopoietic enactivism*, also known as *the enactive approach* (Thompson 2007; Di Paolo 2005; Di Paolo et al. 2010). This approach aims to understand how organisms *qua* biological being enact a domain of significance. Its main assumption is that life and mind are continuous in the sense that the processes that underlie life are the same processes that underlie cognition. In order to continue to exist, an organism has continuously create the possibilities for its own survival. Organisms thus essentially make themselves, i.e., they are *autopoietic*. In the words of Maturana and Varela (1980, p. 79), an autopoietic system 'generates and specifies its own organization through its operation as a system of production of its own components'. In order to maintain its organisation, the organism has certain requirements. As a result of these requirements, otherwise meaningless aspects of the environment become meaningful for the organism. In this way, it enacts a domain of significance. For example, Thompson (2007) describes how, for a bacterium, sucrose

has significance as a nutrient. This significance is a relational property, as it is dependent on properties of the sucrose but also on the metabolism that underlies the bacterium's autopoiesis. Therefore, 'Sucrose has significance or value as food, but only in the milieu that the organism itself brings into existence' (Ibid., p. 258).

The second approach is *sensorimotor enactivism* (Hurley 1998; O'Regan & Noë 2001; Noë 2004, O'Regan 2011). This approach limits itself to explaining the phenomenal properties of perception. The main idea is that perception does not happen in the brain, but is rather something we do (Noë 2004; Myin 2016) based on a particular kind of knowledge: 'vision is a mode of exploration of the world that is mediated by knowledge of what we call sensorimotor contingencies' (O'Regan & Noë 2001, p. 940). Sensorimotor contingencies are 'the structure of the rules governing the sensory changes produced by various motor actions' (Ibid., p. 941). In other words, perception consists in exploring the environment, which is made possible by an implicit knowledge of how our actions affect our sensory states. What we perceive then, is not limited to the stimulation of the retina, nor does it consist in building representations. To take a favourite example of this approach, when we see a tomato, we see a three-dimensional object. This phenomenal experience is explained in terms of the practical knowledge we have of what would happen if we were to walk around the tomato, or rotate it, and so on.

The third approach is *radical enactivism*, or *REC* (Hutto & Myin 2013; 2017). Hutto (2015, p. 1) describes the task REC sets itself as follows:

> The main work of radical enactivism is to RECtify existing varieties of enactivism and other cognate approaches so as to strengthen and unify them into a single collective that can rival classical ways of thinking about mind and cognition.

The primary arguments put forward by REC are thus negative, as evidenced for example by the formulation of the hard problem of content. Rather than offer an alternative to other variants of enactivism and enactivism-friendly approaches, REC aims to RECtify other approaches, i.e., to rid them of any inadvertent vestiges of representationalism. For example, Hutto and Myin (2013) have argued that sensorimotor enactivism's claim that that vision is *mediated by knowledge* of sensorimotor contingencies could lead to a representational reading. Another example of RECtification is REC's recreation of teleosemantics without the semantics. According to what REC calls *teleosemiotics*, Fodor's worry that selectionist explanations are extensional can be accommodated by showing how evolved biological functions can account for a basic kind of intentionality, called *Ur-intentionality*, which is non-representational (Hutto & Myin 2017).

1.1.7 Ecological psychology

Like enactivism, ecological psychology starts from the idea that cognition and perception need to be understood in action-oriented terms.[8] E.J. Gibson (1988, p. 5) explains:

> We don't simply see, we look. The visual system is a motor system as well as a sensory one. When we seek information in an optic array, the head turns, the eyes turn to fixate, the lens accommodates to focus, and spectacles may be applied and even adjusted by head position for far or near looking.

To look is to act. Organisms do not passively wait for useful information to be brought in through sensory channels. Instead, ecological psychology foregrounds the fact that organisms engage in exploratory activity, and that this activity is a constitutive part of the perceptual system. Perceptual systems are described in terms of perception-action cycles.

8 Lobo, Heras-Escribano, and Travieso (2018) provide a good introduction to the history and philosophy of Ecological Psychology.

A key insight of ecological psychology is that once you factor in exploratory activity, the information that is available to the organism is very different than cognitivists assume. According to cognitivism, information about the world is ambiguous and therefore needs to be enriched by cognitive processes. The example we discussed was the information projected on the retina, which leads to the idea of the retinal image. On the ecological approach, however, the assumption is that all the information the organism needs to guide its action is there already. Ecological information is described as invariant patterns in energy arrays. For example, while an object's shape might be ambiguous when visually perceived under a certain angle, this ambiguity can be resolved by further sampling the optical array, that is, by moving around the object. In this way, the *activity* of the organism resolves potential ambiguity.

Learning about the world is conceived as a process of 'differentiating previously vague impressions', not of 'enriching meagre sensations' (J.J. Gibson & E.J. Gibson, 1955, p. 34). Adolph and Kretch (2015, p. 131) give an example:

> A good, intuitive example of increased specificity through differentiation is wine tasting; […] Over years of practice, small differences – differences that truly exist in the chemical signatures of different wines – become more easily distinguished; an expert taster can identify a Bordeaux from France and distinguish it from a similar mix of Cabernet Sauvignon, Merlot, and Cabernet Franc grapes from California.

Learning wine tasting does not consist in becoming able to add something to the information available, that is, in inferring characteristics of the world based on incomplete and ambiguous information; rather it consists in becoming able to differentiate between different sources of information that are out there. If there was no chemical difference between different sort of wines, wine tasters would not be able to differentiate between them.

This perceptual learning is conceptualised as *fine-tuning* an organism to the relevant information (E.J. Gibson & Spelke, 1983). Through exploratory activity, the organism can pick up on these differences that are out there, and gradually attune to them. In the case of the human, perceptual learning often occurs under the guidance of somebody who is already attuned to the environment. Here we speak of the *education of attention*, which can be thought of as a process of *guided rediscovery* (Ingold, 2001; Zukow-Goldring & Ferko, 1994). In the case of learning to taste wine, for example, an expert wine taster can direct a novice's attention to differences in taste between different kinds of wine, differences that the expert already notices.

The main innovation of ecological psychology is a redescription of the environment that an organism can perceive. The environment of an animal does not coincide with the world as described by physics. It differs from the physical world because it consists on the ecological scale and is determined by characteristics of the animal. In line with the interconnection of perception and action, the environment does not show up for the organism as a collection of objects, but rather as possibilities for action called *affordances*. J.J. Gibson (1979, p. 127) describes this as follows: 'The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill'.

For example, when we approach a ladder we perceive that we can climb it. Whether a person can climb it is determined *both* by the properties of the ladder, e.g. the distance between the steps, and by their action capabilities and physical characteristics. For an adult, different ladders are climbable than for a small child, and for a person who has broken both their legs, no ladders are climbable.

There is a debate in ecological psychology regarding the ontological status of affordances. Turvey et al. (1981) define affordances as dispositional properties of the environment that are complemented by dispositional properties

on the part of animals (a class which includes humans), which they call *effectivities*. Chemero (2009, p. 145), however, argues that dispositions do not fail to actualise under the right circumstances. A sugar cube has a disposition to dissolve, and if we put it in water it *will* dissolve. On the other hand, although I have the ability to climb ladders, I will not always do so when I encounter a ladder. Moreover, my ability to climb ladders does not mean that I will never slip and fall in my ladder climbing abilities. So the exercise of abilities can fail in a sense not easily accommodated by a dispositional account.

Chemero (2009) proposes a relational concept of affordances. On this view, affordances are relations between animals and their environment. More particularly, affordances are relations between abilities and features[9] of the environment. A potential problem of such a relational account of affordances is that defining affordances in relation to individual animals violates the realism espoused by ecological psychology. For example, the education of attention model of learning relies on the affordances being there in advance of an individuals ability to act on that affordance. More generally, ecological psychologists are interested in defining affordances as resources available to animals. The solution to this possible objection is to make a distinction between two different senses of environment, which I discuss in the next section.

1.1.8 The ecological-enactive approach

Ecological psychology and enactivism were initially developed in relative isolation from one another. This isolation can be explained because proponents of both approaches have stressed their incompatibility. Varela, Thompson, and Rosch (1991, p. 204), for example, claim that:

9 Chemero (2009) uses the word *feature* instead of *property* to foreground the fact that affordances are features of situations, not properties of things, because, according to Chemero (2009, p. 140), 'the situation as a whole supports (perhaps demands) a certain kind of action.'

> In a nutshell, then, whereas Gibson claims that the environment is independent [from the perceptually guided activity of the animal], we claim that it is enacted (by histories of coupling). Whereas Gibson claims that perception is direct detection, we claim that it is sensorimotor enactment. Thus the resulting research strategies are also fundamentally different: Gibsonians treat perception in largely optical (albeit ecological) terms and so attempt to build up the theory of perception almost entirely from the environment.

At the same time, ecological psychologists have argued that the enactive approach's idea that an organism *enacts* a world entails a 'solipsistic epistemology with abhorrent social consequences' (Swenson 1992, p. 207).

However, the approaches are closely related. Both are radical embodied approaches, in that they explain cognition without invoking representations. Besides this negative commitment, they also share positive commitments, that can be understood against their shared influence from the pragmatism of Dewey and James. Heras-Escribano (2019, §5.1) describes these commitments as follows:

> perception and action are not separated processes, but two sides of the same continuous and dynamic process; the environment is a constitutive aspect of cognition; cognition is the set of skills that are the product of the history of interactions between an organism and its environment.

Moreover, both approaches take an essentially phenomenological approach by relying on Merleau-Ponty (1945/2012). Cognition should not be understood in the Cartesian way, in terms of a subject that is epistemically separated from the world and therefore has to represent that world in thought, but instead as an active body-subject that can interact with her environment directly.

At the same time, each approach foregrounds a different aspect: where enactivism starts from the organism, ecological psychology starts from the environment (Baggs & Chemero 2018). Recently, it has been argued that the two approaches can be reconciled, and thereby provide a unified post-

cognitivist approach to cognition (Heras-Escribano 2019; Baggs & Chemero 2018; Kiverstein & Rietveld 2018).

Fully integrating these approaches a task that goes beyond the scope of this thesis. For the purposes of this thesis, I have relied on one aspect of this synthesis. This is the idea that we have to do justice *both* to the fact that affordances exist independent of individuals, such that they can be resources available to individuals, *and* to the fact that individuals enact their world (Rietveld & Kiverstein 2014; Baggs & Chemero 2008). Moreover, in order to understand language from an ecological-enactive perspective, a definition of affordances has to take into account that the environment of humans is structured in important ways by our practices, i.e., the regular ways we have of doing things in which we are accountable to rules or criteria.

Rietveld and Kiverstein (2014) follow Chemero in espousing a relational account of affordances. However, in order to countenance the objection that a relational account undermines the objective reality of affordances, and therefore the fact that they are available to individuals as resources, Rietveld and Kiverstein (2014, p. 335) propose a distinction between two levels of description:

> 1. The form of life and the patterns of behavior that make it up (a form of life in which individuals have the *potential* to engage with affordances skilfully); and
>
> 2. A particular individual's actual skilled engagement with an affordance.

Their proposal is to define *affordances* in relation to the form of life, namely as 'relations between aspects of a material environment and abilities available in a form of life' (p. 335). Rietveld and Kiverstein (2014, p. 337) explain:

> We suggest then that affordances are not relative to the abilities of a particular individual who actually perceives or detects the affordance. They have an existence that is relative to the skills available in the practice, or to use our preferred way of formulating this, to abilities available in the form of life.

The notion of form of life that is in play here, at least in the human case, 'comprises a multiplicity of sociocultural practices' (p. 330). The sum total of all affordances so defined is called *the landscape of affordances*.

Every particular individual, based on a history of education of attention, is selectively open to a subset of the landscape of affordances, called the *field of (relevant) affordances*. The notion of a field captures the fact that at any given time, a person is selectively open to a number of different affordances at the same time. What stands out as relevant for an individual in a given situation depends on the concerns of that individual. For example, a water bottle always affords drinking, but this affordance will stand out as relevant when a person is thirsty. But other factors also influence the field of relevant affordances, think for example of emotions (when a person is happy other affordances are relevant than when a person is sad), social setting (a statue in a museum does not afford lifting), and so on.

To replace the notion of representation, J.J. Gibson (1966) proposed to see the nervous system as resonating to information in the environment. On the ecological-enactive approach, the internal states of the organism and its brain are taken to be states of action-readiness (Frijda 2007; Bruineberg, Kiverstein, & Rietveld 2016). A state of action-readiness is somewhere in-between an overt action and an ability. The general idea is thus that perceiving an affordance requires abilities. Actually perceiving an affordance leads to a state of action-readiness, which in turn may lead the organism to act on that affordance, if no other states of action-readiness take precedence (see Bruineberg 2018 for an account along these lines).

Foregrounding the fact that affordances in the human form of life should be defined in terms of practices also entails that the skilful acting on affordances of an individual is subject to normative assessment (Rietveld & Kiverstein 2014, p. 333), for the 'normative standards in terms of which an agent's engage-

ment with affordances is assessed as better or worse come from practices or customs belonging to a form of life' (Ibid., p. 335). In particular, Rietveld and Kiverstein argue that 'language opens up the possibility to be held to account by other people in our community for what we say and do'.[10] In this way, acting on affordances can thus be used to explain 'higher forms' of cognition, in particular, typical human forms of cognition such as language.

1.2 Language: from codes to metalanguage

Classical cognitivism is most straightforwardly combined with a particular view of language and linguistic communication called the *code view* (Harris 1990; Love 2004). The code view is not a theory of language, but rather a set of assumptions that have been very influential in theoretical reflections on language in the Western world over the past two and a half millennia. In this second part of the introduction, I introduce the code view and provide some arguments against it. I then introduce an alternative, the metalinguistic approach, according to which language is constitutively dependent on metalanguage, i.e., our ways of talking *about* it.

1.2.1 The code view

In English, we can say things like *he struggled to get his thoughts across*, *putting a concept into words*, *his words conveyed a sense of urgency*, *the paper contained some good ideas*, and so on. Based on a large number of examples like these, Reddy (1993, p. 170) distils a conceptual framework for how English people in general talk about communication, which he calls the *conduit metaphor*:

10 Besides the kind of normativity that is made possible by language, Rietveld and Kiverstein (2014, p. 326) argue that 'a very basic kind of normativity belongs to the engagement with affordances in particular situations more generally'. In this thesis I only take up the question of the normativity in linguistic practices (see also footnote 24, p. 61)

> (1) language functions like a conduit, transferring thoughts bodily from one person to another; (2) in writing and speaking people insert their thoughts or feelings in the words; (3) words accomplish the transfer by containing the thoughts or feelings and conveying them to others; and (4) in listening or reading, people extract the thoughts and feelings once again from the words.

This lay understanding of the function of language has been hugely influential in theories of the function of language. The linguist Harris (1990) claims that this lay understanding lies at the basis of virtually all Western theoretical reflections on language. He calls this the *language myth* (Harris 1998), which consists of two theses:

1. *the telementation thesis*: linguistic communication consists in transferring thoughts from one person to another person or persons
2. *the determinacy thesis*: languages are synchronic systems of fixed codes, that function by correlating public words with private thoughts

These two theses are interrelated. In particular, the determinacy thesis is implied by the telementation thesis. This can be understood once we realise that the telementation thesis starts from the assumption that thoughts are essentially private, tucked away in the mind, and therefore inaccessible to others. Language enables us to encode thoughts into physical objects, namely sound-waves in the case of spoken language or marks in the case of written language, and these physical objects *can* be made known to others. If, however, these physical words are to do their job of transferring thoughts, we need a system that makes sure that when one person encodes their thoughts into words, another person will be able to decode them. In other words, linguistic communication as telementation is only possible when people use *the same* fixed codes for encoding and decoding their private thoughts into public words. For it is only when people share these codes that the hearer will end up with the same thought the speaker intended to convey.

Throughout Western history we find explicit statements of the code view in theories of language.[11] Here is Hobbes (1651/1989, p. 18):

> The general use of speech, is to transfer our mental discourse, into verbal; or the train of thoughts, into a train of words.

And Locke (1690/2004, p. 363):

> Man, though he have great variety of thoughts, and such, from which others, as well as himself, might receive profit and delight; yet they are all within his own breast, invisible, and hidden from others, nor can of themselves be made appear. The comfort and advantage of society, not being to be had without communication of thoughts, it was necessary, that man should find out some external sensible sign, whereby those invisible ideas, which his thoughts are made up of, might be made known to others. […] Thus we may conceive how *words,* which were by nature so well adapted tot hat purpose, come to be made use of by men, as the *signs of* their *ideas*. […] The use then of words, is to be sensible marks of ideas; and the ideas they stand for, are their proper and immediate signification.

More recently, this idea finds expression in the work of De Saussure, the father of modern linguistics (1922/2011, pp. 11–12):

> In order to separate from the whole of speech the part that belongs to language, we must examine the individual act from which the speaking-circuit can be reconstructed. The act requires the presence of at least two persons; that is the minimum number necessary to complete the circuit. Suppose that two people, A and B, are conversing with each other:
>
> Suppose that the opening of the circuit is in A's brain, where mental facts (concepts) are associated with representations of the linguistic sounds (sound-images) that are used for their expression. A given concept unlocks a corresponding sound-image in the brain; this purely *psychological* phenomenon is followed in turn by a *physiological* process: the brain transmits an impulse corresponding to the image to the organs used in producing sounds. Then the sound waves travel from the mouth of A to the ear of B: a purely *physical*

11 In §3.1.1 (p. 76), I give some more examples (see also Harris 2003, pp. 26–28).

> process. Next, the circuit continues in B, but the order is reversed: from the ear to the brain, the physiological transmission of the sound-image; in the brain, the psychological association of the image with the corresponding concept. If B then speaks, the new act will follow – from his brain to A's – exactly the same course as the first act and pass through the same successive phases

The code view has important implications for the study of language. It entails that linguistic behaviour can only be explained with reference to knowledge of *a* language, understood as an abstract and synchronic system of codes, that is by its very nature invariant across all contexts. The study of language, then, can readily ignore the messy nature of concrete linguistic interaction: what matters is the abstract system. This also means that a clean distinction can be made between those aspects of behaviour that are linguistic, and those that are not. In De Saussure's quote, this is even given as a motivation for describing the speech circuit. Harris (1998) calls this tendency to dissociate language from the behaviour of people *segregationism.*

While it may be intuitively attractive, the code view runs into serious problems when used as an *explanation* of linguistic behaviour. Harris (2003, pp. 32–36) argues that the code view embodies an essential tension. On the one hand, it starts from the idea that linguistic communication is by and large successful, while at the other hand, it entails a crippling scepticism with respect to this success. This scepticism originates in the idea that successful linguistic communication relies on shared knowledge of fixed codes. As is made evident by Locke, words are required as external signs of internal ideas *because* the thoughts of one person are not available to another. However, the same holds for the alleged knowledge of fixed codes. As the Saussure argues, the encoding and decoding processes are also psychological process. This means that they are also unavailable to others. In other words, we have no way to infer, based on the linguistic behaviour itself, what code another person is using; they might be

using a code that is very different from the one we use. For example, although someone might open the window when I ask him to, this does not show that his mental representation associated with the word 'window' is identical to the one I have in mind. Given the code view and the privacy of thought for which it is a solution, people can never verify that they indeed share the same codes, and therefore, they can not know whether communication is ever successful. In the words of Harris (2003, p. 24), 'there is something manifestly awkward about an explanation of any human social activity which leaves the participants theoretically unable to grasp what it is they are doing.'

Besides the scepticism with respect to successful communication, there is a more fundamental problem for the code view. According to the code view, a language is an abstract system consisting of fixed codes that associate forms and meanings. In order for this to be a viable theory, there must be some way of identifying these forms and meanings in a context-invariant way. This, however, turns out to be a complicated affair.

First, let us look at the possibility of identifying linguistic forms. In one sense, this identification seems straightforward: it is undeniable that we experience language as consisting of words that are repeated across contexts. Love (2004, pp. 538–539) explains:

> Somewhere near the root of the language-as-code notion is the idea that when I produce an utterance there is some definite linguistic entity over and above my utterance itself, or which the utterance is a repeatable instance. This idea ultimately derives from the ease with which the possibility of repetition conduces to reifying a repetitum. So an utterance comes to be seen as an utterance *of* an abstract linguistic unit of some kind; and it is this abstract linguistic unit, not the concrete utterance itself, that constitutes the relevant item in the linguistic code.

The idea is straightforward: when we speak, listen, write, or read, we often have no problem identifying the words that are used. Every once in a

while we might come across a word we don't know, but in the majority of cases we will have a strong sense of recognition based on a history in which we have came across the *very same* words already. However, if we want to use (knowledge of) linguistic units in order to *explain* linguistic behaviour, we need a principled way of identifying them. And here purely formal characteristics will not do. Take the example of homonyms: the word *bank*, for example, can be used to talk about a financial institution *and* about the side of a river, and therefore constitutes at least two distinct linguistic forms that encode two different meanings. A proponent of the code view can remedy this initial observation by saying that the codes are of course not one to one mappings between forms and meanings, but rather many to many mappings (Cruse 2017, p. 253). This means that a given form can express more than one meaning. But this solution has an important implication: it makes the identification of form-meaning mappings dependent on the identification of meanings. We can only know that the form *bank* is part of at least two form-meaning codes because we know it can be used with at least two distinct meanings. This in turn means that, in order to ground a code view of language, we need a principled way of identifying meanings.

The fact that linguistic forms, such as words, are easily recognisable as being the same, originates 'the nominalistic illusion that any word does revolve around a firm conceptual core' (Bottineau 2010, p. 283). A far-reaching argument against the possibility of identifying context-invariant meanings, however, is the *systematic ambiguity* of words and expressions as discussed by Ryle (1945b/2009, p. 215). The general idea is that words and expressions have a certain 'elasticity of significance'. In different contexts, words and expressions will express different meanings. i.e., are polysemous. In contradistinction to homonyms, such as the example of *bank* discussed above, these different meanings of a systematically ambiguous word are not unconnected, but can be seen

as ‘inflections of the same root’, which can be understood in terms of family resemblances (Tanney 2009). Take the following three sentences:

[1] His arrival was punctual.

[2] He was punctual when he arrived.

[3] He is a punctual person.

It is clear that the meanings of the different recurrences of the form ‘punctual’ in these sentences have something in common. At the same time, however, the word *punctual* is used in different ways. It is something different for a person’s character to be punctual that it is for his arrival to be punctual on a given time. Ryle (1945b/2009, p. 216) concludes:

> It would be absurd to compare the punctuality of a man on a particular occasion with that of his arrival on that occasion; it would be absurd to compare the punctuality of his character with that of his arrival on a particular occasion. […] These and similar absurdities show that the word ‘punctual’ undergoes inflections of significance when applied to different types of subject.

From this, Ryle (Ibid.) draws the conclusion that, ‘where precision is wanted, it is wrong to speak of “the idea” of punctuality’, although this does not imply that ‘punctuality’ thereby becomes a homonym. For Ryle, the unnoticed systematic ambiguities lead to philosophical problems. In our present discussion, however, they form a inextricable dilemma for the code view. For would we now say that the form *punctual* is a code for three distinct meanings?

One might object that this is at best an epistemic problem: although it will be hard to determine in how many ways a particular form is used, there is a fact of the matter, e.g., patterns of historical usage, that enables one to answer this question. However, Ryle gives us good reasons for not accepting this line of reasoning: if it were not for the systematic ambiguity of natural language, our linguistic abilities would not admit of novelty. This conclusion is also reached

by Davidson (1986), who argues that we often have no trouble working out what somebody means when they use a word in a novel way, or even when they use a novel word. What this shows, according to Davidson, is that understanding what somebody says, in any concrete instance, cannot be explained in terms of what we bring to the table, but instead, has to be understood as an ability to converge on a mutual understanding in a given situation.[12] From this, Davidson (1986, p. 265) concludes:

> there is no such thing as a language, not if a language is anything like what many philosophers and linguists have supposed. There is therefore no such thing to be learned, mastered, or born with. We must give up the idea of a clearly defined shared structure which language-users acquire and then apply to cases.

In response to this, a proponent of the code view could argue that while these arguments might be persuasive, they simply do not conform to our best theories of cognition that *do* assume stable mental meanings. For example, Ryle's systematic ambiguity could just be a property of mental meanings. And while Davidson shows that we can often work out what others mean in the absence of fixed codes, this can be accommodated by the code view as an ability to work out the relevant code on the fly in a given instance.

Here we can see how the code view of language hangs together with a classical cognitivist view of cognition. A classical cognitivist view of thought can be used to argue in favour of the code view: if it is indeed the case that our non-linguistic thought consists of meaningful mental representations, the code view is the most straightforward approach to the cognition of language. At the same time, the code view *requires* a theory of cognition that conceives of non-linguistic thought in meaningful terms. The reason for this is that, on a code view, the meaningfulness of language is explained in terms of the meaningfulness of thought, for, in the words of Locke, ideas are the proper signification of

12 I discuss this point in more detail in §2.5.3, pp. 67ff.

words. The code view thus explains language in terms of a semantic triangle (see figure 1). Linguistic symbols are codes for mental representations, which in turn represent objects in the world.

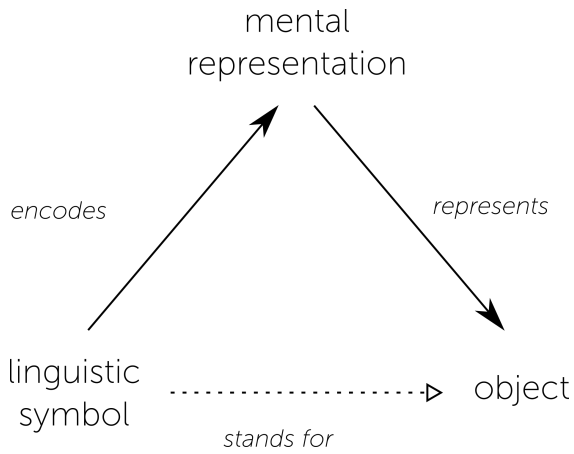

Figure 2: The semantic triangle.

This in turn means that, if we have a different definition of cognition, such as the ecological-enactive concept of cognition, the code view simply will not do. Arguments against the possibility of mental representations, such as the hard problem of content, will therefore also be arguments against the code view of language. For if there were no mental representations, words could not encode them.

The idea that the meaningfulness of language must be explained in terms of the meaningfulness of thought is often taken to be the *only* viable explanation. This is expressed for example by Shapiro (2014, p. 5), who, in a review of Hutto and Myin (2013), claims that without contentful mental states, we would be unable to explain language:

> External symbols acquire their meaning from meaningful thoughts—how could it be otherwise? We use words and other symbols to convey our thoughts to others, and sometimes ourselves. Research in embodied cognition

> indicates that these external symbols might be more thoroughly integrated into the thinking process itself, but none of this is to say that these symbols can be meaningful independently of the thoughts they have been designed to express. So, if the Hard Problem spells doom for contentful thinking, it ought to spell doom as well for our abilities to understand and use language.

The conclusion thus reached by Shapiro seems to pose a dilemma for the ecological-enactive approach: either the quintessential properties of language, such as meaning, content, and linguistic normativity, can be explained in terms of the very same properties that exist at the level of thought, or they cannot be explained at all and thus need to be eliminated from our theories.

1.2.2 The metalinguistic approach

Both horns of the dilemma just identified are problematic for the ecological enactive approach. The dilemma originates in the idea that our theories of human behaviour, including linguistic behaviour, have only two ways of dealing with folk theories: realism and eliminativism. Both of these positions start from treating a folk theory as a proto-scientific theory with ontological commitments (Taylor 1992). This results from the social-linguistic process of *reification*. In the formulation of Van Dijk (2016, p. 994), reification amounts to turning 'a characteristic of an ongoing process into the pre-existing source of that process'. For example, we notice that across different contexts people use the same word, which enables us to make an abstraction, namely the word in question. In a second step, the abstraction is taken to be more basic than the concrete instances that enabled the abstraction. Here it is assumed that each instance of people using the same word needs to be explained in terms of an underlying abstraction. Finally, the abstraction is concretised by retrojecting it onto individuals as a cause of their behaviour (Love 2017). The code view is a prime example of this process. Another good example is the self-proclaimed 'hyper-realist' interpretation of folk psychology by Fodor (1975; 2008), who takes

beliefs and desires to be sentences in a Language of Thought, tokened in the heads of people. Eliminativists agree with the realist framing of the problem. They merely disagree with the realist in that they think these abstractions in fact don't exist, and that the folk theories are radically mistaken.

There is, however, a middle way between realism and eliminativism, which I refer to as the *metalinguistic approach* (Taylor 1990; 2000; 2010; 2013; Harris 1998; Love 1990; 2017). The key insight is to see that metalinguistic practices that attribute properties to language and linguistic behaviour, are not theoretical practices in which we describe a reality independent of those practices; rather, they serve as normative practices that, through regulation and enforcement, create the linguistic reality they describe.

Here is a thought experiment that enables us to grasp this middle way in an intuitive way. Suppose we come across a Martian scientist who is investigating the curious practice that consists in people following a ball across a piece of grassland, only to kick the ball away the moment they finally reach it. Baffled by this form of behaviour, the Martian asks us what the function is of the large numerals on the side of the grassland. They keep track of the score, we tell him. The display currently reads 0 – 2, which means one team has no points, whereas the other has two. Now what should we tell the Martian when he asks us where the points are? Here we would have to tell them to he is making a category mistake: points are not things in the world. So, the Martian asks us, does that mean that the players are mistaken in thinking that they have two points, given the fact that there is no such thing as a point? Here it is obvious that this conclusion does not follow. The points are real, they are just not things. The practice of keeping score does not describe pre-existing points that were there all along; rather, points are created in the practice of keeping score.

According to the metalinguistic approach, our language exists in the same way that the score in a game of football exists. In a way, things like meanings,

rules, promises, and stories only exist because we talk about them. They are not things that exist independent of our everyday ways of talking about them. But this conclusion does not lead to an eliminativist position, for without these metalinguistic practices we are unable to make sense of what we do when we talk to one another. Take the example we discussed earlier: if I ask another person to open the window, and he indeed opens the window, he has understood me. It would be nonsensical for a scientist to say that she discovered that in actual fact this person has not understood me, because he lacks the appropriate mental state of understanding. Similarly, no scientific discovery could show that in actual fact, nobody has ever made a promise, or that, on closer inspection, all nouns turn out to be verbs. The reason is that metalinguistic practices in which we attribute understanding and make promises play a normative rather than a descriptive role. By using these metalinguistic practices, we normatively structure our linguistic behaviour. Taylor (1990, p. 135) explains:

> What we need to examine are our ordinary practices of attributing and justifying attributions of shared knowledge, mutual understanding, and the like. These are normative practices with which we forge the conformity, regularity, and multi-individuality of verbal interaction: i.e., by which we give the individual acts of verbal expression a social instrumentality. But in refocusing our attention on such normative practices, we turn away from speculation about abstract objects and mental states 'underlying' verbal interaction, and concentrate once again on language as a voluntary activity performed by individuals.

Taylor's solution to the problem of the reality of language is metalinguistic, rather than metaphysical: the proper object of study for theoretical reflections on language are concrete and situated normative metalinguistic practices, not abstract objects or underlying mental states. Language has reflexive properties, as expressed in metalinguistic activities, by means of which we attribute properties to language, people using language, and the activities that we can accomplish by using language (Agha 2007). Examples of these kinds of norm-

ative regulation of linguistic behaviour by metalinguistic means include utterances like:

- What I said was *hat,* not *bat.*
- The word kiwi can refer both to a fruit and to a bird.
- A noun is a word that indicates people, places, or things.
- English does not have grammatical gender, as opposed to French.
- Read is pronounced like reed
- She doesn't speak a word of Spanish
- The past tense of *bid* is *bid*, not *bidded*.

This makes language into a special kind of communicative behaviour. For not all forms of communicative behaviour are reflexive in this sense. Harris (1998, p. 27) gives the example of traffic lights:

> Red, amber and green signals play a contrastive communicational role which motorists and other road-users recognize. But the system of lights does not include any 'meta-signals' which mean 'red', 'amber' or 'green'. This is hardly surprising, since directing traffic at intersections is a very restricted communicational function and it is difficult to see what purpose such sophistications would serve.

If we want to explain to someone what the traffic lights mean, we can, and must, resort to language. However, in the case of language, we cannot step outside it to describe on it. A question about questions is itself a question, and an assertion about assertions is itself an assertion. From this, Love (2003, p. 88) concludes, 'language, being language, is on its own. It is interpretatively terminal.'

The metalinguistic approach holds that without metalanguage, 'language as we know it would not be possible' (Harris 1998, p. 28). One particularly insightful example Taylor (2010, p. 19) gives of how this metalinguistic

construction of linguistic practices works, is how a child learns her name (I also discuss this example at p. 65). A child starts out by responding to her name by turning her head when she hears it. Somewhere between 12 and 18 months, the child also learns to utter her own name. This by itself, however, does not mean that the child has learned her name. Dogs can respond to their name, and parrots can be taught to say their own name. However, none of these behaviours show that the child, dog, or parrot, conceives of that particular sound *as their name.* For, knowing what your name is requires you to know *what* a name is, and this understanding cannot be reduced to the mere association between a sound and yourself.

In order to learn what a name is, a child must learn to engage in all kinds of metalinguistic activities. For example, the child must learn to produce her name in response to a question of the form 'What is your name?', she must be able to introduce herself, and so on. If the child is not capable of participating in these reflexive exchanges, then we have no reason to say that she learned her name. If she is capable of participating in these reflexive exchanges, than we must conclude that she knows her name.

A crucial implication of taking metalinguistic practices seriously is that there is no sudden lightbulb moment where a child suddenly grasps a particular concept. Instead, she gradually learns about names by learning to engage in her community's metalinguistic practices. Although 'a child's name does not enter her verbal repertoire *as a name*' (Taylor 2010, p. 25), the child gradually attunes to the metalinguistic practices that constitute our practice of using names, and thus she gradually begins to use her name *as* a name.

A second implication is a radical denial of the ethnocentric fallacy, which is to think that properties of our language use must be universal properties of all language use. Learning what language is amounts to learning what language is *for us*. Taylor (2010, p. 20) explains this in the case of names:

> In Anglophone languaculture this means knowing, for instance, that everyone has a name, that names often have more than one part, that your name identifies you (it declares 'who you are'), that a name cannot be easily changed, that typically when someone calls out your name they are seeking your attention, that putting your name on things is a means of identifying them as your own, and so on. In other languacultures knowing what a name is may involve knowing that saying an adult's name in public is impolite, or that you are not supposed to utter the name of a dead person, or that a name can only be given (or changed) in a special kind of ceremony, or that people have two names each of which is to be used exclusively in particular contexts, or that a person's name tells something about them (such as who their parents are or who their husband is), and so on. In other words, knowing *what a name is* is knowing what a name is *for us*. If you know what a name is, then you have learned what, in our languaculture, names are 'good for': what functions their use serves in our community, and how we value, choose, change, and generally treat them.

On the metalinguistic approach, there is thus no such thing as a name, independent of what we take a name to be.

1.3 Summary of chapters

Besides this introduction, this thesis consists of four chapters that are written as independent research papers, and can thus be read separately. This means that there is occasional repetition, in particular in the introduction of concepts from the ecological-enactive approach. This thesis ends with a general conclusion that also identifies some directions for future research.

In Chapter 2, I propose an account of linguistic knowledge in terms of Rylean know-how. I discuss Nigel Love's seminal distinction between first-order linguistic behaviour and second-order or metalinguistic constructs that we use to describe this first-order linguistic behaviour, such as *word, sentence, meaning,* and *understanding*. Based on Taylor's notion of criterial relations,

linguistic knowledge is defined as knowing-how to provide and recognise criterial support. To be a competent linguistic communicator one has to be sensitive to the criterial relations that originate in metalinguistic practices. I conclude that all first-order linguistic activity has a second-order dimension to it, in the sense that it is directly perceived in terms of the metalinguistic practices that constitute it.

Chapter 2 was previously published as: Van den Herik, J.C. (2017). Linguistic know-how and the orders of language. *Language Sciences*, 61, 17–27. doi.org/10.1016/j.langsci.2016.09.009

In Chapter 3, I develop an ecological-enactive account of utterances of concrete words – words used to indicate observable situations, events, objects, or characteristics. Building on the education of attention model of learning, utterances of concrete words are defined as *attentional actions*: a repeatable form of behaviour performed by a person to indicate (i.e. point out) a particular aspect of the current situation *to* someone *in order to* achieve something. Recent empirical studies show that the attention-directing effects of linguistically constituted categories affect cognitive processes in the absence of overt language use. I argue that this can be understood by relying on the idea that humans continuously self-organise into task specific devices, a form of phenotypic reorganisation. The unfolding of this process is constrained by a history of learning, the actions of oneself, others, as well as aspects of the environment. In this way this chapter proposes a novel explanation for the empirical phenomenon of category effects.

Chapter 3 was previously published as: Van den Herik, J.C. (2018). Attentional Actions – An Ecological-Enactive Account of Utterances of Concrete Words, *Psychology of Language and Communication*, 22(1), 90–123. doi.org/10.2478/plc-2018-0005

In Chapter 4, I propose a way of solving the hard problem of content by showing how children can become competent participants in content-involving practices. I develop an account of the skills a child has to learn in order to develop a sensitivity to the contents of speech acts. In particular, I argue that a child has to be 'calibrated' to measure her world in terms of the normative similarity judgments of her community, a process that can be explained in terms of the education of attention. The calibrated child still has to learn to put her 'measurements' to use in producing and understanding contentful speech acts. The child learns to do so by acquiring reflexive linguistic skills that enable her to participate in the negotiation of correctness conditions.

Chapter 4 is currently *under review.*

In Chapter 5, I defend the thesis that metalinguistic reflexivity is constitutive of linguistic normativity. A potential defeating counterargument against the constitutivity of metalinguistic reflexivity is the regress objection. By discussing the work of Searle, I show that this regress objection originates in the idea that learning language requires learning to follow the constitutive rules of linguistic practices. On this view, learning language is like learning chess: we *first* learn the rules and then put them in practice. I propose an ecological-enactive alternative, according to which learning language can be understood as first learning regular communication behaviour, which can be explained in terms of attentional actions. On this view, it is only because a child *first* behaves in regular ways, that she can *then* retrospectively interpret her own behaviour in normative metalinguistic terms, that is, as being guided by rules. Metalinguistic reflexivity thus enables regulation of communicative behaviour, and thereby constitutes linguistic normativity. I argue that linguistic rules are resources: they are available to participants in order to (re)negotiate properties of language and linguistic behaviour. The account developed in this chapter thus enables us to

understand the constitutive role of metalinguistic reflexivity for linguistic normativity without falling prey to the regress objection.

Chapter 5 is currently *under review.*

2 Linguistic know-how and the orders of language

Humans exhibit linguistic behaviour: we speak, listen, write, and read.[1] Philosophers and linguists typically assume that this behaviour has to be explained with reference to *linguistic knowledge*. This knowledge is traditionally considered to be theoretical knowledge of *a language*, which in turn is understood as a complex system of rules and principles (Barber 2003). On what Matthews (2003, p. 189) calls the *received view*, linguistic knowledge consists of 'an explicit internal representation of these rules and principles.' However, proponents of a radical embodied approach to cognition have recently mounted a thorough attack on the idea that cognitive processes, which include those in linguistic communication, should be explained in terms of internal representations.[2] If there are indeed no internal representations, the received view of linguistic knowledge cannot be correct. However, as of yet, no alternative account of linguistic knowledge has been proposed by these philosophers.

Integrational linguists[3] have similarly argued that linguistic behaviour should not be explained in terms of knowledge of *a language*. Instead, the integrationist 'starts from the premise that communication proceeds by means of signs which are created at and for the moment of communicational exchange' (Wolf and Love 1993, p. 313). This implies that every episode of linguistic

1 Spoken and written language are different phenomena, that constitute different cognitive domains (Kravchenko 2009). A more detailed investigation into the cognitive dynamics of speech and writing would require a separate treatment of these domains. For the purposes of this paper however this is not relevant.

2 See Hutto and Myin (2013) for the most thorough attack on representationalism and Chemero (2009) for a positive account of what a nonrepresentational cognitive science could look like. These accounts can be traced to Enactivism as proposed by Varela et al. (1991) and Ecological Psychology as proposed by J.J. Gibson (1979).

3 Although sometimes distinguished (e.g. Harris 1998, p. 1), I shall use integrationism and integrational linguistics interchangeably.

communication is unique and cannot be explained in terms of (knowledge of) decontextualised rules and principles. However, in explaining linguistic behaviour as context-sensitive language-*making*, the integrationist does not turn a blind eye to our experience[4] of recurrent linguistic units and the stability of their usage. Here Love's (1990) seminal distinction between first-order linguistic activity and second-order constructs comes into play: because we can reflect on first-order linguistic activity and perceive it to be repeatable,[5] we can produce second-order constructs (e.g., word, sentence, noun, verb, hashtag, meaning, understanding, Dutch, English, &c) by means of which we engage in second-order practices (spelling, glossing, defining, writing grammars, explaining the meaning of a word, teaching philosophical writing, &c). A few words about terminology: I consider *second-order practices* to be those linguistic practices that feature *second-order constructs*, which I take to be words that can be used to talk about first-order linguistic activity. Because these practices and constructs are *about* language, I also refer to them as *metalinguistic practices* and *metalinguistic constructs*. I follow Love (1990, pp. 99ff) in using the mass noun language to refer to first-order linguistic activity. This differs from its use as a count-noun – that is, used in conjunction with an article or in its plural form: a language is a second-order construct. These second-order constructs, according to the integrationist, are not merely descriptive, but play a normative role in shaping the production and understanding of language which in turn can explain the experience of linguistic units and stability of first-order practices (Harris 1998). However, in thus redefining the science of linguistics (Davis and Taylor 1990), integrational linguists have paid little attention to what constitutes linguistic knowledge (Taylor 2011).

4 I use experience in the sense of perceptual experience, which can be explained in non-representational terms (e.g. Degenaar and Myin 2014). By linguistic experience I mean perceptual experience of language.

5 To perceive first-order linguistic activity to be repeatable is to 'embrace the possibility of "saying the same thing"' (Love 1990, p. 99).

In this paper I propose an account of linguistic knowledge. As opposed to the received view, which conceives of linguistic knowledge as theoretical knowledge of a language, I propose an account of linguistic knowledge in terms of Rylean knowhow. Based on Taylor's (1990) notion of *criterial relations*, I argue that the normative character of second-order practices is crucial for understanding linguistic know-how. In doing so I also argue against the possibility of making a clean distinction between 'pure' first-order linguistic activity and linguistic activity informed by second-order practices, thereby providing an argument for Pablé and Hutton's (2015, p. 29) claim that 'first and second-order practices are inextricably intertwined'.

This paper is laid out as follows. First, I briefly introduce the integrationist approach to language (Section 2.1) and Love's distinction (Section 2.2). Following this, I distinguish two possible views of Love's distinction: the *optional extension view* and *the constitutive view*, and argue for the latter (Section 2.3). I then propose an account of linguistic knowledge in terms of Rylean knowing-how, paying special attention to how the skilful exercise of this know-how is subject to correction based on criteria. I rely on Taylor's notion of *criterial relations* to show the necessity of second-order practices for linguistic knowledge (Section 2.4). In the concluding section I argue that, if my account of linguistic knowledge is correct, first- and second-order practices cannot be cleanly distinguished.

2.1 The integrationist account of language

According to Harris, the founding father of integrationism, all Western theorizing about language is under the spell of the language myth[6]: 'a sedimented

6 The integrationist unmasking of the language myth resulted in fundamentally *Redefining Linguistics* (Davis and Taylor 1990) and *Rethinking Linguistics* (Davis and Taylor 2003). Integrationists assert that the object of study of orthodox linguistics is created by a particular, culturally determined point of view that necessarily arises out of immersion in Western

form of thinking that has gone unchallenged for so long that it has hardened into a kind of intellectual concrete' (Harris 2001, p. 1). To fall prey to the language myth is to assume: (i) the existence of languages, understood as synchronic systems of fixed codes linking words to private meanings (the determinacy thesis); and (ii) that the sharing of these fixed codes is necessary for linguistic communication, which is then thought to consist in the transference of private meanings from one person to the other by encoding and decoding private meanings into words (the telementation thesis) (Harris 1998, p. 32). In this paper, I do not rehearse the arguments that integrationists have levelled against the language myth, but instead start from the assumption that the integrationists are correct in rejecting the language myth.

What I want to emphasise is the integrationist conclusion that the positing of a language turns any explanation of linguistic behaviour on its head. For 'what requires explanation is misrepresented from the outset by a priori theoretical fiat' (Love 1990, p. 75). The language that an orthodox linguist assumes and then employs as an explicans, is treated as an explicandum by the integrationist (Harris 2003, p. 50; 1998 p. 55).[7] This integrationist insight is articulated by Harris (1998 p. 5) when he states that 'the right theoretical priority is exactly the reverse: languages presuppose communication.'

Whilst opposing the idea that a language can explain linguistic behaviour, the integrationist thus does not turn a blind eye to our experience of linguistic units (words, sentences, common expressions, &c) and the stability of their usage. In doing so, the integrationist walks a fine line between the Scylla of eliminativism and the Charybdis of naïve realism (Cowley 2011a) or linguistic

metalinguistic practices (Love 1995, p. 337; 2007). Orthodox linguistics, then, is dismissed as an extension of lay metalinguistic practices (Davis 2003, p. 3) resulting from the language theorist taking everyday metalinguistic questions as intellectual challenges (Taylor 1992; Harris 1996, p. 149). For the purposes of this paper however, I will not delve into the questions regarding the scientific status or proper object of linguistics.

7 According to Love (1990, p. 107), this means that languages must be naturalised, a project that sits well with the radical embodied project of naturalising all forms of cognition (e.g. Hutto and Myin, 2013).

immanence (Taylor 2010). The immanent realist position results from taking the lay perception of language at face value (Harris 1998, p. 53). Linguistic communicators obviously perceive language as consisting of recurring units and the integrationist is not in the business of convincing anyone that this perception is illusionary or that the accompanying metalinguistic practices are mistaken. That is, the integrationist is not an eliminativist with regards to lay people's metalinguistic practices, but aims to prevent these practices from giving rise to metalinguistic illusions in theorising about language (Taylor 1992; Harris 1996, p. 148). Accordingly, instead of explaining linguistic behaviour in terms of objectively existing underlying invariants, the integrationist turns to the experience of language to explain the emergence of apparent invariants.

This tension with regards to the ontological status of languages in particular and metalinguistic constructs in general is diagnosed by Love (1990) as follows: we – humans exhibiting linguistic behaviour – experience linguistic activity as if it consists of repeatable linguistic units, that is, as if it consists of tokens, or instantiations, of more abstract types.[8] At the same time, Love states that these linguistic units are radically indeterminate because *saying the same thing* – that is, repeating a linguistic unit – depends on 'the kind of sameness required' (Ibid., p. 100). For instance, saying *Schnee ist Weiss*, *snow is white*, and *sneeuw is wit* amounts to saying the same thing for some purposes, whereas, in other contexts, we can legitimately claim that different pronunciations of the 'same' word do not constitute saying the same thing.[9] What counts as a repetition thus cannot be determined in a decontextualised way, but 'is something for speakers themselves to decide in particular contexts' (Ibid., p.

8 Interestingly, we do not have this reificatory tendency with other aspects of human behaviour. Harris (1998, p. 83) likens 'saying the same thing' to 'scoring the same goal in football'. It is obvious that the latter is impossible. No matter how closely the second goal approximates the first, it will be a new goal. Similarly, Harris argues, no matter how closely an utterance resembles a previous one, something new will be said.

9 Cf. Love (2003; 2004) for many more examples that problematise the immanent realist position.

98). So, on the one hand we experience language as consisting of linguistic units, whilst we know at the same time that these units are radically indeterminate from an objective point of view (Ibid., p. 106).

However, this way of describing the experience of linguistic units – as instantiations of abstract types – runs the danger of overintellectualising our experience of first-order linguistic activity. As Harris (2003, p. 57) points out, *re*cognising the 'same' linguistic unit is like recognising the same face or the same shape.[10] I shall therefore refer to these experienced linguistic units as (linguistic) repeatables – also to stress the action required to accomplish repetition.[11] Moreover, the radical indeterminacy of these repeatables from an objective point of view does not entail that linguistic communicators can always *decide* whether or not something is repeated, in particular because they cannot decide how to experience linguistic activity. I return to this point in Section 2.4.3 when I discuss novelty in linguistic activity.

We should therefore not understand repeatables in terms of (experienced) identity relations, but as the '*perception* of relatedness, similarity, and recyclability' (Toolan 1996, p. 267, emphasis added). In contrast to identity relations or type–token relations, the notion of repeatability presupposes a temporal order. For when an earlier utterance is identical to a later utterance, the later utterance is also identical to the earlier utterance – or, alternatively, both are tokens of the same type. If a later utterance is a repetition of an earlier one, however, the

10 One might object that these recognitional capacities require mental representations, and that therefore this approach to linguistic knowledge in this paper turns out to be representational after all. Hutto (2006), however, gives a non-representational account of recognitional abilities. Degenaar and Myin (2014) argue that even in the case of absent or abstract features representations need not be invoked.

11 Note that there are two things what we should be careful not to conflate: on the one hand, we have an individual's experience of linguistic activity as it unfolds. To *describe* this experience, we need not invoke abstractions, for what we experience is not an abstraction but a familiar concrete event, although these concrete events can be described by lay people in terms of linguistic units such as words and languages. On the other hand, we have the language theorist's *explanation* of this fact of experience, which can – and in the case of orthodox linguistics, does – invoke *abstracta*.

earlier utterance is not a repetition of the later one. By definition, you can only repeat what has already happened. This entails that the repeatability under consideration here is itself time-bound (Harris 1998) and therefore cannot be determined in a decontextualised way.

As Toolan (1996, p. 238) reminds us, what we are concerned with is 'the ontological level at which repetition exists'. For whereas we are undoubtedly capable of repeating 'bits of language' we hear someone else speak or we see in writing, the brute fact of this repeatability does not conjure into existence decontextualised linguistic objects that provide objective standards of correctness for determining when repeatables are *really* repeated. Whether a repetition is taken to be a repetition by participants is therefore always a contextual matter. For the integrationists, to insist on one true analysis, 'which "really" reflects the structure of "the language used" is a metalinguistic illusion' (Harris 1996, p. 160).

2.2 Love's distinction

In understanding repetition in terms of underlying abstractions, the utterance – which is a spatio-temporally extended and context-determined event brought forth by a linguistic communicator – gets reified by treating it as an instantiation of underlying linguistic objects (words, sentences, &c). To understand the origins of this metalinguistic illusion, Love (1990, p. 101) introduces a seminal distinction:

> A language[12] is a second-order construct arising from an idea about first-order utterances: namely, that they are repeatable. Such a construct may be institutionalized and treated as the language of a community. But it remains a

12 Love (1990, p. 100) conceives of a language 'as an individual's system of repeatable abstractions underlying language-use'. Of course, the abstracta themselves are not repeated but repeatedly *tokened*.

> construct based on an idea: at no point does it become a first-order reality for individuals. (Although in a society which teaches its institutionalized construct to its members it may be expected to have a large effect on their first-order behaviour, and may perhaps give rise to linguistic theories which project the construct on to them as the basis for their first-order behaviour.)

First-order linguistic activity is understood as the 'making and interpreting of linguistic signs, which in turn is a real-time, contextually determined process of investing behaviour or the products of behaviour (vocal, gestural or other) with semiotic significance' (Love 2004, p. 530). This is an expression of the integrationist idea that as time-bound agents, we cannot step outside the 'time-track of communication', and, therefore, linguistic signs 'cannot exist except in some temporally circumscribed context' (Harris 1998, p. 81). Instead of selecting ready-mades from a (mental) depository of signs, we *make and remake* linguistic signs. Competent linguistic communicators directly perceive this first-order linguistic activity as consisting of repeatables, i.e., as affording repetition – akin to how we directly recognise a familiar face. I shall refer to this experienced stability (as opposed to material stability (Harvey 2015)) as *metalinguistic experience*.[13] What metalinguistic experience amounts to is accentuated when one is confronted by unfamiliar first-order linguistic activity. One simply cannot determine what would count as repeating an utterance in a language one is not conversant with.

In this way, integrationists emphasise the process-like nature, creativity and context-boundedness of first-order linguistic activity. This creativity, however, is not limitless. For our linguistic abilities include that of giving metalinguistic accounts of our communicational behaviour (Pablé and Hutton 2015). We can describe the flow of linguistic activity in terms of second-order constructs (words, sentences, understanding, meaning, &c) because of our meta-

13 This is akin to what Cowley (2011b) calls 'taking a language stance', which amounts to 'treating speech as if it consisted of verbal patterns'.

linguistic experience of first-order linguistic activity as consisting of repeatables. These second-order constructs in turn can shape or inform first-order linguistic activity in as much as they are used prescriptively. They are therefore not merely descriptive, but carry normative force (I return to this in the section on linguistic knowledge). In our Western society, for instance, with its historically unprecedented grade of literacy, these second-order constructs are codified in dictionaries and grammar books and used as pedagogical tools (Love 1995), leading to explicit norms of linguistic correctness.[14]

Love (2007) takes second-order constructs to be *constructs* because they do not refer to objectively given first-order objects. The second-order constructs are therefore simply those a community agrees on. There can be no decontextual authority (Pablé and Hutton 2015), no higher court of appeal (Harris 1998, p. 145), and, specifically, no scientific discovery, which objectively grounds second-order constructs. Note that this does not imply that anything goes. As described above, second-order constructs surely have normative force, but only when enforced. Think for instance of the recent rise of using 'they' as a singular pronoun. Purists might dislike this novel use and condemn it as incorrect. However, as more and more people start using 'they' in this way, it might effectively become a singular pronoun. The absence of a decontextual authority means that there can be no definite answer to the question whether 'they' actually *is* a singular pronoun.

2.3 Second-order practices: optional or constitutive?

From the discussion so far, multiple views of second-order practices are possible. In this section I distinguish between (i) second-order practices as an

14 Harris (1996, Ch. 2 & Ch. 3) traces Western second-order constructs to the pedagogical needs of the ancient Greek grammarians' main task, viz. to offer an education in writing and literature.

optional extension of first-order linguistic activity, and (ii) second-order practices as *constituting* first-order linguistic activity. Call these the *optional extension view* and the *constitutive view*. I shall argue for the constitutive view. Note that it is not my intention to attribute these views to any particular author (I rely on quotes from Love (1990) for both views) but, rather, to distinguish between them for conceptual clarification, thereby setting the stage for the section on linguistic knowledge. After discussing linguistic knowledge, I shall address the question whether we can – from the constitutive view – still make a sharp distinction between second-order practices and first-order linguistic activity.

2.3.1 The optional extension view

Love (1990, p. 99) observes that the priority of language over languages – that is, the priority of first-order linguistic activity over the idea of a language as consisting of repeatables – depends on 'a simple point of logic'. He argues that, both as regards the nascency of language as well as regards a child being initiated into language, there must be a primordial act of understanding. This primordial act of understanding first-order linguistic activity must precede metalinguistic experience of that activity. That is, there must be a primordial utterance which was understood by someone in spite of not 'being able to relate it to antecedently given abstract units' (Ibid., p. 107).

Love then claims that the capacity for reflecting on experience (not on linguistic activity) by abstracting from it is presupposed by any language use. For instance, to be able to talk about cats, 'distinct individual organisms have to be seen as tokens of a type "cat"' (Ibid., p. 97).[15] As the capacity for reflecting on experience was thus necessarily present at the nascency of language, and

15 Here we once more run the danger of overintellectualisation. Nothing seems to be lost if we would say that distinct individuals have to be recognised as cats, rather than that they have to be seen as tokens of a type cat.

linguistic activity itself is experienced, Love concludes that 'the birth of language as an object of contemplation follows hard on the heels of the birth of language itself' (Ibid.).

On the optional extension view, first-order linguistic activity therefore has both temporal and logical priority over metalinguistic experience, and therefore, over metalinguistic practices that depend on metalinguistic experience. This in turn entails that there was a period of pure first-order linguistic activity. Thus, so Love claims, '*perceiving* utterances as manifestations of underlying "sames" is not a necessary condition of any use of language whatever' (Ibid., p. 100, emphasis added).

The optional extension view is based on a notion of understanding on which it makes sense to ascribe understanding of first-order linguistic activity to a person who cannot ascertain this themself – for declaring that one understood something relies on the use of the metalinguistic construct *understanding*. On this view, it is the act of understanding that makes metalinguistic experience possible: 'once one has understood an utterance for the first time, one will entertain the possibility of repeating it' (Ibid., p. 100). In the rest of this section I argue against this notion of understanding, thereby reaching the conclusion that metalinguistic experience and practices are not an optional extension of first-order activity, but are instead constitutive of it.

2.3.2 The constitutive view

The logical and temporal priority of first-order linguistic activity over metalinguistic experience and metalinguistic practices follows from the fact that a primordial act of understanding is presupposed by metalinguistic practices. But what would this act of primordial understanding amount to? Love (1990, p. 107) gives the following suggestion:

> It means, perhaps, that A's vocal noise elicited from B behaviour that suggested to both A and B that associations (images, memories, etc.) somehow evoked in B by A's noise were similar to those associations of the noise for A.

But how similar would the associations of A and B have to be to constitute understanding? The only way this suggestion could provide a criterion for understanding is for there to be an objective measure of the similarity of A's and B's associations as well as a cut-off point above which we could speak of understanding. But then understanding would be a telementational notion: understanding would be achieved when B's associations are sufficiently similar to the ones A originally 'had in mind'.

The first step towards a non-telementational interpretation of understanding is therefore to notice that in our normal usage of the term, we do not use any mental criteria for understanding. Taylor (1990, p. 130) gives the example of asking someone to open the window. In normal circumstances, if the person we ask indeed opens the window,we conclude that they understood our request; no further knowledge of their mental state is required for this judgement (cf. Love 2004, p. 532).[16] We manifestly do not treat understanding as 'an unobservable, private, mental event' (Taylor 1986, p. 179).

An alternative integrationist interpretation of understanding is available. Toolan (1996, pp. 262–263) proposes that understanding – rather than being dichotomous in the sense of 'an abrupt or sudden illumination' – is 'a gradual, incremental grasping of things and their uses, or "things in use"'. Toolan's alternative has important implications for Love's distinction. If understanding consists in the grasping of things in use, then understanding requires these things to be experienced by the participants. This in turn entails that, rather than following hard on the heels of the birth of language, metalinguistic experience

16 Of course, we seldom make this explicit judgement. Moreover, this judgement is defeasible. Take the contrived example of a person who always opens the window, regardless of what is said.

of vocal noises[17] – that is, coming to treat them as 'repeatables' – is a requirement for first-order linguistic activity. This re-evaluation of metalinguistic experience as a necessary condition for the genesis of language does not require assuming additional capacities, for Love already assumed the necessity of reflecting on experience for anything that we would call language. The constitutive view thus controverts the optional extension view merely in negating the temporal priority of first-order linguistic activity over metalinguistic experience. On the constitutive view, metalinguistic experience is constitutive of any behaviour that we would call linguistic, where I understand constitutive to mean that metalinguistic experience plays a crucial role in producing any behaviour that we would call linguistic. Only when a person starts to recognise vocal noises, that is, experiences them as repeatables, and starts using them in stable ways, can we call their behaviour *linguistic*.

But even if metalinguistic experience is constitutive of first-order linguistic activity, we might still suppose that explicit metalinguistic practices need not be. This would be feasible if there can be a period, either in sociocultural or ontogenetic development, in which first-order linguistic activity is accompanied by metalinguistic experience without there being explicit metalinguistic practices.

It is exactly this possibility that is investigated and rejected by means of a thought-experiment by Taylor (2000). He asks us to envision humans that engage in first-order linguistic activity without there being any metalinguistic practices. Such a linguistic community would not employ metalinguistic constructs such as *word, sentence, noun, verb, mean, talk, speak, promise, answer, suggest, describe, refer,* and therefore would not be able to talk about language at all, as we often do, when we for instance say things such as *what did you say?*, *do you understand?*, *please stop talking!*, and *what's the answer*

17 I follow Love in talking about vocal noises. The same argument applies to forms of linguistic communication that do not rely on vocalisations, such as gestural languages.

to this question? As Taylor (2000, p. 489) argues, much depends in this thought-experiment on 'what we are willing to call "language"'. Even if the humans were capable of using some sort of regularised signalling behaviour to influence each other's behaviour, it would surely be much different from what we would normally call language.[18] For this alleged behaviour would lack even the most basic properties that we attribute to language. For instance, Taylor describes, there would be no possibility of standardisation and error. Without the metalinguistic resources to call some behaviours 'correct' and others 'incorrect', there could not be criteria for correct linguistic behaviour.[19]

A similar line of reasoning holds for the notion of understanding. In discussing the two views of second-order practices so far, we tacitly assumed understanding to be a purely descriptive term. That is, we took the response of opening the window, after a request to do so was made, to demonstrate understanding in a straight-forward way. However, in the absence of a metalinguistic notion of understanding, 'it is not clear how a hearer's response to something said could have counted as a case of understanding or of not-understanding' (Taylor 2000, p. 490). That is not to deny that we – who know how to wield the concept of understanding – could attribute understanding to the imagined metalinguistically impaired people. But in doing so, Taylor claims, we would fall prey to the *ethnocentric fallacy* in assuming 'that the reflexive linguistic distinctions which our culture applies in evaluating and characterizing communicational behaviour must also be applied – and if not explicitly, then implicitly – by the members of every culture' (Ibid.).

To conclude this section, metalinguistic practices constitute our practice of understanding (and, *mutatis mutandis,* meaning, talking, promising,

18 Taylor (2000, p. 493) discusses the example of vervet monkeys' alarm calls. Although these calls surely have a communicative function, it is unclear what purpose it would serve to attribute metalinguistic properties such as truth or reference to these calls, when the monkeys themselves are obviously unable to conceive of the calls in these terms.

19 Nor, incidentally, could a distinction be made between linguistic and non-linguistic behaviour.

describing, &c). Without these metalinguistic practices we would not be able to disagree as to whether someone understood someone else, which entails that the distinction would have no force in our first-order activities. It is therefore not only metalinguistic experience of linguistic activity as being repeatable, but the metalinguistic practices themselves that constitute first-order linguistic activity. Although people that lacked metalinguistic practices might develop some kind of signalling system, it would not be recognisable to us as linguistic. Reflexivity – understood as linguistic activity turning back on itself through metalinguistic practices – is therefore a necessary property of anything we would call language.[20] What at first sight might be conceived of as pure first-order linguistic communication – talking *about* something, promising, telling the truth, claiming, describing, referring to something, &c – is thus shown to be constituted by second-order practices. In the next section, I show how this conclusion feeds into an account of linguistic knowledge.

2.4 Linguistic knowledge

Taylor (2011) claims that there is a surprising absence of discussion in the integrationist literature on language acquisition and linguistic knowledge. This question is especially pertinent to integrationists and proponents of a radical embodied approach to cognition, as we have seen in the introduction, because they cannot adopt the received view of linguistic knowledge.

Taylor and Shanker (2003) distinguish between *what* the child acquires, and *how* the child acquires this. In this penultimate section, I focus on the *what*-question – understood as the question 'what constitutes linguistic knowledge?' I particular, I argue that linguistic knowledge should be conceived of as practical

20 This necessity of reflective practices first argued for by Taylor (1990; 1992) became an important part of integrationism, as evidenced for instance by its inclusion in Harris' *Introduction to integrationial linguistics* (1998, pp. 25ff).

knowledge – or knowing-how – rather than theoretical knowledge – or knowing-that. I first introduce the distinction between knowing-how and knowing-that, after which I show how linguistic know-how is dependent on metalinguistic practices as congruent with the constitutive view. In particular I argue that Taylor's (1990) notion of *criterial relations* is crucial to understanding linguistic knowledge.

2.4.1 Knowing how and knowing that

Ryle (1946) argued for a distinction between knowing-that – theoretical knowledge, conceived of as considering propositions – and knowing-how – practical knowledge or a direct exercise of intelligence in practical performance. In exercising know-how, Ryle maintains, intelligent performance is not the outcome of occult episodes of considering propositions, but is a manifest property of activities. That is, the activities themselves are conducted intelligently or stupidly (Ryle 1949, p. 27). Ryle's master argument against intellectualists that give priority to theoretical knowledge is that because the entertaining of propositions is itself an activity, conceiving of know-how in terms of prior theoretical knowledge leads to a regress. For if intelligent performance requires a temporal prior reflection on how to act, conducted by means of considering propositions, this reflection on how to act itself can only be intelligent if preceded by another instance of reflection: a reflection on how best to reflect on how to act (Ibid., p. 31).[21]

Hanna (2006) builds on Ryle's account of knowing-how. She starts from Dummett's (1978) observation that the response *I don't know; I've never tried*, is *prima facie* acceptable when asked if one can swim, whereas it would not be when asked if one can speak Spanish. Dummett intended this example to

21 Stanley and Williamson (2001) argued that this Rylean argument is mistaken, and that know-how always requires knowing that. See Simpson (2010, pp. 632ff) and Noë (2005) for rebuttals.

demonstrate that *qua* rational activity, speaking Spanish cannot be a skill. However, so continues Hanna, this is to misconstrue what is meant by practical knowledge. For the term 'swim' in the example has two possible interpretations: first, merely being able to hold one's head above water, and, second, the skill of swimming as required to pass swimming classes, such as performing the breaststroke. Whereas one can genuinely be ignorant whether one would be able to stay afloat, performing the breaststroke is an acquired skill that admits of criteria; when we interpret swimming in the latter way, the nescient response is unjustified.

Accordingly, if we conceive of the distinction between practical and theoretical knowledge as a distinction between the kind of 'thrashing about'[22] one undertakes in trying to stay afloat on the one hand, and obvious rule-following activities such as playing chess on the other hand, then surely linguistic communication is more like the latter. However, setting up the debate in these terms excludes a third category, that of skills.[23] For in contrast to merely staying afloat, performing the breaststroke is subject to evaluation and correction, and it is the possibility of correction that allows us to say that an action is performed skilfully. I therefore define skills as those activities we perform that are subject to correction by others.[24]

22 I borrow this term from Fodor (2008, p. 13), who explicitly opposes the Rylean conclusion of the priority of practical knowledge over theoretical knowledge. Note that Fodor's (1975, p. 58) rejection of the primacy of theoretical knowledge leads to the absurd conclusion that we cannot acquire a first language. Interestingly, this conclusion is based on an account of language learning that cannot but view this process as a metalinguistic endeavour of hypothesis formation and confirmation, where the metalanguage is mentalese.

23 According to the Pablé and Hutton (2015, p. 68) 'Knowledge is a matter of an individual's integrational –or communicational – proficiency, which "comprises the whole range of knowledge, abilities and experience that [he/she] can bring to bear on the communication situation in question." […] (Harris 2009, p. 70)'. A comparison between what I define here as skills and communicational proficiency lies outside the scope of this paper.

24 In (radical) embodied approaches to cognition, the notion of skill is often used in a much broader sense. Crucial here is whether there are sources of normativity that are non-social in nature. Chemero (2009, p. 145), for instance, sees affordances as relations between abilities and features of the environment, while maintaining 'that there is something inherently normative about abilities'. Enactivists such as Di Paolo (2005) explain non-social normativity in terms of the self-production of organisms. See Heras-Escribano et al. (2013) and Heras-

This correction in turn is possible because there is a conventional way in which the activity is performed. In line with the reflection on experience Love identified as necessary for the genesis of language, a particular swimming behaviour is conceived of as an instance of performing the breaststroke, and the perceived (dis)similarities with other instances of performing the breaststroke provide criteria for evaluating and correcting behaviour.

When we say that there is a *conventional* way of performing the breaststroke, we need not assume that there is a convention that guides performance. Instead, this is merely to point towards the fact that when we perform the breaststroke, we are subject to correction by others only if they conceive of it as an attempt to perform the breaststroke.[25] When we think of a convention as an entity, we can come under the impression that it is the cause, or among the causes of, people behaving similarly. Instead, I propose, we should conceive of conventions as a *coming together* (*com-venire*), that is, as the emergence of relatively stable patterns of behaviour. Following Millikan (1998), I shall refer to these stable patterns as *natural conventions*. Millikan (1998, p. 162) understands natural conventions as 'patterns that are "reproduced"', where 'the fact that these patterns proliferate is due partly to weight of precedent, rather than due, for example, to their intrinsically superior capacity to perform certain functions'. I therefore take *conventional regularities* to be those regularities that proliferate in part due to precedent.

According to Millikan, the explicit rules we can formulate to describe these natural conventions are just that: descriptions. The natural conventions by themselves therefore carry no prescriptive force. It is only when the behaviour of others is corrected based on an awareness of these conventional regularities,

Escribano and De Pinedo (2015) for a Wittgensteinian critique of these notions of normativity. For the purposes of this paper, I do not address the question whether non-social forms of normativity exist, and therefore limit my discussion to social forms of normativity.

25 Note that in consonance with earlier discussion, this experience of something as being repeatable need not imply that it is named by participants. In general, an inability to formulate explicit rules does not negate skilfulness (cf. Noë 2012, pp. 118ff).

that they acquire prescriptive force, though always by proxy of the corrector (cf. Wittgenstein 1983, VI §42).[26] So we have to say that the conventional regularities need not only be regularities, but also conceived of as being regularities and therefore repeatable in order for them to be counted as skilful behaviour. Note that none of this implies that the kind of rules under consideration here have to be effable for the rule-followers. In terms of the previous section, metalinguistic awareness of 'things' in use does not imply accompanying metalinguistic practices.

Before we turn our attention to the role criterial relations play in linguistic knowledge, I want to say a few words about explicitly formulated rules. Linguistic prescription based on explicit rules occurs abundantly in highly literate societies – based *inter alia* on codification in dictionaries and grammar books. In this process of codification, natural conventions are reified into what Millikan calls *stipulated conventions*, which refers to explicitly stated rules.[27] This does not threaten an account of linguistic knowledge in terms of know-how. First of all, this process of reification depends on practical linguistic knowledge. Moreover, the reified rules are not occult (mental) entities, but are formulated in natural language and therefore publicly available. As such, an explicitly stated rule is not an explanatory hypothesis but a guide for conduct and a standard of correctness (Hacker 2014). Finally, similar to natural conventions, explicitly stated rules carry prescriptive force only in so far there are used to correct the behaviour of others, or, derivatively, of oneself.

26 Note that this observation does not negate the fact that *once* someone mastered a particular practice, they can correct themself based on observed natural conventions. A discussion on this self-corrective process is, however, outside the scope of this paper.

27 Prescription on the basis of explicit rules of course need not be based on previous conventional regularities – think for instance of a person coming up with the rules of a new game.

2.4.2 Criterial relations

Besides conceiving of an utterance as repeatable, and the perception of (dis)similarities between repetitions, we still have to explain how language users can correct one another based on the appropriateness of a particular linguistic act in context. In the rest of this section, I argue that Taylor's idea of *criterial relations* provides insight in the constitution of natural conventions regarding the use of linguistic repeatables. Central to this notion is that we conceive of metalinguistic practices as normative practices in which the experienced stability of language is forged (Taylor 1990).

According to Taylor (1990, p. 136), a criterial relation holds between assertions. He introduces the notion as follows: 'Part of knowing the meaning of the word *understand* is ordinarily taken to include knowing that the assertion of "he gave a correct explanation of E" provides criterial support for the justifiable assertion of "he understands E".' Understanding what a word means – understood as grasping a repeatable in use – is thus constituted by knowing how to provide criterial support for its application and recognising when criterial support is provided.[28] These criterial relations can thus be conceived of as norms or rules (natural as opposed to stipulated conventions) specifying correct use of the word. However, whether or not an assertion provides criterial support for another is not determined in a decontextualised fashion. Criterial relations can only provide standards of correctness when recognised by linguistic communicators. This means that what criterial relations apply is always a contextually determined matter (Ibid., p. 144). Crucially, a community must place a weight on the maintenance of the criterial relation, and it will usually do so if this maintenance is conducive to furthering its purposes.

28 Once again, in line with Section 3.1, I understand constitutive to mean that knowing-how to provide and recognise criterial support plays a crucial role in any behaviour we would attribute understanding to.

A particular enlightening example of these criterial relations is given by Taylor (2000; 2010; cf. Shanker and Taylor 2001) in discussing the practice of naming.[29] He starts from the observation that many parents hold that one of the first words their child learns is their own name. But, when the child is able to make vocal noises that are similar to the ones others make when they use the child's name, and the child responds to hearing its name by directing its attention to the speaker that used it, these behaviours do not show that the child conceives of that repeatable as its name. Many animals can be conditioned to respond to and produce particular vocalisations, and this fact does not show an understanding of the practice of naming. That is, a 'child's name does not enter her verbal repertoire – either productively or receptively – as a name' (Taylor 2012, p. 9). For besides being able to produce a vocalisation that admits of relevant similarities to her name, the child has to be 'reflexively enculturated' by knowing how to provide criterial support for utterances like *x is my name* (Taylor 2000, p. 496). Recognising that another person saying *I'm Roy* when you first meet them provides criterial support for saying *His name is Roy* is to go beyond the simple response to and production of a vocal noise. Whilst this example may seem trivial, it manifestly relies on the metalinguistic practice wherein the metalinguistic construct *name* is used. Note also, that in line with the discussion about understanding, in a culture in which the metalinguistic construct *name* is absent people would not have names – at least not what we understand by names.

To understand what it is to have a name thus includes participating in reflexive practices like telling others your name when you meet them for the first time, responding appropriately to questions like 'who is the one called Roy?', understanding that you can name your stuffed animals, &c. This shows

29 These practices are of course culture-specific. Think for instance of cultures where your name changes once you reach adulthood, cultures in which a value is placed on avoiding the names of recently deceased, &c.

that merely engaging in the right kind of first-order linguistic activity – where 'right kind' is determinable from the perspective of other participants – is not enough: the child has to become conversant with the appropriate second-order practices in order to understand what it means to have a name. In other words: the *prima facie* first-order practice of naming is constituted by second-order practices of reflecting on naming, in the sense that second-order practices play a crucial role in first-order practices.

It is crucial to note that, in attuning to the criterial relations that constitute the practice of naming, there is no nonarbitrary point where the child can be said to have completely mastered the practice. Even adult speakers are – in line with Putnam's (1975) idea of the *division of linguistic labour* – unaware of all metalinguistic practices that constitute the first-order practice of naming. Even in the everyday phenomenon of personal names we can imagine people who are unaware that one can change one's name, adopt a *nom de plume*, call loved ones by a pet name, &c. Moreover, the criterial relations that constitute a particular practice are ever-changing, which makes any envisaged complete understanding a practical impossibility. A useful metaphor is provided by Ryle (1945/2009), who compares being at home in language as knowing your way around a familiar village. Both are acquired through one's daily walks and are 'knowledge by wont and not knowledge by rules' (Ibid., p. 211). And in both cases, complete knowledge is impossible. This entails that there is no stable state that a person can achieve such that it can be counted as having fully mastered a particular linguistic practice.

Whilst linguistic knowledge, understood as knowing-how to give and recognise criterial support, is necessary for the participation in first-order linguistic communication, linguistic knowledge cannot be reduced to rigid application of criterial relations. Crucial here is the observation that what criterial relations apply is context dependent and subject to continuous change.

Moreover, although criterial relations are often made explicit in metalinguistic practices, this need not be the case. Finally, criterial relations do not hold only between assertions, as non-verbal behaviour can provide criterial support (e.g., a person's nodding whilst you are talking to her provides criterial support for her understanding you).

2.4.3 Linguistic knowledge and novelty

An interesting phenomenon that shows the limitations of explanations of linguistic behaviour in terms of criterial relations is that of novelty in linguistic activities. This deviance in the application of criterial relations can of course be affected through explicit metalinguistic discourse. Take the case of Humpty-Dumpty, who claims that 'When *I* use a word, it means just what I choose it to mean.' This leads to the situation in which Alice does not understand his utterance *There's glory for you*, which does not surprise Humpty Dumpty: 'Of course you don't – till I tell you'. Only by explicitly telling her that he means *There's a knock-down argument for you* by *There's glory for you* – that is, by relying on explicit metalinguistic discourse – can Alice understand him. However, more interesting for our current purposes is people getting away with what Humpty Dumpty attempts to do *without* thereby relying on explicit metalinguistic discourse.

Davidson (1986)[30] shows this possibility when discussing understanding of utterances in context. The phenomenon that interests Davidson is the fact that successful communication is rarely frustrated by novelty in linguistic activity, such as demonstrated by malapropisms, incomplete or garbled utterances, unfa-

30 Goldstein (2004) also notices the similarities between integrationist thought and the Davidson (1986) that wrote 'A nice derangement of epitaphs'. However, Goldstein does not part with traditional ways of conceiving of language (Harris 2004). Of particular interest to integrationists should be Davidson's (1986, p. 265) confutation of the language myth in his conclusion that 'there is no such thing as a language, not if a language is anything like what many philosophers and linguists have supposed', which he defines as 'a clearly defined shared structure which language-users acquire and then apply to cases'.

miliar words, neologisms, slips of the tongue,&c. In these cases novelty is produced – knowingly or not – either through converting familiar words or constructions 'to a new use by an ingenious or ignorant speaker' (Ibid., p. 259), or, through sheer invention. From this, Davidson concludes that our ability to understand one another cannot be explained in terms of what we bring to the table, since this cannot account for the specificity of the occasion (Simpso 2010, p. 639); instead, we converge on a mutual understanding[31] in a particular context. In other words: 'a speaker may provide us with information relevant to interpreting an utterance in the course of making the utterance' (Davidson 1986, p. 260).

This conclusion reached by Davidson thus shows how first-order utterances *themselves* get to do some of the work that we might originally have thought to be reserved for explicit metalinguistic discourse. By being understood and accepted[32] by all participants *any* first-order utterance can have second-order effects which can be transient (in the case of slips of the tongue) or longer lasting (as in the coinage of a neologism). That is, every first-order linguistic act has the potential to change the criteria by which we conceive of that very act, which means that understanding should be conceived of as a context-dependent convergence on a mutual understanding, that is, not as the transference of some-thing from one person to another.

Note also that novelty in linguistic activity is – like second-order practices – not an optional extension. This is true of course from a sociocultural perspective, since language had to originate from somewhere, but also from an ontogenetic point of view. For understanding is most apparent in a person's

31 Davidson couches his discussion in terms of a convergence on a passing theory. However, by theory Davidson means what would amount to a satisfactory description of what a linguistic communicator can do, without implying that they actually know the theory in any sense of the word (Davidson 1986, p. 256).

32 course, novelty can be understood but not accepted, e.g., when corrected. Crucially, however, correcting another person relies on understanding to a degree what they wanted to say. If you simply do not understand at all what another person says you cannot correct what they said.

ability to use a word in a novel context and in novel phrases. That is, merely being able to repeat verbatim shows no understanding at all. It is only when someone can say something 'in their own words' or apply a concept to a novel situation that we are convinced that they understood something.

2.4.4 Implications for linguistic knowledge

The idea of a context-dependent convergence on a mutual understanding has three important implications for linguistic knowledge. *First*, as criterial relations are contextual, and language admits of novelty, there can be no rules for arriving at a mutual understanding. Understanding happens in the unfolding of an episode of communication, and cannot be explained solely by what each participant brings to the table. Like perception in radical embodied approaches, understanding is something that we do, but, crucially, something we do together.[33] The absence of rules for arriving at a mutual understanding is further corroborated by the fact that understanding is often dependent on knowledge that is not traditionally understood to be linguistic. In recognising the prevalence of novelty, Davidson (1986, p. 265) thus concludes that 'we have erased the boundary between knowing a language and knowing our way around in the world generally'. This conclusion opposes the received view mentioned in the introduction on which linguistic knowledge is clearly delineated from other kinds of knowledge, because it is theoretical knowledge of *a language*. Note further that this conclusion is in agreement with the integrationist insight that it is impossible to distinguish in an absolute sense those aspects of a communicational episode that are linguistic from those that are not. It is therefore in principle impossible to delineate sharply linguistic knowledge from non-linguistic knowledge.

33 Even in the *prima facie* example of solitary understanding, such as understanding a text one is reading, one can only be certain one has understood it when one agrees with others in discussing the text.

Take knowledge of English spelling, a prima facie paradigm example of linguistic knowledge. Knowing-how to spell English words, however, means knowing that there is a difference between American English and British English, and that, depending on the context, either *behavior* or *behaviour* is correct. Knowing how to spell correctly thus includes knowing when to use which form, and this in turn relies on knowledge that is not *prima facie* linguistic.

Second, criterial relations hold not only between assertions. In producing linguistic behaviour, we rely on a host of resources in making ourselves understood. Criterial relations therefore hold not only between utterances, but also between *prima facie* non-linguistic actions and utterances. In the example we looked at earlier, opening the window at request showed an understanding of that request. The fact that the addressee starts acting without requests for elucidation thus provides criterial support – in this particular context – for her having understood it. In realising that the distinction between linguistic knowledge and non-linguistic knowledge is not clear-cut, we thus also have to acknowledge that criterial relations hold not only between utterances, but between any aspect of our behaviour and utterances.

Finally, in highlighting the context-dependency of converging on a mutual understanding, what it is that we understand cannot be the utterance itself – for there is no such thing as the utterance itself without theoretically misbegotten decontextualisation. Instead, what we understand is a person trying to get a point across by means of language. So, instead of understanding *language*, we understand *others* (Simpson 2010). This further underlines the blurring of the distinction between linguistic knowledge and knowing our way around in the world generally.

Summarising this section, I argued that linguistic behaviour is grounded in skilful use of linguistic repeatables; this behaviour is skilful because it is

subject to correction. This correction is based in the (dis)similarities perceived between occurrent and previous utterances based on metalinguistic experience, as well as criterial relations to which people attune through metalinguistic practices. What criterial relations apply is not decontextually determined. In any linguistic episode, criterial relations can be changed and constituted without the need for explicit metalinguistic discourse. This shows a blurring of the distinction between linguistic and non-linguistic knowledge, as well as a wider application of the notion of criterial relations. In the concluding section I revisit Love's distinction from the perspective of this account of linguistic knowledge.

2.5 Love's distinction revisited

In this paper I proposed an account of linguistic knowledge in terms of knowing-how, which firmly places linguistic knowledge within the reach of theories that propose a radically embodied account of cognition. In this final section, I revisit Love's distinction from the perspective of the account of linguistic knowledge developed in this paper.

I start with the relatively trivial observation made by Wittgenstein (1953, §97) that metalinguistic constructs, such as *language*, cannot be but 'humble' words like any other. This entails that all second-order practices are necessarily also first-order practices. In other words, although arising out of reflexivity, metalinguistic practices are not outside language in the way that a tennis commentator is outside the match he talks about (Harris 1998, p. 25). As opposed to tennis, 'language, being language, is on its own. It is interpretatively terminal.' (Love 2003, p. 88; 2007).

As mentioned in the introduction, Pablé and Hutton (2015) claim that 'first and second-order practices are inextricably intertwined'. We are now in a position to elucidate this claim. It is crucial to realise that the ability to parti-

cipate in metalinguistic practices does not leave the first-order activities untouched. Think for instance of learning English grammar. To be competent in English grammar is not merely to take a metalinguistic perspective on first-order linguistic activity or linguistic inscriptions every once in a while; one's own first-order linguistic activity has to be in accordance with English grammar to count as being competent. To say that metalinguistic practices are normative with regards to first-order linguistic activity is to take account of the shaping force that they have – subject of course to participants' recognition of the pertinence of the particular metalinguistic practice.[34]

This point is also made by Noë[35] (2012, p. 3), who describes the relation between first-order linguistic activity and second-order practices as follows: 'The first-order practice contains within it the second-order practice. For thinking about language and language-use is one of the basic and unavoidable things we use language to do. There is no such thing as the naïve, unreflective, theoretically unbiased user of language.' The impossibility of an unreflective user of language is also the conclusion reached in this paper: to be able to participate in linguistic practices at all, one has to be able to participate in metalinguistic practices. If we conceive of linguistic knowledge as including knowing-how to provide and recognise criterial support, this means that a competent linguistic communicator has to talk and write in such away that she can provide this criterial support. This in turn requires an ongoing reflective attitude towards one's own and others' first-order linguistic activity. Being able to participate in metalinguistic practices is therefore not properly described merely as being able to switch rapidly between unreflective first-order linguistic activity and detached metalinguistic reflection on that activity. To be a

34 To be competent in English grammar does not entail, for instance, that one's linguistic behaviour is *always* in accordance with it. Everybody that has transcribed spoken language verbatim knows that spoken language often does not consist of well-formed sentences.

35 Noë (2009, p. 198) acknowledges the influence of Roy Harris on his own thinking about language.

competent language user is perforce to be a competent critic of language (Noë 2012), and one can only be critical when one recognises when one needs to be. Claiming that metalinguistic practices are constitutive of linguistic knowledge, in the end, boils down to claiming that to be a competent linguistic communicator, one has to have a reflective – or metalinguistic – attitude towards what one is doing *whilst* being engaged in linguistic activities. It is to perceive directly when a person uses a word incorrectly, to see the spelling error as one reads the text, to understand that someone unfamiliar introduces themself when they utter a name, to notice when you don't understand what is being said, &c. To be a competent linguistic communicator one has to be sensitive to the criterial relations that originate from metalinguistic practices in the course of first-order linguistic activity. All first-order linguistic activity thus has a second-order dimension to it, in the sense that it is directly perceived in terms of the metalinguistic practices that constitute it.

3 Attentional actions – An ecological-enactive account of utterances of concrete words

The cognitive sciences are currently making a *pragmatic turn* (Engel et al., 2013). The dominant *representational theory of cognition*, where cognition is conceived of as computation over mental representations, is gradually giving way to the ecological-enactive approach, which foregrounds the action-oriented, embodied, and situated nature of cognition (see §3.1.3 for a short introduction). This pragmatic turn has implications for understanding linguistic behaviour. In this paper, I reconceive the cognitive function of utterances of concrete words, words that concern observable situations, events, objects, or characteristics. There are two reasons for starting with concrete words. First, concrete words are a paradigmatic example in the dominant code-view in the representational theory of cognition, according to which words function as codes for mental representations (see §3.1.1). Second, the empirical research on the relation between language and cognition on which this paper draws often relies on these concrete words, in particular colour words (see §3.2).

In this paper, I propose a way to understand utterances of concrete words from the ecological-enactive approach by starting from the *constraint view*. On this view, language is reconceptualised as a mode of action that functions as enabling constraints on cognitive and interactive dynamics. Building on the ecological-enactive conception of attention, I define an attentional action as a repeatable form of behaviour performed by a person to indicate a particular aspect of the current situation to someone in order to achieve something. I argue that a concrete word, whether performed as a single-word utterance or in the context of a grammatically structured utterance, has to be understood in terms

of its constraining effect on ecological-enactive attention. In other words, utterances of concrete words are attentional actions.

This paper serves two purposes: first, it undermines the familiar scope objection against the ecological-enactive approach, according to which it can only explain basic cognition and cannot explain 'higher' cognition such as linguistic cognition. Second, it provides a novel explanation for recent empirical evidence with respect to the role of language in categorical colour perception.

This paper is structured as follows: first, I introduce the code view and the ecological-enactive approach. I propose an account of a child's first utterances of concrete words as attentional actions based on the constraint view of language and a discussion of learning language in line with the education of attention model of learning (§3.1). Then I discuss some recent empirical findings that show verbal colour categories to have an effect on putatively non-linguistic colour perception (§3.2). In line with this evidence, I propose an account of the cognitive effects of attentional actions in terms of constraints on the phenotypic reorganisation of task-specific devices (§3.3). In the final section, I situate the account of attentional actions in a larger theoretical perspective on language by means of remarks on the context-dependency and reflexivity of language, grammatical structure, and written language (§3.4).

3.1 Words: from codes to constraints

3.1.1 The code view

Entrenched in our thinking is a picture of language as a carrier of thought. We find articulations of this picture in everyday English expressions such as putting a concept into words, or a text containing new ideas.[1] In theoretical reflections on language, this picture manifests in the idea that words encode mental mean-

1 Reddy (1993) lists a great number of examples from everyday English.

ings. This has been dubbed the *code view* (Harris 1990; Linell 2005; Love 2004; Kravchenko 2007). On this view, communication consists in the encoding of mental meanings in public symbols and the subsequent decoding of these public symbols back into mental meanings.

The code view lies at the root of many theories of language. Notable historical proponents of the code view include Aristotle (1975, p. 43), who claimed 'spoken sounds are symbols of affections in the soul', and John Locke (1690/2004, p. 363), who took words to be 'sensible marks of ideas'. De Saussure (1922/2011, p. 66) defined the linguistic sign as a connection between a concept and a representation of a linguistic sound. He described linguistic communication in terms of the speaking circuit: a concept 'unlocks' a sound representation, which is transmitted to the organs used in producing speech, where they are transfused into actual sound patterns which then travel to another person in order to undergo the reverse operation (Ibid., pp. 11–12). The second half of the 20th century witnessed the rise of *the representational theory of cognition*, which enabled the formulation of new versions of the code view. According to one version of this theory, thought itself is language-like, and understanding public language is essentially a code-like translation to a language of thought, called *mentalese* (Devitt 2006; Fodor 1975; 2007; Pinker 1994; 2007). Other versions of the representational theory of cognition rely instead on the broader notion of semantic or conceptual representations. For example, Cairns and Cairns (1976, pp. 17–18) claim that 'The listener, B, must decode A's message by converting the sounds into a semantic representation'. Nowadays, the idea that words encode semantic representations is a staple in linguistics textbooks. In *The Blackwell Handbook of Linguistics*, for example, Cruse (2017, p. 253) states:

> Each of us has in our cognitive system some kind of inventory of all the words that we know, together with all the information – semantic, grammatical, and

> phonetic / graphic – necessary for their correct use. […] The inventory is accessed via written or spoken forms every time we hear or read something in a language we know, and via some kind of semantic representation every time we produce language.

In *The Lexicon: An Introduction*, part of the *Oxford Textbooks in Linguistics series*, Ježek (2016, p. 5) claims that the 'existence of a word' is the result of 'the direct association of a concept with a lexical form'. Proponents of the code view thus differ in what they take words to be encoding: concepts, ideas, semantic representations, etc. For the purposes of this paper I shall use *mental representation* as an umbrella term to denote this panoply of mental meanings. The code view can then be defined as follows: A word is a code for (a set of) mental representations. Note that due to synonymy and polysemy, the mapping between words and (sets of) mental representations cannot be a one-to-one mapping, but instead is a many-to-many mapping (Cruse 2017, p. 252).

I introduce the code view in order to use it as a foil. The reason for this use is that the code view has had a deep influence on Western theoretical reflections on language and remains the dominant view, although there are of course alternative views.[2] The purpose of this paper is neither to determine how widespread the code view is, nor to provide arguments to criticise it, but instead to use it as a contrast. This enables bringing out the defining features of the constraint view to which this paper aims to contribute.

3.1.2 Mental representations: can't have, don't need

Proponents of the pragmatic turn in the cognitive sciences claim that cognition is embodied. Yet, there is substantial disagreement about what this claim means (Anderson 2003; Kiverstein & Clark 2009; Wilson 2002; Wilson & Golonka

2 Examples of alternatives include Ryle, who does not believe that there are mental meanings (Tanney 2015), the later Wittgenstein (1953/2009), Voloshinov's (1930/1973) Marxist philosophy of language, Maturana's (1978) biological approach to language, Bakhtin's (1982) dialogism, integrational linguistics (Davis & Taylor 1990; 2003) and the distributed language perspective (Cowley 2011a).

2013). In its weakest interpretation, an 'embodied concept is a neural structure that is part of, or makes use of the sensorimotor system of our brains' (Lakoff & Johnson 1999, p. 104). As opposed to *classical cognitivism*, mental representations are no longer amodal. But cognition is still a computational process over discrete mental representations on this weak interpretation. The strongest interpretation, known as *radical embodied cognition*, aims to provide a thoroughly non-representational approach.[3] Note that a non-representational approach forswears *mental* representations. This does not entail forswearing *public* representations such as maps, assertions, diagrams, charts, or flatpack furniture assembly instructions.

The account in this paper is based primarily on two research programmes that espouse a form of radical embodied cognition: ecological psychology (Chemero 2009; J.J. Gibson 1979; Rietveld & Kiverstein 2014; Turvey, Shaw, Reed, & Mace 1981) and enactivism (Hurley 1998; Hutto & Myin 2013; O'Regan & Noë 2001; Stewart 2010; Noë 2012; Varela, Thompson, & Rosch 1991). I therefore speak of the *ecological-enactive approach*. Note that both enactivism and ecological psychology are research programmes within which different theories are espoused, of which I will not provide an exhaustive taxonomy here.

A proponent of the ecological-enactive approach to cognition can employ two lines of argument with respect to the representational theory of cognition: *can't-have* and *don't-need* arguments (Hutto & Myin 2013). Can't-have argu-

3 Named as such by Clark (1997, p. 148), who defines it as follows: 'Structured, symbolic, representational, and computational views of cognition are mistaken. Embodied cognition is best studied by means of noncomputational and nonrepresentational ideas and explanatory schemes.' Proponents that fly under the banner of radical embodied cognition include Chemero (2009) and Hutto and Myin (2013; 2017). The word radical here should not be understood as a radicalisation of embodied cognitive science, but instead in the sense of having different roots. These theoretical roots include American pragmatism (Peirce, James, Dewey, and Rorty), behaviourism (Ryle and Skinner), ecological psychology (James and Eleanor Gibson), theoretical biology (Maturana and Varela), dynamic systems theory (Thelen & Smith, 1994; Van Gelder, 1998), biophilosophy (Jonas), and existential phenomenology (Heidegger and Merleau-Ponty).

ments are negative. They aim to show that there are currently no naturalistically credible theoretical resources available for accounting for mental representations. Hutto and Myin (2013), for instance, argue that there is no satisfactory account of how mental representations get their representational contents. If a can't have argument is successful, representational approaches to cognition are off the table. However, as Chemero (2013, p. 3) points out, can't have arguments are likely to be met with a 'collective shrug'. That is, these arguments against the representational theory of cognition tend not to convince proponents of that theory. Don't-need arguments, on the other hand, are positive. They aim to show that mental representations need not be invoked in order to explain a particular cognitive phenomenon. If a don't need argument is successful, this does not imply that the non-representational explanation should automatically be preferred over an alternative representational explanation. However, meta-theoretical considerations – such as ontological parsimony and explanatory power – can then be used to mount arguments in favour of either explanation.

Note that the two strategies can be employed independently. It is possible to argue that we can't have mental representations without supplying a non-representational alternative explanation, and it is possible to propose a non-representational explanation without arguing that that we can't have mental representations. The intended contribution of this paper is only to provide a don't need argument. This is done by showing a possible account of utterances of concrete words that does not invoke mental representations, and show how it can account for recent empirical evidence.

3.1.3 The ecological-enactive approach in a nutshell

The guiding idea of the enactive approach is that perception and action are intimately intertwined. In the dynamic coupling between organism and environment, a feedback loop is set up (see Figure 2). This is called a *sensorimotor*

loop (Stewart 2010). Cognition is defined as *perceptually guided action* (Varela et al. 1991, p. 173), and thus becomes something the organism does in interaction with its environment (Myin 2016; Noë 2012). Cognitive processes do not take place inside the head of the organism, but are relational in nature.

Ecological psychologists similarly stress the intertwining of perception and action in the concept of *affordance* (J.J. Gibson 1977). Organisms perceive what the environment *affords* doing, that is, possibilities for action. For example, a flight of stairs is perceived as climbable and a cup is perceived as graspable.[4] The perceived affordances depend *inter alia* on aspects of the environment, the skills of the organism (not everything that is liftable for a world champion weightlifter is liftable for me), as well as the material possibility we have in exercising these skills, for instance the availability of artefacts. As Stewart (2010, p. 29) remarks, 'a snow-covered mountain becomes an entirely different place if you have skis on your feet (and if you know how to ski!)'.

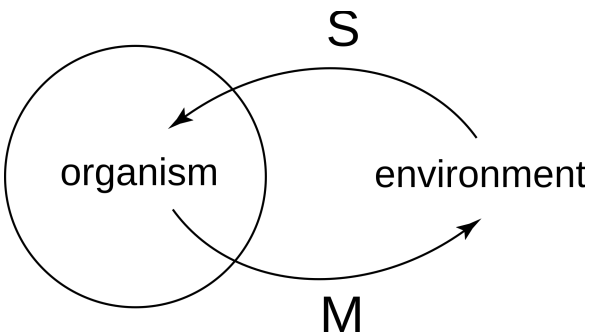

Figure 3. The sensorimotor loop: a feedback loop consisting of sensory (S) and motor (M) processes (based on Stewart (2010, Figure 1.1)).

4 A good example of this is the explanation Zhu & Bingham (2010) give for the the size-weight illusion. This illusion consists in the fact that a smaller object that has the same mass as a larger one is incorrectly judged to be heavier. However, the illusion disappears if we take people to perceive the 'throw-ability' of the objects. A larger object has to have a greater mass to be just as suitable to throwing as a smaller object. Zhu and Bingham found that the larger object is indeed perceived to be just as heavy as the smaller object when it is equally 'throwable'.

We never perceive only one affordance. An individual always perceives a *field of affordances*, that consists of a myriad of possibilities for action afforded by the environment (Rietveld & Kiverstein 2014). Not all of these affordances will invite behaviour equally. Which affordances are more inviting will depend on a host of factors. Whether something is perceived as liftable will for example depend on social setting (a statue in a museum does not afford lifting) and the degree to which a person feels fatigued. Even things such as mental health, for example, have an important role to play: for a depressed person, much less of the environment will be perceived as liftable (De Haan, Rietveld, Stokhof, & Denys 2013).

3.1.4 Ecological-enactive learning as the education of attention

On the ecological-enactive approach, the goal of cognitive processes is not truthful or accurate representation, but perceptually guided action. The notion of representation is replaced with *attunement* (J.J. Gibson 1966). How does a child attune to its environment? The ecological answer is through the *education of attention* (Adolph & Kretch 2015; E.J. Gibson 1953; E.J. Gibson 1963; J.J. Gibson 1966). Attention is understood as selective openness[5] to the field of affordances in relation to a task or goal (E.J. Gibson & Rader 1979; Rietveld & Kiverstein 2014). Attention can only be evaluated with respect to a particular task or goal. In line with the definition of cognition as perceptually guided action, goal-directed behaviour is enabled by a person's sensorimotor skills for acting on relevant affordances. The education of attention thus consists in the development of skills that allow for the context-dependent perceptual guidance of goal-directed behaviour. In other words, as a person is attuned to her environment, she thereby learns how to act on those affordances that are relevant.

5 See Bruineberg and Rietveld (2014) for an ecological-enactive account of the neural dynamics underlying this selective openness.

In the case of humans, the education of attention mostly takes place in social interaction through processes of *guided rediscovery* (Ingold 2001; Zukow-Goldring & Ferko 1994). This means that a master of a practice shows or makes present affordances to a beginner in such a way that the beginner can learn a particular skill or activity. Note that this process need not be explicitly didactic, but can simply consist in doing something together (Reed 1995). In child development, the education of attention highlights the active role of both caregiver and child. The caregiver is responsible for setting up the physical conditions that allow the child to learn.[6] Crucial here is the Vygotskian idea of the *zone of proximal development*, which comprises activities that the child can engage in only with the help of the caregiver (Reed 1995). Besides physical structuring, the education of attention also relies on normative and cognitive structuring of the situation by the caregiver (Williams 2010b). That is, the caregiver's competence in her community's practices allow her to determine what the child is, could, or should be doing, and encourage and discourage the child's behaviour on this basis.[7]

These conditions also hold for the development of more intellectual skills. In learning to talk, a child's utterances do not start out as complete speech acts. Instead, the caregiver treats the child as if she is already competent and on this basis encourages and discourages the child's behaviour.[8] For example, a

6 For example, Zukow-Goldring and Ferko (1994, pp. 172–173) describe a caregiver teaching a child how to roll a ball. The caregiver sits the child down with legs apart, a posture that provides a guide for the trajectory of the ball whilst at the same time allowing the child freedom of movement from a stable position. The caregiver then begins the activity by gathering attention through verbal means (*Look!*), accompanied by moving the ball in the child's line of sight. After the child's attention has been attracted, the caregiver rolls the ball over to the child, and helps the child to roll the ball back, for instance by placing the infants hands on the ball and guiding the return. In this way, the child can gradually learn the activity of rolling a ball.

7 For example, Rączaszek-Leonardi, Nomi, and Rohlfing (2013) describe how a six-month infant that is being changed accidentally touches a diaper. The mother responds to this contingency by placing the diaper in the child's hand while saying 'yes you can start by holding the diaper'. In this way, the infant's essentially random arm movement is incorporated in the goal directed activity.

8 Goldstein, King & West (2003) show that 8 month old children are already highly sensitive to

child uttering something vaguely like /ˈteɪbəl/ in the presence of a table can be encouraged by a caregiver by saying *Very good! That is indeed a table!*, even though nothing in the behaviour of the child shows that she intended to say of is thing that it is a table. Treating the child *as if* she is making an assertion is crucial here. If we were to take the child to be engaged in mere meaningless babbling, there would be no point at all in encouraging or discouraging her behaviour.

In line with the intimate intertwining of perception and action, successfully participating in her community's practices requires the child to see her world in the correct way (Gallagher 2005). She has to be able to pick up on those affordances that are salient with regard to the practices of her community. Through the education of attention, the teacher gradually introduces the child in the practices of her community. Educated attention is manifest in altered selective openness: a person is now able to pick up on and skilfully act on affordances of her environment that earlier did not show up for her.

3.1.5 What a child learns as she learns to speak

On the code view, language learning consists in forming the correct associations between words and mental representations.[9] Ryle (1945a) made a distinction between knowing-that and knowing-how, roughly the distinction between knowing a proposition and having a skill. The code-view assumes linguistic knowledge to be a species of *knowing-that*: knowing a word is knowing *that* it stands for (a set of) mental representations. This does not entail that the code view implies that learning language can have no effect on how the world is

positive responses to their vocalisations, a sensitivity that goes beyond mere imitation.

9 Papafragou (2005, p. 357), for example, states that 'the main task of the [language] learner is to figure out which aspects of the input language correspond to which nonlinguistic conceptual primitive notions – or combinations thereof'. Another example, that shows clearly the idea that meanings come before words, is given by Clark (2009, p. 8): 'One issue for language acquisition is how children find out which meanings there are words for; another is just how they map each meaning to the right word.'

represented. It does entail that the child has to mentally represent the world *before* she learns language. On an ecological-enactive approach, this need not be the case. Emphasising the normative structure of the education of attention means that we do not have to assume that the child can already perceive in a categorical way. Instead, by means of the education of attention, the child learns how to 'use words' and *thereby* learns to perceive categorically. In keeping with the relational perspective of the ecological-enactive approach, categorical perception does not rely on mentally representing categories, but consists instead in having the skills for making distinctions between things in the world in line with a community's standards.

This is in line with the observation that a child learns to communicate in 'the midst of "doing"' (Bruner 1990, p. 70). A child is not a passive observer, but participates *in* activities and it is in these activities that she starts 'doing things with words' (Rączaszek-Leonardi 2009, p. 170). Viewed from this perspective, the language learner does not acquire knowledge-that *about* language (Taylor 2013, p. 317). She does not need to learn that a particular words stands for particular things in the world. Instead she comes to *know-how* to do things by talking (Van den Herik 2017 [Chapter 2]). The question of the acquisition of a language is substituted for asking 'how the child comes to enter the linguistic community' (Reed 1995, p. 2) which she does by learning a repertoire of social skills (Reed 1996; Verbrugge 1985).

This reconceptualisation of language learning has two important implications. First, it shows that linguistic behaviour is always *time-bound* and *context-bound* (Love 1990). According to what Harris (1998, p. 81) calls the principle of co-temporality, what is being said is immediately relevant to the current context, and makes sense or fails to make sense within that particular context. In other words, successful linguistic communication consists in the ability of two speakers to converge on a shared understanding (Cuffari, Di Paolo, De Jaegher

2015; Davidson 1986), an achievement that relies on bringing all kinds of knowledge to bear (Van den Herik 2017 [Chapter 2]). Second, language 'must be based on social coordination' (Cowley 2011b, p. 11). This is evident in ontogenetic development of linguistic abilities: children learn language in a triadic relation, with a caregiver jointly attending to things, in which the focus is on what one can do with language. Language is therefore essentially *dialogical* (Cowley 2011a). For example, prior to learning its first words, infants already participate in interactions that involve turn-taking behaviours under the guidance of a caregiver (Rączaszek-Leonardi 2016).

3.1.6 Utterances of concrete words as attentional actions

To recapitulate, according to the ecological-enactive approach, cognition is a loopy, dynamic, and relational affair. Through the education of attention the developing child gradually attunes to her environment, a process that relies constitutively on both physical and normative structuring of the learning situation by a caregiver that is already competent in the community's practices. However, this leaves open the question what we do with words.

The fact that the ecological-enactive approach eschews mental representations means that the code-view will not do. I start from a different picture of language: *the constraint view* (Rączaszek-Leonardi 2011; 2016; Rączaszek-Leonardi & Kelso 2008; Verbrugge 1985). On the constraint view, language is viewed as a system of social actions that function by constraining unfolding cognitive and interactive dynamics. Because of this reliance on unfolding dynamics, linguistic actions neither have to stand for the processes they constrain, nor do they have to map like a code to the results of their constraining action. In line with the action-oriented nature of cognition, language is reconceptualised as a mode of action. Conceiving of language in terms of constraints brings together the two implications just discussed: ongoing processes are

constrained by utterances in real time (language is time-bound and context-bound), and constraints presuppose a communication situation on which the constraining action can take place (the social and dialogical nature of language).

The education of attention model of learning can explain how language gets to have these constraining effects. Language can fulfil the constraining role because it consists of recognisable and repeatable forms of behaviour, which I refer to as *repeatables* (Van den Herik 2017 [Chapter 2]). Within the education of attention, the repeatables we construe as words 'anchor' attention[10]: by being easily recognisable, they provide a fixed point that allows for the stabilisation, or sedimentation, of the education of attention.

Different kinds of linguistic actions have different kinds of constraining effect. For the purposes of this paper, I further look into the constraining effects of concrete words from the perspective of ontogenetic development. I start from concrete words because they are a paradigmatic example for proponents of the *code-view*, and empirical research on the relation between language and cognition, which I will consider in the next section, usually involves only concrete words.

The ecological-enactive alternative I propose, however, does not start from 'concrete words' in the sense that code views do. It starts from linguistic actions, or utterances. From the perspective of the child there are no words, there are only repeatables: forms of behaviour that can be *re*cognised from one instance to another, and that can be reenacted. In linguistic behaviour, concrete words usually do not occur in isolation, but rather in the context of a grammatically structured utterance. Notable exceptions are the naming games caregivers play with children, where they point to an object and utter the word for that object, and the child's first utterances. A child's production of language starts

10 I take this image picture of words anchoring attention from Clark's (1996) reference to Jackendoff (1996). Given the ecological definition of attention in play in this paper I put it to different use.

out with single-word utterances before she starts combining words into structured sequences (Clark 2009, p. 115; Fenson et al. 1994).

For the purposes of this paper I start by looking at a child's single word utterances of concrete words. The hypothesis I propose is that a child's first utterances of concrete words are attentional actions. An attentional action is defined as a repeatable performed by a person to *indicate* a particular aspect of the current situation *to* someone *in order to* achieve something. In this paper, I only claim that we can understand a particular kind of linguistic utterance, namely those featuring concrete words, as attentional actions. I do not mean to say that only those utterances are attentional actions, or even that only utterances can be explained as attentional actions. It seems defendable to say, for example, that ostensive gestures are also attentional actions. For the purposes of this paper, however, I will not pursue the question how generalisable the account of attentional actions is.

Let us unpack this definition. It builds on the notion of indication developed by Reed (1995). The child's utterance is not a label, or a name, but a skilful way of directing ecological attention. In other words, we can understand an attentional action as a constraint on attention. While Reed limits the indicational phase to roughly the second year of a child's development, I argue that this account of concrete words as attentional actions can be generalised to adults' grammatically structured utterances. I return to this issue in §3.4.2. In first instance, indication is a second-personal skill. We direct somebody else's attention, and whether the attentional action will succeed depends as much on the person that performs it as the person to which it is addressed. Derived from the primordial second-personal action, the person that produces the attentional action can coincide with the addressee, as when we talk to ourselves to direct our own attention.

Note that the definition states that only aspects of the current situation can be indicated. Crucially, the current situation is not limited to what is currently perceivable. A toy that is put away in the closet can be indicated in the same way as a toy that is in view. For the ecological-enactive approach 'there is no sharp line to be drawn between that which is and that which is not perceptually present' (Noë 2012, pp. 25–29). In general, non-representational approaches such as the ecological-enactive approach often face the objection that they cannot deal with 'representation-hungry' phenomena, such as cognition involving *absentia* or *abstracta* (Clark and Toribio 1994). See Kiverstein and Rietveld (2018) and Degenaar and Myin (2014) for ecological-enactive solutions to this objection. It is outside the scope of this paper to go into these solutions. For now it is important to see that whereas a code-view conceives of language learning in terms of learning *the* meaning of a decontextualised word, from an ecological-enactive perspective the situated nature of linguistic actions is foregrounded.

There are four elements involved in an attentional action: the person that performs the action, the person that is addressed, the indicated aspect of the situation, and the goal of the attentional action. Attentional actions can be metaphorically 'anchored' to certain aspects of the situation in the unfolding of the education of the attention. In §3.3 I look at the cognitive mechanism that enable this 'anchoring' in more detail. For now, it is important to note that attention is understood in ecological terms, that is, as the selective openness to the field of affordances in relation to a task or goal. To direct attention by means of an attentional action is thus to foreground an affordance of the current situation, some possibility for action. By thus promoting the salience of certain affordances present in the current situation, the chance that they will be acted upon in the unfolding of the situation is increased. These skills, required for constraining attention and having one's attention constrained, are brought to

fruition in the education of attention, which occurs in the context of caregiver and child doing something together.

Realising that attentional actions stem from, and are performed as part of recurrent and structured social situations is crucial for dealing with Quine's question (Reed 1995): how does a child learn to refer to the same things that their caregivers refer to? Quine (1960) identified the fundamental problem of the indeterminacy of reference. This indeterminacy of reference spells problems for a code-view account of language learning. The argument is as follows. The code view assumes that a child has to learn what words refer to, which is accomplished by associating the word with the correct mental representation. Adults teach a child a word by pointing out an object and saying the word. This ostensive gesture, however, is not enough to specify what is referred to. The reason for this is that pointing out a dog is indistinguishable from pointing out this particular dog, or a standing-rather-than-a-sitting dog, or all and sundry undetached parts of a dog, or any other fanciful fictions. There is thus a referential ambiguity inherent in ostensive gestures. This means that the child has insufficient evidence for determining what the referent of the ostensive gesture is, and therefore cannot decide on the basis of the ostensive gesture alone what mental representation needs to be associated with the word 'dog'.

In order to be able to answer Quine's question, code-view theorists usually assume that there are a priori constraints on the hypotheses that children entertain with respect to word-meanings (Clark 2009, p. 124). For example, children are thought to assume that words refer to a whole objects (*the whole-object assumption*) and taxonomic kinds, rather than thematically related objects (*the taxonomic assumption*; Clark 1981, p. 40; Markman 1981). By means of these *a priori* constraints, the referential ambiguity is dissolved and a child is able to arrive at the correct hypothesis, for example that 'dog' refers to dogs.

If we, however, assume that utterances of concrete words are attentional actions, Quine's question can be evaded. To understand why this is the case, we have to realise that the environment of the child is highly structured, and consists of recurring situations, routines, and activities. It is in these activities that a child starts responding to and performing attentional actions. The task that the child is faced with is becoming an active participant in the activities she engages in with her caregivers. Unsurprisingly, the first twenty words that a child produces typically pertain to salient aspects of recurrent situations that are highly relevant to the infant, such as people (*daddy, mommy*), games and routines (*bye, night-night*), animals (*dog, kitty*), and toys (*book, balloon*) (Fenson et al. 1994). Of these twenty first words, fourteen are concrete words.[11]

An interesting observation is made by Clark (2009, p. 88), who notes that when children start using language, they use a lot of deictic terms like *that* and general purpose verbs like *do*. As she points out, 'without contextual details … it is usually impossible to interpret such utterances', whereas, in 'context, with the aid of joint attention, it is normally quite clear what children are talking about when they do this.' The crucial insight is that this line of reasoning holds for all of the child's initial linguistic behaviour, and continues to hold, although to a lesser degree, even for fully competent adult linguistic behaviour.[12] This is because attentional actions function similar to ostensive pointing gestures: they are performed in a context in order to foreground a particular aspect of the situ-

11 The concrete words are *daddy, mommy, dog, ball, baby, book, shoe, kitty, bird, eye, balloon, bottle, banana*, and *juice*, the other words are *bye, hi, uh oh, no, woof*, and *night night* (Fenson et al. 1994, Table 15, p. 93).

12 For an extension to competent adult language use, we can remind ourselves of Ryle's (1945b, p. 215) claim that all words incorporate a systematic ambiguity: 'A given word will, in different sorts of context, express ideas of an indefinite range of differing logical types and, therefore, with different logical powers.' The example he gives is that of punctual, which 'can be used to characterize a person's arrival at a place, the person who arrives there, his character and even the average character of a class of persons'. Comparing these different kinds of punctuality would be absurd, according to Ryle, and thus he concludes, 'where precision is wanted, it is wrong to speak of "the idea" of punctuality' (*Ibid.*, p. 216). This conclusion is in line with our rejection of the code-view.

ation, and without contextual details, they too are impossible to interpret. Quine's question spells trouble only if we take the ostensive gesture as it is used in the education of attention as evidence or determining the reference. If we instead understand the attentional action to, as it were, take the place of the ostensive gesture, we do not have to answer Quine's question.

In other words, the attentional action inherits the referential ambiguity of the ostensive gesture by means of which the education of attention is initially achieved. This is possible because an attentional action is a species of ostensive gesture. As Baggs (2015, p. 260) puts it, speaking is 'a technique for pointing'. The referential ambiguity only emerges from an observer-perspective, once an observer asks himself what *the* reference of *this* word is. Learning to perform an attentional action does not require taking a stand on what a word refers to. Determining the extension of an attentional action (that which can be indicated by performing the attentional action) would be like determining the extension of the pointing gesture abstracted from the context in which the pointing gesture occurs: a pointless exercise.

Crucially, the referential ambiguity usually does not show up from a participant's perspective, that is, from the perspective of the child or the care-giver in the education of attention, and is not detrimental to communicational success. The reason is that the object of indication is individuated, not by decontextualised referential knowledge (of the form *'dog' refers to dogs*), but by the unfolding of the (recurrent) situation, in which only certain affordances are relevant, and thus can become more salient as a result of the attentional action.

At the same time, to highlight the situatedness and ostensive nature of indication by means of attentional actions is not to say that there is no potential for misunderstanding or conflict. It is crucial to note here that *indicate* is used as a 'verb of success' (Ryle 1949/2009, p. 114). As mentioned, whether someone succeeds in indicating something to someone is not only dependent on her beha-

viour, but lies in the successful coordination of behaviour. In other words, 'indication of a single object, place, or event, is an *achievement* of a dyad, not a given.' (Reed 1995, p. 11). In order to play her part in this achievement, a child needs to be able to successfully coordinate her behaviour with others by means of attentional actions. This means that the success conditions are not specified by the attentional action, but emerge in the unfolding of the situation and are to be decided upon by the interacting dyad.

To summarise, on the constraint view, concrete words do not stand for anything. Learning to use these words does not consist in acquiring knowledge-that about their meaning. Instead, learning to talk is learning a repertoire of social skills. A child's first utterances of concrete words are used to indicate aspects of the current situation to others in order to achieve something, and are therefore understood as *attentional actions*. An attentional action is like an ostensive gesture: it points out something. Learning takes place by means of the education of attention, where attentional actions are 'anchored' to aspects of the environment. In order to understand better how this 'anchoring' functions, we now turn to empirical research into the effect of colour words on categorical colour perception.

3.2 The case of colour

Categorical colour perception has been the preferential paradigm for studying the influence of language on putatively non-linguistic processes, because it is a prime example of categorical perception that allows for easy cross-cultural comparison. In particular, research has focussed on category effects, which can broadly be understood as any effect of verbal categorisation on colour cognition; for example, an effect of verbal categorisation on speed or accuracy in discriminating or remembering colours. In this section, I report the *category*

effects found in this research, in order to draw some implications for understanding utterances of concrete words as attentional actions in the next section.

3.2.1 Basic colour words and cultural differences

Research into categorical colour research relies on the notion of basic colour words. Colour words are basic when they are monolexemic, not subsumed under other colour words (such as *navy*), applicable to all objects (which excludes *blonde* for example), and psychologically salient for a community (Berlin & Kay 1969). In English, these criteria are met by the words *white, black, red, green, yellow, blue, brown, purple, pink, orange,* and *grey*. Other languages, however, employ different colour words that categorise colour space in different ways.[13] For example, the Dani from New Guinea employ just two basic colour words, *mili* and *mola* (Rosch-Heider & Olivier 1972), and the Berinmo have five basic colour words including *nol* which spans colours ranging from *yellow* to *purple* (Roberson, Davies, & Davidoff 2000).

3.2.2 Russian blues

To investigate category effects, Winawer et al. (2007) devised a speeded colour discrimination task that relies on the fact that Russian speakers have basic colour words for lighter blues (*goluboy*) and darker blues (*siniy*), where the English basic colour word *blue* spans these two categories. Participants were shown a triad of blue colour squares and had to indicate as quickly and as accurately as possible which of the bottom two squares was the same colour as the top square (see Figure 3). In so-called cross-category trials, the distractor square

13 There is a debate whether the development of colour terms across linguistic communities is universal. Berlin and Kay (1969) for instance, argued that all linguistic communities go through the same seven stages, starting out by making a dark-cool/light-warm distinction, then adding red, and so forth until they finally make it to the universal end-point which is expressed in the eleven English basic colour words. There however, seem to be counter examples to this alleged universality (Roberson, Davies, & Davidoff 2000). Moreover, the alleged universal order is not found in language learning (Andrick & Tager-Flusberg 1986; Pitchford & Mullen 2002).

was across the *siniy/goluboy* category border, as determined for that each Russian speaker individually by means of an elicitation procedure, whereas in within-category trials the distractor square was in the same verbal category as the target square. Of course, for English speakers all squares always belonged to the same basic verbal colour category, *blue*.

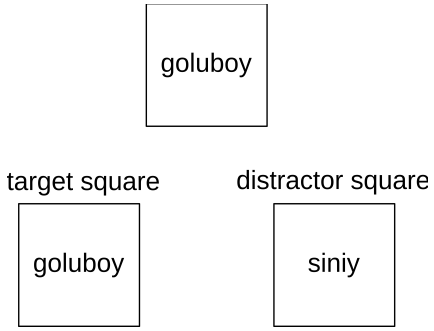

Figure 4. Example of a cross-category trial in Winawer et al. (2007).

The results of this speeded discrimination task were that Russian speakers' performance showed a category effect: their reaction times were significantly shorter in cross-category trials than in within-category trials, an effect that was absent for English speakers. In other words, when the target and distractor square fall on different sides of the *siniy-goluboy* category boundary, Russian speakers are able to identify the target square faster than when both the target and and the distractor square fall in the same category. Interestingly, when asked to differentiate between *light blue* and *dark blue*, English participants drew almost the same border as Russians did between *goluboy* and *siniy*. This shows that English participants can visually distinguish light and dark blues in similar ways to Russian participants.

This experiment evaded the critique levelled against earlier research. For example, Roberson, Davidoff, Davies, and Shapiro (2005) asked participants to categorise a collection of colour swatches, and found that people tended to

categorise the colour chips in accordance with verbal categories. They interpreted the results as showing a category effect: they claimed that it showed that to the participants, colours from the same verbal category looked more alike than colours from a different verbal category. However, this effect can also be explained in terms of a deliberate linguistic strategy: participants could simply put together those colours for which they use the same colour word (Pinker 1994). In the experiment of Winawer et al. (2007), this strategy was not possible: the task did not involve subjective judgements of colour similarity, but consisted in finding the correct answer to a simple discrimination task. Moreover, the effect was measured in the reaction time of participants, a measure that is taken not to be under the participants' control.

Further evidence that category effects are not due to a deliberate linguistic strategy is given by Thierry et al. (2009). They devised an experiment that showed that the category boundary in Greek between basic colour words for light and dark blue has a category effect on implicit colour discrimination in a colour oddball detection task. In this experiment, participants were shown a sequence of stimuli differing in colour and shape and were instructed to press a button when they saw a particular shape. This means that changes in colour were irrelevant to the task. However, Thierry et al. were able to find a category effect using an electrophysiological measure: the Greek distinction between light and dark blues led to a larger visual mismatch negativity ('an index of automatic and preattentive change detection' (p. 4567)) in the case of a deviant colour stimulus as compared to English participants for whom the deviant stimuli fell in the same verbal category (cf. Athanasopoulos et al. 2009).

The category effects just discussed could be explained in two different ways. First, the effect might be due to 'warping' of perceptual space on longer timescales. On this explanation, development in a culture with specific verbal colour categories leads to lasting effects on colour perception. Second, the effect

might unfold on much shorter time-scales. On this explanation, linguistic processes have an online modulatory effect on cognitive processes. A way to decide between these two competing explanations is to compare the interference of verbal dual tasks and other cognitively demanding duals tasks on the category effects. If the category effects obtain due to warping of perceptual space on longer timescales, a verbal dual task should not lead to more interference when compared to a equally demanding non-verbal dual task. If, however, the effect is due to online modulation a verbal dual task should lead to interference over and above that of non-verbal dual tasks.

Earlier studies found that the category effects were indeed subject to verbal interference in this way. For example, Roberson and Davidoff (2000) found that the category effect that obtained in a two-alternative forced-choice recognition experiment[14] could be eliminated if participants were asked to recognise the colour swatch they had just seen after reading aloud for five to ten seconds (cf. Pilling & Davies 2004). This result only shows verbal interference in the case of colour memory. Winawer et al. (2007) also found that their observed category effect was eliminated by a verbal-interference dual task (silent rehearsal of digit strings), but not by a spatial-interference dual task (memorising a spatial pattern). The elimination of the category effect is thus not due to some general cognitive load limitation. However, verbal interference does not eliminate all category effects (Pilling, Wiggett, Özgen, & Davies 2003), a finding that suggests that category effects also play out on longer timescales.

The linguistic origins of the category effects are further corroborated by findings showing that they are strongest in the right visual hemifield, which projects contralaterally to the putative language-dominant left hemisphere

14 This means that participants were shown a colour swatch, and after a delay of five or ten seconds were shown a target and distractor square, and had to indicate which of these was the same as the initial colour swatch.

(Drivonikou et al. 2007; Franklin, Drivonikou, Bevis et al. 2008; Franklin, Drivonikou, Clifford et al. 2008; Gilbert, Regier, Kay, & Ivry 2006; Regier & Kay 2009; but see Suegami, Aminihajibashi, & Laeng 2014).

3.2.3 Categorising and anomic aphasia

Further evidence of the transient effect of language on perception comes from studying individuals with *anomic aphasia*, that is, who have difficulty naming, among other things, colours. Here the influence of linguistic categories on putatively non-linguistic tasks is shown to be more dramatic: without an active command of colour words the task of colour categorisation is impossible (Davidoff & Luzzatti 2005; Dummett 1975). Davidoff and Roberson (2004) describe the problems that patient LEW has with sorting colour chips into categories. Instead of relying on categorical perception, he had to rely on discrimination:

> if colour samples were presented for which the within-group similarity was much greater than the between-group similarity (i.e., narrow ranges of reds, greens, yellows, and blues), he sorted them into four groups without error. Nevertheless, his performance was abnormal because he used a slow pairwise comparison for each stimulus; the colour groups did not 'pop-out'. His abject failure was for tasks where within-group colours had a wide range of lightness and saturation; in those situations, assessing visual similarity is extremely difficult. (Davidoff & Roberson 2004, p. 139).

In the first task, the greater within-group similarity allowed LEW to arrive at the grouping using only the ability to perceptually discriminate between two colour samples when they were present at the same time. When the within group similarity was lower, LEW was presented with a problem that is unsolvable through discrimination abilities alone. This closely resembles the *sorites* paradox (Davidoff & Roberson 2004): we can make a row of colour chips ranging from red to blue, and make sure that the difference in hue

between all the chips is equal. By relying only on colour discrimination, there is no point discernible where the category boundary should be. And indeed, when LEW was confronted with such a task, he drew the conclusion that all colour chips belonged to the same colour category. He arrived at this conclusion by looking for colour chips that appeared identical to him, and then moving them to a group, and then repeating this procedure until all the colour chips were in the same group.

3.2.4 Categorical perception beyond colour

Categorical colour perception is usually taken as a paradigm for studying the effects that verbal categories have on putatively non-verbal cognitive processes. However, if the category effects were only obtained in categorical colour perception, they would at best provide a cognitive curiosity. Fortunately, there is a growing literature which aims to extend the findings from the research on categorical colour perception to categorical perception in general (see Lupyan 2012 for an overview). Gilbert, Regier, Kay, and Ivry (2008), for example, found that the the lateralisation of the category effect also occurs for the perception of animal figures (cats and dogs). This asymmetry between hemifields was, like the category effect in colour perception, subject to interference in the case of verbal dual task, but not a spatial dual task. A second example is the category effect that Boutonnet, Dering, Viñas-Guasch, and Thierry (2013) found using electrophysiological measures for the categorical distinction English speakers draw between cups and mugs, a distinction that Spanish speakers do not make. A final example is an experiment performed by Lupyan and Spivey (2010), where participants had to attend selectively to four fives presented together with four 2s while fixating their gaze on the cross (see Figure 4). As soon as a small dot appears next to one of the 5s, the subjects have to press a button. The subjects were found to perform faster on trials where they heard a recorded

voice saying ‘attend to the five’ as compared to trials where they heard a recorded voice saying ‘attend to the category’. As the task of attending to the 5s remained the same for a 45-minute period of time, hearing the word ‘five’ was completely redundant. This result seems to suggest that hearing the word ‘five’ directed the participants’ attention to the 5s.

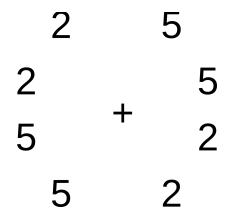

Figure 5. Display used by Lupyan and Spivey (2010, based on their figure 1).

3.3 Attentional actions as constraints on phenotypic reorganisation

In this section, I combine the account of utterances of concrete words as attentional actions based on the constraint view with the empirical results discussed in the previous section. In line with the *don’t-need strategy* as discussed in §3.1.2, I provide an account that can potentially explain the effects suggested by the empirical research without invoking representations.

The empirical results suggest that there are indeed category effects of verbal categories on colour perception. At the same time, the fragility of these category effects, as manifested in their being subject to verbal interference, suggests that these category effects are the result of a complex interplay of different timescales. The empirical evidence also shows that the category effects based in verbal categorisation are found in the absence of overt language use. A potential way of understanding the mechanism behind this is the ecological notion of a *task-specific device* (Bingham 1988; Runeson 1977; Wilson &

Golonka 2013). The guiding idea is that for the purposes of a particular task, an organism self-organises into a task-specific device which is assembled of resources distributed over body, brain, and environment.[15]

One way of understanding the assembly of task-specific devices is by means of *phenotypic reorganisation*. On this view, 'people turn themselves into walkers, throwers, graspers, and so on, and in so doing, they perceive the world in relation to what they have become.' (Proffitt & Linkenauger 2013, p. 172). As can be seen in this quote, Proffitt and Linkenauger focus on basic sensorimotor interactions with the environment. Their claim is that visual ecological information for affordances[16] is scaled based on morphology, physiology, and target-directed action. For example, whether something is graspable for a particular person depends, amongst other variables, on the size of their hand. Therefore, when a person is engaged in *grasping* something, the information for *graspable* is scaled based on hand size, among other things. In this way, the purposes of the organism, in the sense of the activities the organism is currently engaged in, 'mandate a goal-directed phenotypic reorganization' (Ibid., p. 180). Note that the ongoing phenotypic reorganisation is envisaged as a dynamic process of self-organisation.

For our present concerns, we can extend the notion of becoming a task-specific device by means of phenotypic reorganisation to categorical colour perception by taking the social context into account. Information then is scaled

15 For a discussion of the role the brain may play in this process of self-organisation and how neuroscience might develop methods to study this, see Anderson (2014, pp. 272–280) and Van Elk, Slors, and Bekkering (2010). See Bruineberg, Kiverstein, and Rietveld (2016) for an ecological-enactive interpretation of the free energy principle that similarly describes the self-organisation of the brain-organism-environment system.

16 Ecological information is information without content, that is non-semantic information, that enables an organism to pick up on affordances. The basic idea is that structures in ambient energy arrays are informative to an organism because they specify certain aspects of the environment. In other words, ecological information is information *for* affordances, not information *about* the world. Ecological information is not transmitted, nor is it processed or stored by a cognitive system (J.J. Gibson 1979). For a discussion see Van Dijk, Withagen, and Bongers (2015).

based not on characteristics of the individual, such as hand size, or actions performed individually, but rather based on community level patterns of categorisation as inculcated through the education of attention and the (attentional) actions of others. As we noticed earlier, the normative structuring of the education of attention entails that the child need not already be able to perceive categorically in order to be taught to do so.[17] What task a child is engaged in is determined initially by means of the normative and physical structuring of the learning situation by someone who has already mastered the relevant practice. If a caregiver introduces the child to colour words, she educates the child's attention by structuring the learning situation. For example, by drawing the child's attention to coloured objects, by showing the sorting of coloured objects, and so on. The caregiver also provides feedback on the child's behaviour. In this process, the child attunes to the culturally determined colour categories. The use of colour words as attentional actions is crucial here, for they provide the caregiver with the means to 'calibrate' the child. Hearing the word 'red' thereby comes to acts as a constraint on ongoing phenotypic reorganisation, making the child into a better red detector.

J.J. Gibson (1966, p. 52) defined the education of attention as 'a greater noticing of the critical differences with less noticing of irrelevancies' and 'a progressive focussing or centering of the perceptual system'. The current account suggests that this progressive focussing needs to be understood in a dynamic way: as a person interacts with his or her environment, there is a continuous phenotypic self-reorganisation into task specific devices. This phenotypic reorganisation is constrained by the current situation, which in the human case often includes attentional actions. But on this account, attentional actions are not the only constraints that shape phenotypic reorganisation in culturally

17 This of course does not entail that the child requires no abilities whatsoever. For example, learning colour words requires that the child is able to discriminate between different coloured objects. But colour discrimination does not necessarily rely on categorical colour perception, as the example of LEW discussed in §2.3 shows.

specific ways. The example of categorical colour perception suggests that attending to something that has a particular colour can act as a constraint on the ongoing phenotypic reorganisation. Attending to something red can have a similar constraining effect as hearing the colour word 'red', where both lead to phenotypic reorganisation as a red detector. In other words, and in line with the loopy nature of cognition, the ongoing phenotypic reorganisation is perceptually guided. On the level of neural dynamics, the effect of verbal interference can be explained by adopting a neural reuse perspective. Anderson (2016, p. 1) defines neural reuse as a form of neuroplasticity that entails that a 'diverse behavioral repertoire is achieved by means of the creation of multiple, nested, and overlapping neural coalitions, in which each neural element is a member of multiple different coalitions and cooperates with a different set of partners at different times.' A verbal interference task might rely on neural resources that would otherwise be used in the phenotypic reorganisation as a categorical colour perceiver. As Lupyan (2012, p. 4) remarks, once we have acquired categorical perception, much of our experience might best be described as 'hybrid visuo-linguistic experience.'

The current account has no direct implications for phenomenal experience. One possibility is given by Anderson (2014, p. 271), who states that 'Language probably does not make the world look different, but it can make some things easier to see'. Another possibility is that language does make the world look different, and thereby makes some things easier to see. For the purposes of this paper I will not explore the different ways of drawing the phenomenological implications of the current account.

In line with the idea that organisms are always selectively open to a field of affordances (Rietveld & Kiverstein 2014), any organism self-organises into many different task specific devices at any given time. In line with the neural reuse perspective, we should be careful not to conceive of these task specific

devices as neural modules, but rather as dynamically self-organising assemblies of resources distributed over body, brain, and environment. For example, a person is never only a grasper, although the anticipation of the act of grasping will lead to a shift in attention, which can be understood as a phenotypic reorganisation that enables an enhanced selective openness to those aspects of the environment that are relevant for grasping. The phenotypic reorganisation into task-specific devices is thus not an all or nothing affair. It is not the case that hearing the word 'red' either results in the phenotypic reorganisation into a red-detector, or fails to do so. Instead, we can think of attentional actions as transiently modulating phenotypic reorganisation. This modulation can then be up-regulated, for example by means of attentional actions, or down-regulated, for example by means of a verbal interference task (cf. Lupyan 2012).

This account of task specific devices, when combined with neural reuse, shows how perception can be constrained by processes at multiple time scales. The resulting picture is the following: by means of the education of attention, individuals learn to perceive categorically as expressed in structures in the dynamic unfolding of phenotypic reorganisation. This phenotypic reorganisation is perceptually guided and constrained by the actions of oneself as well of those of others. One example of these constraining actions are what I have called attentional actions. Although I have taken colour as a case study, the results for other forms of categorical perception discussed in the previous section suggests that this account might be applicable to categorical perception at large. For example, hearing the word 'chair' makes the person into a better 'chair detector' (Lupyan & Swingley 2012).

3.4 Language beyond attentional actions

So far, I have argued that utterances of concrete words are best taken to be attentional actions (§3.1). I then discussed research on the effect of community colour categories on individual colour cognition (§3.2), and proposed an ecological mechanism based on the idea that attentional actions function as constraints on the dynamic unfolding of phenotypic reorganisation (§3.3). In this final section, I situate the account of utterances of concrete words as attentional actions in a larger theoretical perspective on language by means of three remarks on (1) the context-dependency and reflexivity of language, (2) an extension of the current account to grammatically structured language, and (3) written language.

3.4.1 Context-dependency and reflexivity

It is important to stress that the current account highlights the context-dependency of linguistic actions. Attentional actions are not *stored* and *retrieved*, but *produced* to constrain processes that are unfolding *now*. In other words, we do not *use* language, understood as the tokening of culturally determined types, we *make* language in line with a history of learning (Harris 1980). Although it is certainly the case that we can use words in many *different* contexts, this only shows our ability to recontextualise. And this ability to *re*contextualise does not entail a decontextualised meaning. To be able to walk on many different surfaces does not entail an ability to walk *in abstracto*; similarly, being able to use a word in different or novel contexts does not entail an ability to use that word in abstracto, that is, it does not require knowledge of a decontextualised 'meaning'. This point is captured by Voloshinov (1930/1973, p. 68) when he says that 'the task of understanding does not basically amount to recognizing the form used, but rather to understanding it in a particular, concrete context'

(cf. Davidson 1986). Recognising the context-dependent nature of linguistic actions also does not negate the fact that we have normative practices that aim to regularise and standardise linguistic actions across contexts. As we have seen, the normative structuring of the initiate learning situation by the caregiver is critical for the child's education of attention. Other paradigmatic examples include extensive institutionalised language education and the writing of dictionaries and grammar books and the prescriptive use made out of these codifications. These latter examples are an expression of the reflexivity of language, that is, our ability to talk about talking (Harris 1998; Taylor 1992; 2000; 2010; 2013). I do not object to taking this reflexivity and the normative enforcing of linguistic behaviour it affords seriously. What I do object to is to take 'products of these [reflexive] processes as *realia*, and to retroject them on to languagers as the basis for their languaging activities' (Love 2017, p. 113). To understand this, take for example the practice of promise making. This is a reflexive practice because it consists in taking some linguistic act to be a promise. This is to say, if we were unable to say of some act that it was a promise, and we were unable to determine whether a promise was kept or broken, etc., we could not have the practice of promise making. At the same time, making promises does not have to be grounded in the existence of mental promises: the normative import of making a promise is not guaranteed by a mental state, but instead, by actually making the promise. Similarly, our practices that stabilise and normatively enforce the meanings of words do not have to be grounded in mental meanings. To reify promises or meanings is a prime example of the fallacy of misplaced concreteness (Whitehead 1929). In line with this focus on context-dependency, the constraining effect of attentional actions will also be context dependent. So, for example, the constraining effects of being asked the reflexive question *how do you spell the word 'green'* will be very different from a person telling you to *look at that green one*. For the

purposes of this paper, I have only explored situations which are like the latter, in the sense that they involve what we might call, in line with the notion of indication in the definition of attentional actions, the *indicative use* of a concrete word. The question whether the reflexivity of language can be explained in terms of constraints I leave for a future occasion.

3.4.2 Grammatical structure and attentional actions

In my discussion on attentional actions, I have mainly considered a child's single-word utterances of concrete words. This leaves open the question how we should make sense of concrete words as they are used in grammatically structured utterances. A tentative proposal is to see such utterances as enabling more fine-tuned attentional actions. As we have seen, attentional actions are performed by someone to indicate an aspect of the situation to someone else in order to achieve something. A child's attentional action *ball* is used to indicate a ball in order, for instance, to request a ball that is already the focus of attention, to initiate a game that child and caregiver regularly play together, to inquire as to the whereabouts of a particular ball that the child lost, to mark the sudden unexpected appearance of a ball, and so forth. In this single word attentional action, only the object of indication is thus explicated.

The tentative proposal is that in structured attentional actions, concrete words still serve this indicative role. This means that the discussion on attentional actions as constraints on phenotypic reorganisation also applies to structured attentional actions. For example, an utterance of the form *could you give me the ball* is akin to a child's single-word utterance 'ball' in the sense that it is an attentional action by means of which a ball is indicated in order to achieve something (in this case the other person handing me the ball). However, whereas only the structured environment reduces potential ambiguity in the case of the child's single word attentional action, this potential ambiguity is reduced

by linguistic means in the case of the structured attentional action. If I'm asking another person *Could you give me the ball*, this is more constraining than if I simply say *ball*; but both are attentional actions that indicate the ball in order to achieve something.

In terms of constraints, a structured attentional action is thus more constraining than a single-word attentional action, but the same basic structure is preserved when the child makes the transition to structured attentional actions. That is, structured attentional actions still function akin to ostensive gestures.[18] Of course, not all structured linguistic actions are attentional actions. However, providing criteria for delineating which structured linguistic actions are attentional actions and which are not falls outside the scope of this paper.

One might object that an extension to structured attentional actions fails the criterion of providing a don't-need account, as linguistic structure requires mental representation. And indeed, the received view of the epistemology of language *qua* linguistic structure is that it requires an explicit representation of the rules that together can be said to constitute the grammar of a language.[19] However, non-representational alternatives have also been proposed. Christiansen and Chater (2015, p. 17), for example, suggest that linguistic structure can be understood in terms of a history of processing operations that constrain current processing. They use a metaphor also used by enactivists of 'laying

18 Kukla (2017) goes as far as saying that speech acts that make highly theoretical claims still have an ostensive dimension. She bases her account on Heidegger and Haugeland's interpretation (2013, p. 67), who claims that in 'making an assertion a speaker lets what is being talked about show itself from itself, by pointing it out—putting it on exhibit, so to speak. If, for instance, I discreetly mention that your shoelace is untied, I draw the shoelace to your attention so that you can see, "from" the shoelace itself, that it is untied. By pointing out the untied shoelace (something I could also do without words), I let it be seen— let it show itself from itself.'

19 Matthews (2003, pp. 188–189) defines the 'Received View' of linguistic knowledge as follows: 'knowing a language is a matter of knowing the system of rules and principles that is the grammar for that language. To have such knowledge is to have an explicit internal representation of these rules and principles.' Devitt (2006), similarly, claims that 'the received view' is that 'language processing involves metalinguistic representations of the syntactic and semantic properties of linguistic expressions'.

down a path in walking' (Varela, Thompson, Rosch 1991). According to this metaphor, we conceive of the unfolding of current processes as being constrained by the traces left by earlier processes. An example is a desire path: when a person walks on the grass, she leaves a visible trail that can guide the next person's walking behaviour. By many people thus walking in each other's footsteps, the path becomes more visible and thus constrains subsequent walking behaviour more strongly. Based on this history of processing account of linguistic structure, we can conceive of linguistic structure as providing further constraints. The resulting image is thus that when a child learns to employ structured utterances, her attentional actions become more strongly constraining. I realise that what I have said here with respect to linguistic structure is very brief, but it is merely intended to provide a direction in which the current account can be extended.

3.4.3 Written language

The third remark situating the account of attentional actions concerns the extension to written language. First note that written language contributes considerably to the normative reflexive enforcing of linguistic behaviour. For example, without written language we could not write dictionaries, and we would thus be unable to 'look up' the meaning of a particular word (Ong 1982). Moreover, for us, hyper-literate human beings, our experience of spoken language is shaped profoundly by our facility with written language (Harvey 2015). On the code view, a spoken utterance can encode *the same* mental meaning as a written sentence. The constraint view, however, highlights the fact that the production of spatio-temporal patterns in spoken or sign language and the production of spatial patterns in written language belong to different ontological categories, are used to do different things, and constitute different cognitive domains with different cognitive dynamics (Kravchenko 2007; 2009;

Linell 2005; Love 2007). Perhaps the most conspicuous difference is that for writing and reading the principle of co-temporality, as discussed in §1.5, does not hold. Understanding written concrete words from a constraint-view thus requires an additional account, that is outside the scope of the current paper.

3.5 Conclusion

I argued that on an ecological-enactive approach, utterances of concrete words are best seen as attentional actions that are used to indicate situations, events, objects, or characteristics thereof in order to coordinate behaviour. The results from empirical research show that the attention-directing effects of socially constituted categories affect cognitive processes in the absence of overt language use. This can be understood by relying on the idea that humans continuously self-organise into task specifics devices, a form of phenotypic reorganisation. The unfolding of this process is constrained by a history of learning, the actions of oneself, others, as well as aspects of the environment. In this way this paper proposed a novel explanation for the empirical phenomenon of category effects. Finally, I situated the account of concrete words as attentional action in a larger theoretical perspective on language. Although a small step, the account of spoken concrete words as attentional actions shows how the ecological-enactive approach to cognition can be extended to explain linguistic behaviour.

4 On the way to solving the hard problem of content

Content is a central term in the philosophy of cognition and language. In general terms, some state or act has content iff it takes some thing to be a certain way where that thing need not be that way. To have content means to be correct under some conditions, and incorrect under other conditions. This means that a contentful state or act is *semantically evaluable* as being true, accurate, or satisfied: it has truth, accuracy, or satisfaction conditions (Hutto & Myin 2018; Neander 2017; Rescorla 2016). Uncontroversial examples of contentful acts are speech acts such as assertions (truth conditions) and commands (satisfaction conditions).

Besides these uncontroversial examples, there is considerable disagreement as to the scope of the concept of content. The representational theory of mind takes content to be the mark of the cognitive: it explains *all* cognition in terms of contentful mental representations (e.g. Fodor 2008; Adams & Aizawa 2008). On this view, the content of speech acts is derived from nonderivative mental content (Haugeland 1990). While in the 1980s and 1990s there was a sweeping optimism that mental content could be naturalised by reducing it to some natural phenomenon, there is currently widespread agreement that the reductionist hopes have not materialised, and the endeavour of naturalising content 'bears all the hallmarks of a degenerating research program' (Kriegel 2011, p. 3; cf. Hutto & Myin 2018; but see Shea 2013 and Miłkowski 2015 for an opposing assessment).

The cognitive sciences are currently making a *pragmatic turn* (Engel et al. 2013), in which content plays a very different theoretical role. In this paper I

start from the *ecological-enactive approach,*[1] which explains basic human behaviour in terms of dynamically unfolding skilful interactions with the sociomaterial world *without* invoking content. Cognition is defined as perceptually guided action, which means it is skill-based, ecologically and socially situated, and embodied. On this approach, content only comes into play in content-involving practices,[2] such as claim-making practices (Hutto & Myin 2017, p. 145). Given this ecological-enactive explanatory reversal, content is not used to explain cognition, but, rather, cognition is used to explain the origins of content. This means that the ecological-enactive approach does not require an account of naturalised content to bootstrap explanations of cognition. Instead, the ecological-enactive approach proposes to account for the natural origins of content, i.e., to explain 'how it is possible that content could arise in the natural world.' (Hutto & Satne 2015, pp. 529–530). Accounting for the origins of content consists in answering two genetic questions: [a] how did content-involving practices emerge, and [b] how do children become participants in these practices? It is vital that these questions are answered, for without these answers the ecological-enactive approach cannot account for quintessentially human forms of cognition.

This paper contributes to an ecological-enactive account of the origins of content from an ontogenetic perspective; it aims to provide a partial answer to [b]. I start from the perspective of the child and ask what skills a child has to

1 The ecological-enactive approach (Heras-Escribano 2019; Rietveld, Denys, & Van Westen 2018) combines insights from *Ecological Psychology* (J.J. Gibson 1979) and *Enactivism* (Varela, Thompson, & Rosch 1991; Noë 2012; Hutto & Myin 2013).

2 Following Hanna (2006, p. 275), I define a recurrent activity as a practice iff (1) there are public criteria or rules involved to which participants' behaviour is accountable, (2) there is a point to the activity, (3) and the participants understand the point of the activity and put the activity to use in furthering their own purposes. Not all practices are content-involving. For example, in some communities it is a practice to greet another person by means of a handshake. A person could deviate from this practice, for example by inadvertently overlooking an outstretched hand, or by raising their hand in preparation of a high five. While this act could be called 'incorrect', we could also call it 'inappropriate'. Crucially, it does not consist in taking something to be a certain way such that it need not be that way: it is not semantically evaluable. Therefore, it has no content.

learn in order to develop a sensitivity to the correctness conditions of the speech acts of herself and of others, such that she can become a participant in content-involving practices. In other words, I'm interested in the ontogenetic origins of *content-sensitivity*.

I shall not try to answer [a] in this paper – although I will briefly discuss what the implications of the current paper are for answering this question in §4.5. As Hutto & Satne (2015) point out, the key is to start from an approach that provides the theoretical resources to describe contentless social interaction. The ecological-enactive approach provides these resources. For the purposes of this paper, I shall not defend the ecological-enactive approach, but rather build on it.

The account is developed as follows: In order to become a competent participant in her community's practices, a child has to attune to her community's normative similarity judgements as they ground content-involving practices. In Section 4.1, I show how a representational approach solves the problem of normative similarity. Then I introduce Williams' (2010a; 2010b) non-representational calibration account. In Williams' account, a novice is 'calibrated' by a master to 'measure' the world as others in her community already do. In Section 4.2, I show how Williams' calibration account can be cashed out in the ecological-enactive approach by relying on the ecological concept *education of attention*. Besides grounding the calibration account in a theory of cognition, the education of attention model of learning shows the limitations of the calibration metaphor. Section 4.3 is devoted to a thought experiment that aims to show that calibration by itself is insufficient for content. In Section 4.4, I argue that in order to develop a sensitivity to content, a child has to acquire reflexive linguistic skills that enable her to participate in the (re)negotiation of correctness conditions. Content-sensitivity is expressed as a reflexive attitude, which can be understood as a non-focal responsiveness to

correctness conditions: a person's linguistic behaviour displays content-sensitivity if they are poised to deploy their reflexive linguistic skills when the need arises. The paper ends by showing some directions for future research.

4.1 Normative similarity and the calibration account

Content-involving practices are grounded in a background of normative similarity judgements, judgements that concern sameness with respect to categorisation (Williams 2010b, p. 79). For example, for two people to understand each other when using the word *red*, their judgements with respect to which objects are red need to be roughly aligned: they have to agree that fire trucks and apples are similar in the sense that they are both red. These similarity judgements are normative because the community's patterns of similarity judgements set a standard of correctness. A central problem in understanding a child's initiation into her community's practices is thus how a child attunes to her community's normative similarity judgements. The expression of normative similarity judgements comes in many forms. For example, in order to behave appropriately at the dinner table, the child has to recognise that different situations are the same in that they are instances of having dinner, and adjust her behaviour, for example by remaining seated throughout the dinner. Other examples of the expression of normative similarity judgments are calling this colour ☞■ *black,* stopping for a red traffic light and only moving the bishop on a chess board diagonally.

For the purposes of this paper, I am interested in how a child learns to use her first words that refer to situations, objects, and events, which are *verbal* expressions of normative similarity judgements. In this Section, I first briefly describe how a representational approach deals with this question. The reason for doing this is that this approach is still dominant in thinking about how chil-

dren learn language. Getting clear on the assumptions that underlie this approach helps me in spelling out the ecological-enactive alternative. I then introduce Williams' calibration account, which, as it steers clear of the representational assumptions, is a good starting point for the ecological-enactive approach.

On a representational approach, solving the problem of the verbal expression of normative similarity judgements is relatively straightforward, for it assumes that a child is *already* capable of making similarity judgments: to have a mental representation of some thing is already to categorise that thing. The problem faced by the child is conceptualised as learning what a particular word refers to by associating the correct mental representation to the word. The guiding assumption is that successful communication is guaranteed by shared referential knowledge: it is assumed that in order for communication to be successful, people have to reach a consensus on reference (Maturana 1978). In representational terms, this consensus is achieved when two people associate the same mental representation with the same word. This is facilitated by means of *ostensive definition*, where an adult ostends something and utters the corresponding word. As is well known, ostensive gestures underdetermine reference: for example, the same ostensive gesture can be used to point to Lassie, a dog, a mammal, a dog time-slice, and so on. In order to circumvent these Quinean qualms, children are assumed to be able to arrive at the correct mental representation by means of innate constraints on hypothesis formation for word-meanings (e.g. Clark 2009; Markman 1981; see Van den Herik 2018, §1.1 [Chapter 3, p. 76] for an extensive discussion). On a representational approach, learning words that express normative similarity judgments is thus a mapping problem guided by innate constraints that guarantee that the mapping relations will be the same across individuals.

As the ecological-enactive approach cannot rely on mental representations, the problem of accounting for the verbal expression of normative similarity judgements must be approached in a different way. The child does not face a mapping problem from words to categories she can already distinguish. Rather, the child has to learn to make normative similarity judgements that will constitute the categories for her. Williams' *calibration account* (2010a) provides a solution to how children attune to normative similarity judgements without making representational assumptions.[3] Her account is therefore useful for thinking about the problem of normative similarity from within the ecological-enactive approach.

According to Williams (2010a), the problem of how children attune to normative similarity judgments needs to be understood in terms of the initiate learning situation in which the master-novice relation is in play. The guiding metaphor is that a master of the community's practices 'calibrates' a novice to 'measure' her world by means of normative similarity judgments. Instead of ostensive *definition*, the initiate learning situation relies on ostensive *training* of the novice. The idea is that the novice's behaviour is malleable such that a master, by encouraging or discouraging the novice's behaviour with respect to some object, situation, or event, can make sure the child's behaviour will be similar to the master's.[4]

According to Williams (2010a), *social triangulation* can explain how a master of the practice can calibrate the novice to make the correct normative similarity judgements. Social triangulation, as described by Davidson (1992), consists of situations in which a *child* and *caregiver* interact with something together. In this *referential triangle*, using Davidson's example of a table, there

3 Williams' account is based on the later Wittgenstein. As I want to avoid exegetical issues, I refrain from discussing Wittgenstein explicitly in this paper.

4 A child also learns from observing and imitating others. However, as ostensive training relies on the normative structuring of the initiate learning situation by the master, imitative learning *by itself* is not sufficient for attuning to normative similarity judgements.

are three similarity patterns at play (see figure 5): (1) the caregiver verbally responds similarly to tables in different situations, i.e., she uses the word *table*, (2) the child verbally responds similarly to tables in different situations, and (3) the caregiver finds the responses of the child similar to her own.[5] Because of this triadic structure, Davidson argues, the caregiver can determine what the child responds to: the object at the intersection of lines 1 and 2 in figure 5.

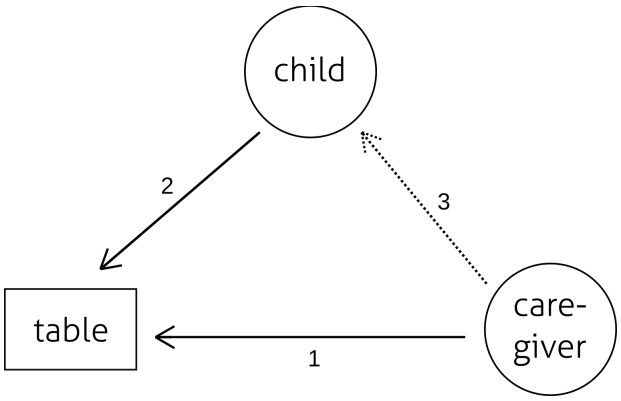

Figure 6: Social triangulation with three similarity patterns: (1) the caregiver verbally responds similarly to tables in different situations, i.e., she uses the word *table*, (2) the child verbally responds similarly to tables in different situations, and (3) the caregiver judges the responses of the child to be similar to her own. Note that social triangulation unfolds across situations. This time dimension is not represented in the figure.

Williams adds that there is an important asymmetry in initiate learning situation. For the novice, she says, 'there are no concepts, no grammatical background, no articulated environment' (Williams 2010b, p. 363). The ostensive training of the novice is possible only because the master normatively structures

5 Besides the similarity patterns, both participants have to be aware of the sharedness of the attention, which can be understood as a primary and embodied being-together. Note that this awareness of sharedness should not be understood as involving an explicit determination of the other's perspective (Fuchs 2012).

the initiate learning. This structuring provides the normative background against which 'the words and actions of the novice count, by courtesy, as reports [...] or distinguishing red from green' (Ibid.). The calibration account thus provides a very different answer to the Quinean question how we know what the child responds to with her verbal behaviour. The calibration account claims that there is nothing about the child that fixes the reference of her utterances. It is only because of the normative structuring that is provided by the master's attunement to her community's normative similarity judgements that the child's utterance of *table* can be taken, by the master, to be directed at tables, and not, say, four legged objects, undetached table parts, table timeslices, or other fanciful fictions. This normative structuring is evident in the behaviour of the master: she calibrates the novice based on whether the child's verbal expression of normative similarity judgements is in line with her own normative similarity judgements.

This asymmetry in the initiate learning situation has an important implication: the verbal behaviour of the novice is initially only meaningful against the normative background provided by the master (Williams 1994, p. 180). This conclusion holds not only for determining the referent of an utterance by a child, but extends to a determination of which speech act a child performs. Suppose for example that a child brings forth a sound reminiscent of /teɪb(ə)l/ in the vicinity of a table. There is nothing in the behaviour of the child to suggest that her intention is to say of this particular thing *that* it is a table – her holophrastic utterance is not an elliptical declarative sentence.

Here it is instructive to imagine what would happen if it were not a young child learning her first words, but rather a competent adult that uttered the word /teɪb(ə)l/ in a way that would make us question what the point of this utterance was. We would simply ask him what he meant, or provide a possible understanding of the utterance ('Would you call *that* a table?'). In the case of

the child, however, the complex language skills required to answer these kinds of questions are as of yet beyond her grasp (Williams 2010a, p. 201). The caregiver thus has no means to determine whether the child intends to assert that *this* is a table. In §4, I will defend the claim that these reflexive language skills, in which language is used to talk about language, are crucial for the child's development of content-sensitivity. For now, I want to show that there is no fact of the matter, independent of the background provided by the master, of what the novice intends to do (Williams 1999, p. 203). The produced sound *counts* as an utterance of the word *table*, that refers to the table, only by courtesy of the master. It is therefore only by courtesy of the caregiver that the child can do things such as talk about tables (Williams 2010b).[6]

Fortunately, at this stage the master need not worry about what the novice intends to do. As Dummett (1993, p. 157) remarks, 'the adults are not […] much interested whether his remarks have point or relevance, but only whether he makes them in the *right* circumstances, those that warrant them'. In order to do so, the master doesn't have to determine the intention of the child. She only has to calibrate the novice's verbal expression of normative similarity judgements to be in line with her own, which are already in line with the standard of the community. The master thus acts as a proxy in inculcating the community's normative similarity judgements in the novice. The master is capable of doing so because the community's normative similarity judgements have already been inculcated in her.

To sum up: where representational approaches individuate *what* the child responds to by means of the child's mental representations, the calibration account points to the normativity of practices in which the child is situated.

6 Sellars (1971/1975, p. 320) makes a similar point: 'We can imagine a child to learn a rudimentary language in terms of which he can perceive, draw inferences, and act. In doing so, he begins by uttering noises which *sound like* words and sentences and ends by uttering noises which *are* words and sentences. We might use quoted words to describe what he is doing at both stages, but in the earlier stage we are classifying his utterances as *sounds* and only by courtesy and anticipation as *words*.'

These practices enter into the initiate learning situation through the master and the way she normatively structures the initiate learning situation.

4.2 Calibration as the education of attention

Although the calibration account describes the structure of the situation that enables calibration, it remains silent on how it works. The ecological-enactive approach can elucidate the cognitive mechanisms underlying calibration with the concept *education of attention* (E.J. Gibson 1969; J.J. Gibson 1979; Adolph & Kretch 2015). Besides grounding the calibration account in a theory of cognition, this combination of ideas enables refining the calibration account, in particular by showing the limitations of the calibration metaphor. In this section, I first introduce the ecological-enactive approach and the education of attention, after which I draw the implications for the calibration account.

The ecological-enactive approach starts from the idea that we perceive possibilities for action in our environment, called *affordances*. Acting on affordances is always perceptually guided: if you do not perceive anything, you cannot do anything. Action thus presupposes perception. Similarly, perception presupposes action: we do not wait passively for affordances to appear, we actively explore our environments. We squint our eyes, run our hand across a surface, turn our heads, manipulate objects, and so on. Perception and action are thus inextricably intertwined. The internal states of organisms, including brain dynamics, are states of action-readiness, states that are in-between overt action and ability (Frijda 2007; Bruineberg, Kiverstein, & Rietveld 2016). Besides triggering a readiness to act, perceiving an affordance can motivate to act on that affordance (Withagen, de Poel, Araújo, & Pepping 2012). Whether a perceived affordance motivates acting on it depends on which activity one is currently engaged in. If I give a lecture, I can see a bottle of water affording

drinking from without being motivated to drink from it; perceiving this affordance will become motivating when I feel I have a dry throat. We always perceive multiple affordances at the same time, that is, we are *selectively open to a field of affordances* (Rietveld & Kiverstein 2014). Attention can thus be defined as selective openness to the field of affordances in relation to the activities one is currently engaged in (Van den Herik 2018, p. 95 [Chapter 3, p. 82]; E.J. Gibson & Rader 1979).

Learning to act on a particular affordance requires becoming selectively open to the aspects of the environment that are relevant for that affordance by actively exploring them. This perceptual learning is conceptualised as an education of the attention, 'a greater noticing of the critical differences with less noticing of irrelevancies' and 'a progressive focussing or centering of the perceptual system' (J.J. Gibson 1966, p. 52). Where a representational approach assumes that a cognitive system must learn to enrich sensory input to infer representations, the ecological-enactive approach sees learning as a process of selection and differentiation, that is, as discovering distinctiveness and invariance *in* the environment (E.J. Gibson 2000). The development of normative similarity judgements and their verbal expression can be understood as an education of attention with respect to the categories constituted by normative similarity judgements.

It is important to realise that on the ecological-enactive approach, word-learning is not acquiring referential knowledge. The child's verbal behaviour is described in terms of *attentional actions,* actions aimed at directing the attention of others (Van den Herik 2018 [Chapter 3]). The guiding idea is that the child learns to use repeatable vocalisations to *indicate* things in her environment to others. The child's use of language is thus indicational. Language is in first instance mode of action aimed at manipulating the attention of others in order to engage in joint activities. This means that the child does not acquire knowledge

about language, she comes to know-how to do things with words.[7] The child's utterances, understood as attentional actions, function as ostensive gestures, they are 'a technique for pointing' (Baggs 2015, p. 260). They can fail in the sense that the coordination of behaviour breaks down, but this does not show that the ostensive gesture or attentional action was *incorrect*.

One important insight from studying verbal development from an ecological-enactive perspective follows from the rather banal observation that children's attention is educated in everyday situations in which they are doing things together with a caregiver (Reed 1995; Zukow-Goldring 2012). The child is not a passive observer, or spectator, but learns to speak by 'doing things with words' (Rączaszek-Leonardi 2009, p. 170; see also Bruner 1990, p. 70). Calibration is thus not something the child passively undergoes. It is rather expressed in her burgeoning ability to engage in practices. Ingold (2001) therefore refers to the education of attention as a process of *guided rediscovery*. This means that the caregiver structures the environment of the child in such a way that she can find out, for herself, what the caregiver already knows.

A second insight is that, unlike the readings of a lifeless measurement device such as a thermometer, normative similarity judgements are situated and dependent on the activities one is currently engaged in. For example, something that is not a table can count as a table in particular contexts, such as when a child uses a small stool as a table for his dolls. The normativity in normative similarity judgements is thus always a *situated normativity* (Klaassen et al. 2010; Rietveld & Kiverstein 2014).

Let us look at a real life situation to see the education of attention at work. Consider the following example, given by Bottineau (2010, p. 288), in which a mother introduces her two-year-old daughter to snow:

7 The ecological-enactive approach accounts for linguistic behaviour without invoking mental representations of linguistic rules and principles, but instead sees linguistic competence as embodied know-how. See Van den Herik (2017) [Chapter 2], Lance (2017), Simpson (2010), and Hanna (2006) for accounts of linguistic knowledge along these lines.

> Look! This is snow. Look! It's white, it's cold, it sticks, you can pick some and make a ball, and throw it at Daddy, look, splash! And they all burst out laughing.

As Bottineau (2010, p. 289) describes, before the mother initiates this interaction, the child 'is not supposed to have even begun to detect the snow'. The mother *shows* the snow to the child, thereby *making it present* for her. Ostensive training has a transformative potential: in some cases, 'through the ostension, he not only comes to see what he did not before, but he comes to be able to see what he could not before' (Kukla and Lance 2014, p. 40). The mother sets up a referential triangle with her daughter. Then, the mother shows the snow in its different sensorimotor and cultural dimensions. In this way the mother *educates the attention* of the child by setting up conditions in which the child can come to *re*cognise snow and *re*discover the affordances the snow offers within her community's practices, such as engaging in a snowball fight, and directing attention to the snow by saying *snow*. The child does so as an active participant in the social interaction. To say that the attention of the child is educated therefore means that the child *recognises* snow and is able to respond to it in ways that are appropriate in her community: in other words, she is calibrated. The role of the caregiver in this process of the education of attention is that of active guidance. The caregiver encourages and discourages the child's behaviour, based on what the caregiver's own normative similarity judgements tell her is the appropriate way to behave.

Recognising the indispensability of the active role of the child shows the limitations of the calibration metaphor. Calibrating a child is unlike calibrating a lifeless measuring device. For although it is certainly the case, as the calibration account shows, that the novice's verbal behaviour is in the first instance only meaningful against the normative background provided by the master, we should be careful not to describe the behaviour of the novice in terms of mind-

less stimulus and response. As a child comes to know-how to do things with words in a shared world, she is already participating in practices, even if she still lacks the skills required to be counted as a master of them. One of the defining characteristics of human learning is that we do things before we can do them, in an environment that is cognitively, normatively, and physically structured by a caregiver (Reed 1996, p. 149). Calibration relies *both* on the normative structuring of the initiate learning situation by the master and on the malleable behaviour of the novice. These two factors combined enable the child to gradually grow into her role as a competent participant in her community's practices.

4.3 Why calibration is not sufficient for content-sensitivity

While we should not underestimate the calibrated child, and describe her behaviour in terms of mindless stimulus and response, we should also not overestimate her behaviour. For although the calibrated child's speech acts *can* be treated by the caregiver *as if* they are already contentful, the fact that they are meaningful only against the normative background provided by the caregiver means the child's speech acts are not yet contentful. In this Section, I motivate the idea that calibration by itself is not sufficient for content-sensitivity.

In order to understand why calibration is not sufficient for content, a physical measurement device can serve as an analogy. Think for example of a calibrated mechanical scale, where a dial will show the weight of the object on the scale. The workings of such a calibrated scale, by itself, do not involve content: the scale makes no claims with respect to the weight of the objects placed above it. This *hard problem of content* is discussed at length by Hutto and Myin (2013, Ch. 4–6). The root problem they identify is the fact that a

correlation does not constitute content. Proper calibration only ensures that the position of the dial reliably correlates with the weight of the object on the scale. There is no doubt that this correlation is necessary for the scale's role in the content-involving practice of ascribing weights to things – in fact, the scale is designed so that this correlation holds. Also necessary, however, is the skill of a person in *using* the scale to perform measurements of weight, and on this basis, come to a contentful judgment. A similar argument can be made in the case of the calibrated child. The child's verbal expressions of her similarity judgements reliably correlate to the situations that warrant them – that is simply what it means to be calibrated. However, the child lacks the skills to put these 'measurements' to use in producing and understanding contentful speech acts.

At this point, one might argue that the child *does* put her 'measurements' to use. For did I not just discuss that the child starts out by doing things with words? Unlike the scale, the child uses her calibrated 'measurements', i.e., her verbal expressions of normative similarity judgements, in order to coordinate her behaviour with others. However, this by itself is not sufficient for content-sensitivity. A factor that complicates the matter at hand is that we have ways of talking about non-linguistic behaviour and ways of talking about contentful linguistic behaviour, but, in the words of Davidson (1999, p. 11), 'what we lack is a satisfactory vocabulary for describing the intermediate steps'. Although the ecological-enactive approach aims to provide a vocabulary for these intermediate steps, it comes very natural to us to describe the calibrated child's verbal behaviour in terms of contentful speech acts. If a child produces the sound /teɪb(ə)l/ in the vicinity of a table, we readily describe this as the child meaning 'this is a table'. In order to circumvent this complication, it is therefore instructive to perform a thought experiment in which there is calibration, and coordination of behaviour that is dependent on this calibration, *without* a background of content-involving practices.

Suppose we come across a social animal that builds constructions; beavers, for example, who build dams.[8] We hear a beaver produce a vocalisation in the presence of a second beaver, after which the second beaver leaves and returns with a branch which it drops in front of the first beaver, who uses it to continue building the dam. This is a recurrent phenomenon: every time a beaver produces this vocalisation, a second beaver searches for a branch in the forest and brings it back to the beaver that produced the sound. Suppose finally that we see older beavers calibrating their young: an older beaver produces the vocalisation, and then encourages the young beaver to accompany it to the forest, to gather a branch, leads it back, and so on.[9]

What should we make of this beaver behaviour? One way of interpreting it would be the following: [A] the first beaver notices that an additional branch is currently required to continue construction and aims to inform other beavers of this fact, who, given their pro-social nature act on learning about this. The vocalisation then would have the content 'a branch is needed', which is correct in conditions in which a branch actually is needed to continue construction and incorrect otherwise. A second way of interpreting the vocalisation would be the following: [B] the first beaver notices that an additional branch is required and commands the other beaver to bring him one. Here the content of the vocalisation would be roughly 'get another branch!', with the satisfaction condition being that the other beaver gets the required branch. Many other interpretations are possible, but for present purposes two will suffice.

[A] construes the vocalisation as a kind of assertion about the current status of the construction project, whereas [B] construes the vocalisation as a command that makes no reference to the current situation. The question is now:

8 My beaver thought experiment is based on Kukla and Lance's (2014, p. 36) super squirrel thought experiment. I have extended it and tailored it to the current discussion on content.

9 This example could be made more complex by supposing that the beavers have more vocalisations with other social effects, however, this would not change the conclusions reached about their behaviour.

is there a way of deciding which interpretation is the correct interpretation? Note that the interpretations have different consequences for evaluating situations. For instance, suppose that we observe a beaver producing the vocalisation and another beaver getting a branch even thought the construction work is finished. On [A], the beaver was incorrect in making the vocalisation: no additional branch was needed; the behaviour of the addressed beaver is irrelevant. However, on [B], no such semantic error has to be assumed. The first beaver simply issued a command in slightly different circumstances than it normally would, and the command was satisfied by the second beaver's behaviour.

I would propose that there is no answer to the question what the beavers are doing in terms of our practices of asserting and commanding. The reason for this proposal is that the criteria we have for distinguishing asserting from commanding do not apply to the beavers' behaviour. Just like the calibrated child, the beavers lack the reflexive communicative skills that enable them to differentiate between the different possible interpretations of their own vocalising behaviour. The differentiation *we* can make therefore has no import for the beavers. In fact, in the absence of reflexive communicative skills, no behaviour can provide evidence for one interpretation over the other. For instance, if we sometimes observe that the second beaver does not respond to the vocalisation by retrieving a branch, this can be described either as understanding the assertion-like vocalisation but not being motivated to act, or as disobeying the command-like vocalisation. In other words, there is no way of deciding which interpretation of the vocalisation is correct because the beavers' vocalising behaviour does not conform to the criteria we have for making the distinction. Therefore, the distinction does not apply: the beaver vocalisation is neither an assertion nor a command, although it does enable the beavers to coordinate their behaviour.

Even if the first problem could be solved, a second problem arises when we try to specify the content of the beaver vocalisation. Interpretation [A] was that the vocalisation meant *branch needed*. However, interpretation [A'] is also possible: *construction material needed*. The correctness conditions of [A] and [A'] are different. This is important if new situations occur, for example the introduction of plastic into the beaver habitat. If a beaver was to bring back a piece of plastic suitable for construction of the dam in response to the vocalisation, [A] would entail that the beavers erroneously took the piece of plastic to be a branch, whereas [A'] would not entail that an error is made, for the piece of plastic *is* construction material.

Moreover, in case the beavers' coordination of behaviour breaks down, they lack the means to negotiate correctness conditions of their behaviour. If a beaver uses the piece of plastic the second beaver brought in constructing the dam, this does not mean that the content of his vocalisation was *construction material needed* all along, for we could interpret the same situation by saying that the other beaver, incorrectly, takes this piece of plastic to be a *branch*. Since the beavers have no way of negotiating the conditions under which their vocalisation would be correct, they have no way of specifying the content of their vocalisation, which is to say, they have no way of *saying what they meant* when they produced the vocalisation. This means that, although the beavers' vocalisation can contribute to successful dam building, it is not the case that beavers can be correct or incorrect when producing or responding to the vocalisation in the way required for the constitution of content.

One possible counterargument can be made by introducing the idea of normal conditions. Here the argument would be that the content of the beaver's vocalisation is fixed by the 'normal' conditions in which the beavers used the vocalisation. Following Fodor (1984, p. 243), we can respond to this argument by saying that there is no way of specifying what the normal conditions are

without assuming a description of those conditions. For while we can specify extensionally what the beavers' vocalisations correlated to in a particular period of time, this does not give us a content for their vocalisation, as the extensionally specified correlatum can always be described by many different contents.

Fortunately, these considerations do not affect the beavers, for they need not care about getting things right in a way that requires content. They need only care about whether the dam gets built. In other words, as long as there is a regularity in the coordination of behaviour of the beavers, the beavers do not require a sensitivity to correctness conditions of their vocalisations. This does not negate the fact that *we* could attribute content to their vocalisations, just as *we* can say of the malfunctioning thermostat that it falsely claimed that the room was too cold when it turned on the heater in the middle of a hot summer's day. But this possibility of attribution does not show that content is involved.

The calibrated child is in a similar position to the beavers in our thought experiment. She has, through ostensive training, attuned to her community's normative similarity judgements and expresses these verbally, which enables her to coordinate her behaviour with others. Whether her verbal behaviour is appropriate is determined by the caregiver, who relies both on the normativity of sociocultural practices and on whether the situation warrants the response. The child, however, has no means to ascertain, or even entertain, the correctness of speech acts, be it of herself or of others, which is exactly the reason for conceptualising the initiate learning situation in terms of calibration.

4.4 The role of reflexivity in the development of content-sensitivity

In the last Section, I argued that a calibrated child, in order to develop content-sensitivity, still has to learn to put her calibrated 'measurements' to use. By

means of the beaver thought experiment, I argued that using verbal expressions of normative similarity judgements in order to coordinate behaviour with others is insufficient for constituting content. In this Section, I argue that a child has to acquire reflexive linguistic skills in order to put her 'measurements' of the world to use in ways that display content-sensitivity. I first introduce the notion of reflexivity, then I show how the ecological-enactive interpretation of the calibration account can explain the development of reflexive linguistic skills, and in doing so account for the development of content-sensitivity.

One of the defining, though often underappreciated, properties of linguistic communication is that it is *reflexive* (Taylor 2013; 2010; 2000): we do not just talk about the world, we also talk about talking, and we do so quite often. Take the following examples:

- Answer the question please!
- What do you mean by that?
- I did not mean to insult you.
- How do you say 'hello' in French?
- He spoke in a strange accent.
- The conversation was very engaging.
- I do not like talking to him.
- Did you understand what he said?
- He suggested we go that way.
- Most of what he said was a lie.

This reflexivity is expressed in metalinguistic activities,[10] which are 'a vast range of meaningful behaviors that typify the attributes of language, its users, and the activities accomplished through its use' (Agha 2007, p. 17). When metalinguistic activities form a pattern and consolidate, they become a metalinguistic practice. To typify something is not merely to describe it, but to suggest treating it in a particular way. Take the practice of making promises for instance.

10 In using the term 'metalinguistic' I do not wish to invoke a Tarskian picture of two languages, a metalanguage and an object language. Instead, I highlight the constant and fluent shifting between talking about other things and talking about language as distinctive for linguistic behaviour.

Starting an utterance with the words *I promise that*… typifies that linguistic act as a promise, and thus suggests treating that utterance as a promise. Of course, we can also typify a linguistic act in retrospect (*You promised me you would do that!*), or question whether a certain typification is adequate (*A: I'll try to do that. B: Is that a promise?*). I define reflexive linguistic skills as those skills that enable one to make these metalinguistic typifications.

Having reflexive linguistic skills is a criterion of being a competent participant in linguistic practices. As Noë (2012, p. 3) formulates it, 'to understand a word is to know how to use it, and that means, among other things, knowing how to explain its use to another, how to answer the question *what does this mean?*' (see also Hacker 2014, §5). If we were to come across a person who produces sounds that resemble English sentences, but who could not explain what he meant upon our request, we would not be able to ascertain whether he understood what he was saying. In other words, being able to speak requires being able to explain what you mean when you speak, a reflexive linguistic skill *par excellence*.

Taylor (2013) proposes that acquiring reflexive linguistic skills can be seen as a special case of Williams' calibration account (see figure 2). As verbal behaviour is just another observable event the child regularly encounters, the process of metalinguistic calibration can be similar to the initial calibration process.[11] From the perspective of the novice, being taught metalinguistic activities enables her to 'measure' the verbal behaviour of others and herself in metalinguistic terms: utterances being questions, excuses, compliments, promises, requests, commands, assertions, baptisms, &c. This in turn enables her to do such things as explaining what she means, repeating what she just said when asked, questioning a statement made by someone else, wondering whether someone is being serious or making a joke, giving a reason for an earlier utter-

11 Clark (2006) similarly highlights that utterances are objects we can be directed at.

ance, debating whether a previous response was warranted by the situation, &c. In other words, by being taught reflexive language use the child can start to use her community's metalinguistic typifications in order to coordinate her linguistic behaviour with others. This metalinguistic calibration, like the initial calibration, does not involve the child acquiring knowledge *about* language, or, as Taylor (2013, p. 317) puts it, she is not a little linguist. Her growing metalinguistic competence consists in coming to know-how to do things with metalinguistic words.

Getting to grips with reflexive language use marks a major breakthrough in the child's development. The reason for this is that it enables the child to relate to her own relation to the world, afforded by the fact that the child's own utterances show up for the child as objects to be 'measured' in metalinguistic terms. Crucial here is that the metalinguistic education of attention constitutes utterances as objects of attention (see figure 6). As the child learns to do things such as to question the correctness of people's utterances, she similarly learns to question the correctness of her own utterances. This enables the child to understand that her utterances can be mistaken with respect to the way the world is.

Reference to 'the way the world is' does not entail that the child must somehow have direct access to a determinate world. Instead, it means that the child comes to understand, in a practical way, that others can have a different take on the same thing, that she could change her mind based on further observations, that she might be persuaded to change her mind by talking to others, that things might look different under different conditions, &c. Like other forms of social behaviour, content-sensitive behaviour consists of the coordination of behaviour with respect to the world. But content-sensitivity means that in this social coordination we are 'beholden to how things actually are' (Kukla & Lance 2014) in the sense that we understand, in a practical way, that what we are inclined to say can be mistaken.

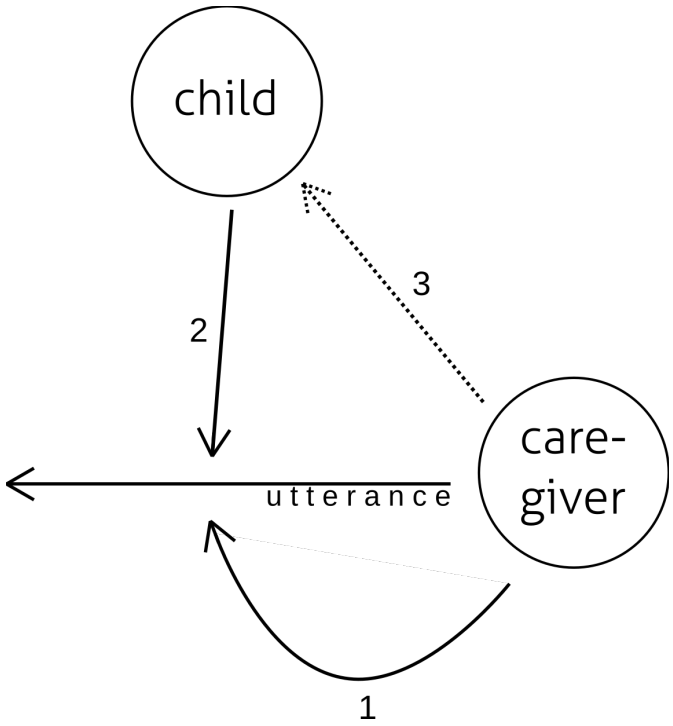

Figure 7: Metalinguistic education of attention constitutes an utterance act as an object of attention. Caregiver and child both metalinguistically respond similarly to an utterance, for example by calling it a promise (1+2), and the caregiver judges the responses of the child similar to her own (3).

Important to note here is that the child's nascent content-sensitivity is grounded in the child's initial calibration, but cannot be reduced to it. For judging someone's verbal expression of a normative similarity judgement to be incorrect in a referential triangle relies on being able to make the correct normative similarity judgment. For example, to be able to judge that someone incorrectly says of this shape ☞■ that it is a circle requires that one is inclined to call it a square.[12] The child experiences cognitive friction as her speech acts

12 This does not entail that content-sensitivity is a form of verificationism. Making a contentful

are corrected or commented upon by her caregiver, and it is this cognitive friction that enables the child to learn gradually to evaluate her own utterances in metalinguistic terms. This is impossible to do without reflexive linguistic skills, and this is exactly what separates the child from the beavers from our thought experiment. While her own normative similarity judgements provide the anchoring that enables the child to experience cognitive friction, it is by means of her reflexive linguistic skills that she can begin to entertain the fact that her utterances can be incorrect with respect to the way the world is. The cognitive friction thus results both from the fact that the verbal expression of normative similarity judgements between people do not always coincide, and from the reflexive resources available to the child to deal with these differences.

Content-sensitivity is not limited to explicitly using one's reflexive linguistic skills. Instead, content-sensitivity is marked by a *reflexive attitude*: an ongoing non-focal *responsiveness*[13] to the correctness conditions of (verbal) behaviour. Let us look at an example to understand the idea of non-focal responsiveness. Any person reading this text will have spent considerable time in their life becoming a competent speller. Much of this learning takes place by means of explicit (classroom) instruction, a paradigmatic metalinguistic activity. During this instruction, spelling is the focus of attention. And while being a competent speller surely includes an ability to participate in these explicit spelling activities, like spelling a word out loud when asked to do so, competence in spelling is obviously not limited to verbal behaviour marked by an explicit metalinguistic dimension. For being a competent speller is also expressed in one's writing being in accordance with the rules of spelling, as well as noticing spelling errors as one reads. The responsiveness to correct

assertion requires an understanding under what conditions that assertion would be correct, not the ability to verify if these conditions actually obtain.

13 I borrow this terminology from C. Taylor (2016, p. 7), who claims that language 'involves a sensitivity to the issue of rightness', where he understands sensitivity as a 'nonfocal responsiveness'.

spelling is thus non-focal in the sense that, once mastered, spelling usually is not the focus of attention even though it guides one's behaviour. Of course, spelling can become the centre of attention, for example when one is unsure about the spelling of a word while writing, or when one spots a spelling mistake.

Content-sensitivity similarly consists in a non-focal responsiveness. Although the child's initiation into content-involving practices is effected by means of explicit metalinguistic activities, content-sensitivity is expressed in the ability to monitor the unfolding of communicational activity for possible breakdowns in communication such as misunderstandings, disagreements, errors, &c, and being able to employ reflexive linguistic skills in order to negotiate these breakdowns. This ability consists in being poised to, and knowing-how to do things such as question, disagree with, argue against, justify, proof, and doubt contentful speech acts of others *when the need to do so arises.*

In other words, the metalinguistically calibrated child no longer merely relates to the sociomaterial world: she now relates to her own and others' relation to this world. In this way, metalinguistic calibration thus enables us to explain in non-mysterious terms how a child comes to understand that she and others have a perspective on the world: by means of the metalinguistic education of attention someone's perspective, as expressed in their speech acts, becomes an object of attention measured in metalinguistic terms. It is by acquiring this reflexive attitude that the child's perspective as well as the perspectives of others show up for the child. This reflexive attitude is thus a *metaperspective* that enables the articulation of first- and third-person perspectives (Fuchs 2012). This understanding is crucial for developing content-sensitivity, for it is only when one understands that a perspective on the world can be correct or incorrect that the question arises under what conditions that perspective, as expressed in verbal behaviour, is correct.

Developing content-sensitivity is a gradual affair. There is no point at which a child has mastered all reflexive linguistic skills relating to content-involving practices. Metalinguistic practices that enable the determination of content are a mixed bag, ranging from the everyday to the highly specialised, and we are initiated into these practices in a piecemeal fashion. Think for example of the highly specialised jargons of science and philosophy, and the particular metalinguistic ways in which philosophers determine the truth conditions of their assertions. Acquiring the skills required to participate in these metalinguistic practices enable a student to negotiate the correctness conditions of utterances in new ways, and they therefore contribute to her content-sensitivity. Here once more the normative structuring we discussed in introducing the calibration account comes into play. As the child is calibrated with respect to language, her developmentally earlier ways of 'verbal' responding, which counted as speech acts only by courtesy of the master, will come to rely less and less on the structuring provided by the master. However, it is impossible for any person to be wholly independent of other people, for the skills the child acquires as she develops content-sensitivity are *social* skills, geared towards participating in a practice with other people. The fact that the child's development is supported by the normative structuring of others means that the child's indicational use of language can *gradually* transform to full blown speech acts. That others treat her *as* a competent participant enables her to develop the necessary metalinguistic skills *before* she has mastered participation in those practices that require content-sensitivity. It is therefore only in virtue of treating the child *as if* she is already content-sensitive that the child can develop to actually become content-sensitive.[14]

14 It is interesting to note that the initiation into highly specialised practices can also be understood as an initiate learning situation with a master-novice structure and continues to rely on treating initiates as being proficient while they are not yet. In engaging a neophyte student's linguistic actions, for example, the philosopher also has to treat the student *as if* she has already mastered philosophical concepts. For if we treat the neophyte student as if she does not employ the philosophical concepts when she uses the words used to express these

Finally, it is important to realise that content-sensitivity is not a merely receptive capacity. The child is not sensitive to correctness conditions that exist independently of her own activity. By her initiation into metalinguistic practices the child becomes capable of judging whether other people's utterances are correct or incorrect, and this is what marks her transition from novice to master. By using her reflexive linguistic skills, the child can question the master's authority, the authority that enabled the master to calibrate the novice, and become an real participant in her community's practices.

4.5 Limitations and open questions

In this final Section, I discuss some of the limitations of the present account and directions for future research. First of all, for the purposes of this paper, I assumed that content-involving practices are already in place. However, fully accounting for the origins of content includes answering question [a] from the introduction, namely how did content-involving practices emerge in the first place? If the current account of the ontogenetic origins of content-sensitivity in terms of reflexive linguistic skills is viable, this shows that an account of the origins of content-involving practices is not in principle impossible. The reason is that there is nothing mysterious about reflexive linguistic skills. If we can explain animal communicative behaviour in terms of attentional actions, and we realise that communicative behaviour is just another observable aspect of the environment, we can conclude that it would be perfectly possible to develop reflexive communicative skills in a piecemeal fashion. This, of course, leaves open the question what particular reflexive linguistic skills are required for developing content-involving practices, and how the development of these practices this took place. One interesting suggestion here is given by Lance (2017),

concepts from the get go we would not have any reason to agree with or correct the student's assertions.

who argues that content can only arise given a certain context-independence of animal signalling behaviour. He argues that this context-independence can emerge if signalling behaviour has a function in different practical contexts. Reflexive communicative skills could prove useful in understanding how our ancestors might have coordinated their signalling behaviour across different practical contexts, for by these reflexive communicative skills the similarity between different contexts could be marked.

In this paper, I focussed primarily on spoken language. There is a good reason for this: children learn to talk before they learn to write. At the same time, written language plays an important role in the reflexivity of language (Love 1990; 2017). One reason for this is that whereas spoken or gestural language consists of spatiotemporal patterns, written language consists of spatial patterns (Harris 1998). This detemporalisation of language could play an important role in the emergence of content-involving practices. Although outside the scope of the current paper, further research could elaborate on the interrelations between written language and content-sensitivity.

A third direction for future research would be to determine the effects the development of content-sensitivity has on contentless forms of cognition. On the one hand, content-sensitivity could be an add-on to basic forms of cognition, another tool to be added to the cognitive toolkit, leaving contentless forms of cognition more or less intact. Radical Enactivists Hutto and Myin (2017) argue in favour of such a 'multi-storey story' where basic contentless forms of cognition continue to exist besides content-involving forms of cognition. On the other hand, the development of content-sensitivity can be taken to be more transformative. With respect to perception, for example, one could argue that most, if not everything that we perceive involves normative similarity judgements, and that this by itself is enough to describe perception as contentful. Another important limitation of the present account is that I have only discussed the content of

speech acts. This leaves open the question how children acquire the concept of belief, and how the ecological-enactive approach deals with such mental states with content.

Finally, discussing the origins of content also raises the question what content is. Traditional accounts of content take contents to be abstract objects, such as propositions. However, this reified notion of content does not sit well with the ecological-enactive approach. This is one of the reasons I have framed the question in terms of content-sensitivity, which concerns the skills required to participate in content-involving practices. One important consideration here is that the practice-based approach to content proposed in this paper need not conceive of content as being fully determinate. Instead, content is determined partially, provisionally, and defeasibly by the participants in a particular encounter and subject to constant (re)negotiation. Although answering this question would require a separate paper, a tentative suggestion would be to interpret content in a way reminiscent of adverbial theories of perception (cf. Crane & French 2017, §3.2). On such a view, we could say that a child that has developed content-sensitivity can speak in contentful ways, where this is cashed out in terms of her possibility to use her reflexive linguistic skills when the need to do so arises. The upshot of this would be that we can talk about content without having to introduce abstract objects.

4.6 Conclusion

Through the metalinguistic education of attention, the child comes to measure linguistic behaviour in metalinguistic terms, which enables her to relate to her own relation to the world. In this way we can explain in non-mysterious ways how a child comes to be sensitive to the correctness conditions of speech acts. By mastering the reflexive linguistic skills to distinguish between correct and

incorrect responses, the child develops a reflexive attitude, expressed in a non-focal awareness to correctness conditions. The development of this reflexive attitude allows her to determine when her metalinguistic abilities need to be brought into play. This enables the child to participate in the ongoing negotiation of correctness conditions in content-involving practices. Although the current paper doesn't give all the answers, it suggests that an ecological-enactive account of quintessentially human forms of cognition is not in principle impossible.

5 The constitution of linguistic normativity

Linguistic practices involve linguistic normativity: words can be used correctly or incorrectly, linguistic actions can be used to undertake commitments, we can criticise the use of an expression, and so on. Linguistic behaviour is therefore subject to normative assessment, which is possible because we have criteria for differentiating between correct and incorrect use of language, between appropriate and inappropriate speech act, between commitments kept and broken, and so on. In this paper, I argue that metalinguistic reflexivity is constitutive of linguistic normativity. By this I mean that the normativity of language requires metalinguistic practices, by means of which we attribute properties to language, language users, and the activities we can accomplish by using language. I discuss the regress objection, and in response to this objection I sketch a way to account for the ontogenetic and sociogenetic origins of linguistic normativity in light of this constitutive role of metalinguistic reflexivity. In this sketch I rely on the theoretical resources of the ecological-enactive approach to cognition.

The paper unfolds as follows. Section 5.1 is dedicated to defining the thesis that Metalinguistic Reflexivity is Constitutive of linguistic normativity (MRC). In Section 5.2, I discuss empirical evidence in line with MRC. Section 5.3 gives three arguments in favour of MRC. In Section 5.4, I discuss the *prima facie* devastating *regress objection* to MRC, as well as Searle's (1995, p. 72) response according to which 'language does not need language because it already is language'. I argue that the regress objection originates in the idea that language is an activity that is constituted by rules. In Section 5.5, I propose an alternative response to the regress objection. This response consists in first accounting for regular communicative behaviour, by relying on the theoretical resources from the ecological-enactive approach to cognition, and then showing

how the addition of metalinguistic reflexivity enables the emergence of linguistic normativity. On this account, language qua communicative activity is not constituted by rules. Instead, the addition of rules, formulated in metalinguistic vocabulary, enable people to regulate their communicative behaviour, thereby constituting linguistic normativity.

5.1 Defining

Linguistic behaviour is marked by *reflexivity*: we effortlessly shift between talking about other things and talking about talking itself. This ability to 'turn language on itself', as Davidson (1979) describes it, means that language itself is often an object of attention, not just in poetry and philosophy, but also in everyday situations. This reflexivity is part and parcel of our everyday language use. Take the following examples:

- Do as you are told!
- What did you say yesterday?
- Could you repeat that last word?
- I didn't understand a word he said.
- She is not a talkative person.
- He kept his speech rather short.
- That's not a nice thing to say.
- What do you call this thing?
- What do you mean by that?
- English is easy to learn.
- This book is incomprehensible.
- That accent is from Liverpool.
- Nothing rhymes with orange.
- That conversation got us nowhere.

This reflexivity of language is expressed in *metalinguistic practices* in which we attribute properties to language, the people that use it, and the activities that we can accomplish through its use (Agha 2007, p. 17). A metalinguistic practice consists of recurrent metalinguistic activities, and taking part in these metalinguistic activities requires metalinguistic skills on the part of

the participants. I use *metalinguistic reflexivity* to refer to the reflexivity of metalinguistic practices and the skills required to participate in them.

Metalinguistic reflexivity has received remarkably little attention in theoretical reflections on language.[1] The reason for this is that there are different ways of understanding the role of metalinguistic reflexivity. It can either be understood as an optional extension of linguistic behaviour, or as constitutive of linguistic behaviour; call these the *optional extension view* and *the constitutive view* (Van den Herik 2017 [Chapter 2, p. 53]). On the optional extension view, language and language learning can be explained without mentioning metalinguistic reflexivity. Here I briefly introduce the optional extension view as a contrast to the constitutive view I discuss next.

On the optional extension view, language is mostly experienced as a transparent medium that enables one to directly engage with what is meant. This unreflexive engagement with language is seen as the primary mode of linguistic behaviour on ontogenetic timescales. Instinctive mental capacities such as a *language acquisition device* (Chomsky 1965) or a *language of thought* (Fodor 1975) are surmised to support language 'acquisition' and linguistic performance outside the scope of our awareness. On such a view, metalinguistic reflexivity is often described in terms of *metalinguistic awareness*. This awareness is thought to develop after a child has acquired knowledge of language, as it is thought to

1 A notable exception is Taylor (e.g. 2000), whose work revolves around the notion of reflexivity and its constitutive role language and language learning. My understanding of metalinguistic reflexivity has been shaped decisively by his work. Other exceptions include the integrational linguists (e.g. Harris 1998, p. 28), who claim that, without verbal metalanguage, 'language as we know it would not be possible.'; Jakobson (1956/1985, pp. 120–121), who claims that metalanguage is 'an innermost linguistic probem' for the language sciences, and 'the vital factor of any verbal development'; Lucy (1993, p. 2), who claims that 'ordinary language structure and practice depends in crucial ways on the existence of metalinguistic forms and practices.'; Verschueren (2000, p. 439), who, building on Jakobson, claims that 'metalanguage is an important topic for linguistic research because it reflects metapragmatic awareness [which] is no less than the single most important prerequisite for communication as we know it'; and Cappelen and Lepore (2007, p. vii), who note that it is 'not implausible that a primitive capacity for the metalinguistic kicks in at the inaugural stages of language acquisition'.

rely both logically and developmentally on this knowledge. Cazden (1974, p. 13), who was one of the first to discuss the notion of metalinguistic awareness, writes:

> Metalinguistic awareness, the ability to make language forms opaque and attend to them in and for themselves, is a special kind of language performance, one which makes special cognitive demands, and seems to be less easily and less universally acquired than the language performances of speaking and listening.

Or, on the back of Gombert's (1992) *Metalinguistic Development:*

> At a very early age, the child is able to use and understand language correctly. Later comes the precocious ability to 'reflect' upon and deliberately control its use. Metalinguistic development, or the emergence of a reflective attitude to the comprehension and production of oral and written language, must be distinguished from that of ordinary verbal communication.[2]

This metalinguistic awareness is often operationalised in terms of explicit judgements of more technical aspects of language, such as grammaticality, reference, and ambiguity (Hakes 1980; Van Kleeck 1982; McDaniels & Cairns 1996; see Myhill & Jones 2015 for a recent overview). This leads to the idea that metalinguistic abilities concern decontextualised language, 'divorced from its utility for communication' (Smith Cairns 2015, p. 271), and therefore play no role in everyday language behaviour.

There are two differences between this concept of metalinguistic awareness and understanding metalinguistic reflexivity in terms of metalinguistic activities. *Firstly*, metalinguistic awareness is described in individualist terms, a consequence of the ever popular methodological individualism in the cognitive sciences. In this paper, however, I am interested in metalinguistic activities as a form of social behaviour and the normative role they play. Viewed in this way,

2 See Hakes (1980), Karmiloff-Smith (1986), and Bialystok and Ryan (1985) for similar accounts.

metalinguistic reflexivity is in first instance a property of social activities, not of individual cognisers. This does not mean that cognitive mechanisms are irrelevant. Of course children have to acquire certain skills. But in first instance, these skills are social skills that are expressed in interaction with others. *Secondly*, metalinguistic activities are not in general concerned with decontextualised language. Instead, metalinguistic activities are situated: they always occur in context. Although judgments of grammaticality, reference and ambiguity *are* metalinguistic activities on my understanding, they are a subset of a much broader class of metalinguistic activities.

The constitutive view gives metalinguistic reflexivity a constitutive explanatory role. De Jaegher, Di Paolo, and Gallagher (2010, p. 443) define the situation in which a phenomenon occurs as 'the collection of past and present events, processes and relations that are observed with a phenomenon'. Some of these events, processes, and relations will figure in an explanation of the phenomenon, in which case they are explanatory factors. According to De Jaegher, Di Paolo and Gallagher, these explanatory factors can have one of three different explanatory roles. First, variations in *contextual factors* produce variations in the target phenomenon. Second, *enabling factors* are causally necessary for the phenomenon to occur. This means that the absence of an enabling factor in a given situation prevents the phenomenon from occurring. However, the absence of an enabling factor does not imply that the phenomenon can not occur, as another enabling factor might take its place. For example, learning to write could be enabled by a pencil, but it could also be enabled by a pen. If no pencils are available, a pen will do. However, without a pen or pencil, one could not learn to write, bar of course alternative enabling factors. Finally, a factor is *constitutive* if it is a necessary part of of the phenomenon. As De Jaegher, Di Paolo, and Gallagher (2010, p. 443) explain, what 'exact role an elements plays in X [the phenomenon] depends on how one chooses to describe

and observe X'. This means that the claim that a given factor is constitutive of a phenomenon is a conceptual claim: if our concepts by means of which we describe a phenomenon change, what we take to be constitutive also changes (Di Paolo 2016). Take the explanation of *cognition* as an example. If cognition is defined as computations over mental representations with non-derived content, than perceptual processes are merely *enabling* factors for cognition. They are causally necessary, but not part of cognition itself (e.g. Adams and Aizawa 2010). If, however, cognition is defined as perceptually guided action (e.g. Varela, Thompson, and Rosch 1991), than perceptual processes *are* constitutive of cognition.

For the purposes of this paper, linguistic normativity is the phenomenon. The explanatory factor under investigation is metalinguistic reflexivity. The claim is that metalinguistic reflexivity is constitutive of linguistic normativity. In other words, given MRC, metalinguistic reflexivity, understood as metalinguistic practices and the skills required to participate in them, is a necessary part of the set collection of events, processes, and relations that are observed in relation to the phenomenon of linguistic normativity.

A *prima facie* reason for thinking that metalinguistic reflexivity is important for explaining linguistic normativity is the fact that metalinguistic activities often have a prescriptive function. To talk about language, a linguistic action, or some person in metalinguistic terms is to suggest treating it or her in a particular way. Of course, there is a continuum here. On the one hand, giving a definition for a new word is a wholly prescriptive act: there is no previous usage to conform to, and thus the definition sanctions a particular way of using the word. On the other hand, we could have a statement like *the word red is in the dictionary*, which is as descriptive as metalinguistic utterances get. In between these extreme cases there are intermediate cases. For example, an argument about the meaning of a word usually is not merely a factual disagreement about

past usage of that word; its outcome partly determines how that word *should* be used. In other words, metalinguistic practices normatively structure our linguistic practices (Taylor 2000; Harris 1998).

So far I have understood metalinguistic reflexivity as referring *both* to metalinguistic practices and the skills required to participate in them. We can however be more precise when we make a distinction between those two aspects. *First, sociogenetic MRC* is the thesis that metalinguistic practices are constitutive of the linguistic normativity inherent in linguistic practices. *Second, ontogenetic MRC* is the thesis that a child's metalinguistic skills are constitutive of the development of a sensitivity to linguistic normativity. In other words, only when a child has acquired the relevant metalinguistic skills, can her linguistic actions be justifiably described in normative terms. These two theses are interrelated but also have a certain independence.

On the one hand, sociogenetic MRC does not imply ontogenetic MRC. For example, one could argue that although metalinguistic practices are constitutive of the normativity of linguistic practices, an individual could develop the skills required to competently participate in these linguistic practices, and thus be subject to justifiable normative assessment, without developing the skills required to participate in metalinguistic practices. This view would be analogous to the division of linguistic labour proposed by Putnam (1975). On his account, the word *gold* refers to gold because there are speakers in a community that have a way of recognising gold – they have criteria for distinguishing gold from other substances. However, not everybody has to *know* these criteria in order to be able to use the word *gold* to refer to gold. The fact that *gold* refers to gold can thus be explained by the abilities of a subset of a community, and need not be explained in terms of individual knowledge.

On the other hand, ontogenetic MRC does imply sociogenetic MRC. The reason is obvious: ontogenetic MRC requires that children learn metalinguistic skills. But one cannot learn the skills required to participate in metalinguistic practices if those practices do not exist.

5.2 Empirical considerations

MRC has empirical implications. If MRC is true than it has to to be the case that wherever we find linguistic normativity, we also find metalinguistic reflexivity. The reverse is not true. It could also be that linguistic normativity is constituted by different processes, and that the metalinguistic practices we observe are merely an effect of normative linguistic practices.

Given these empirical implications of MRC, a first function of empirical evidence is that it could provide a counterargument to MRC. If we were to find a community in which there would be no metalinguistic reflexivity, yet there would be linguistic normativity, this would provide a defeating argument against MRC. Here it is worth noting that metalinguistic reflexivity is sometimes linked to written language: some theorists have pointed towards the transformative impact of literacy on language (e.g. Goody 1977; Ong 1982; Harris 2000; Love 2017). Therefore, one might hypothesise that metalinguistic practices are limited to cultures that have a facility with written language. However, oral cultures have rich metalinguistic practices (see for example Feldman 1991). Even the language of the Pirahãs, known for allegedly lacking recursivity, has rich metalinguistic practices.[3] It therefore seems reasonable to assume that wherever we find language and linguistic normativity, we find metalinguistic reflexivity.

3 For example, Everett (2009) translates Pirahãs as saying 'That's *right*' (p. 181) with respect to something Everett said, 'what do you *call* him?' (p. 184), 'The foreigner *said*...' (p. 204), 'We Pirahãs *speak our language* well' (p. 210), 'Now I'm going to *ask* for cloth' (p. 211), and 'Xaió!' meaning correct with respect to a certain way of saying something (p. 227).

A second function of empirical evidence is that it shows the possibility of linguistic practices that are radically different than our linguistic practices, with radically different normative structures. This in turn protects us from the *ethnocentric fallacy*, which Taylor (2010, p. 490) describes as the idea that 'the reflexive linguistic distinctions which our culture applies in evaluating and characterizing communicational behavior *must* also be applied – and if not explicitly, then implicitly – by the members of every culture.' The assumption behind the ethnocentric fallacy can perhaps be explained by the fact that entertaining alternative conceptualisations of language is highly counter-intuitive. The root of this counter-intuitiveness is that our means of conceptualising language by means of metalinguistic practices are not 'outside' language.[4] In this sense, language is a special kind of activity. Love (2003, p. 88, cf. Harris 1998, pp. 26–28) employs a comparison with musical notation. In the case of musical notation, things such as misunderstandings and differences in opinion as to how the medium relates to musical performance can be discussed and negotiated in language. In fact, they *have* to be negotiated in language, for there is no meta-notation: musical notation is not reflexive. In the case of language, however, there is no recourse to another medium to resolve disputes. An assertion about assertions is itself an assertion, and a question about questions is itself a question. As Love (Ibid.) words it: 'Language, being language, is on its own. It is interpretatively terminal.' This means that we lack a method of neutral comparison of different ways of conceptualising language. The illusion therefore arises that our ways of metalinguistically understanding language, in terms

4 At first glance one could think that technical terms that originate in theoretical inquiries into language might provide an alternative conceptualisation. However, a description of the phenomenon relies on lay metalinguistic typifications. For example, although language theorists have tried to define 'word' in very many different ways, it seems impossible to theoretically reflect on Western language without accounting for words. Linguistic inquiry is thus properly viewed as an extension of, not an alternative for, lay metalinguistic typifications. In this way, 'linguistic inquiry is conditional on the reflexivity of of language' (Harris 1998, p. 26).

of talking about things (reference), truth, words and sentences, and so on, are the *only* possible ways to conceptualise language (Lucy 1993; Whorf 1956).

One way of understanding the constitutive nature of metalinguistic reflexivity is to see that metalinguistic practices from other languages do not apply to English. For example, the Ilongot categorise all speech in three categories (Rosaldo 1973): straight speech (*qube:nata qupu*), crooked speech (*qambaqan*), and invocatory speech (*nawnaw*). Straight speech is everyday, ordinary speech, and is used to for gossiping, chatting, or exchanging news. Invocatory speech is 'characterized by frequent and often hyperbolic metaphors, redundant rhythms, and stereotyped lines' (Ibid., p. 197). The best example are long spells used to coerce gods and spirits, but one can also *nawnaw* children, which is to lecture them, or to coerce them by telling them lies ('a child who insists on eating slowly will be told that airplanes, or soldiers, are coming to take away his food' (Ibid.)). Unlike children, adults have to persuaded differently in the egalitarian Ilongot society. Crooked speech is a subtle mode of speaking which enables one to 'hide' from their words, 'a distancing of the speaker from his words' (Ibid., p. 198) by using devices such as metaphor and qualification. For the Ilongot, all speech can be categorised using this tripartite distinction. At the same time, from the descriptions of three kinds of speech, it is clear that this distinction does not apply to English speech.

A second way of understanding the constitutive nature of metalinguistic reflexivity is to see that our metalinguistic practices are not applicable to all languages. For example, in Western thought, one of the primary linguistic phenomena is reference and one of the primary speech acts is the assertion. This is not only the case in theoretical reflections on language: in everyday language we ask what somebody is talking *about*, and we say that a child learns learns language by 'learning words', which amounts to learning what a given word *stands for*. The Ilongot, on the other side, see *tuydek*, or commands, as the

primary speech act, and claim that 'children learn to speak by learning *tuydek*' (Rosaldo 1982, p. 209). This perspective on language and language learning means that the Ilongot do not dissociate language from action: 'knowing how to speak itself was virtually identical to knowing how and when to act' (Ibid.). Based on an analysis of Ilongot metalanguage, Rosaldo (1982, pp. 214–215) concludes that:

> To … "share knowledge", or "tell stories" are, then, in most Ilongot speech, a matter less of representing facts about the world in words, than of articulating relationships and claims within the context of a history that is already known. … [Assertive verbs'] power seems much closer to the force that Searle assigns to our "declaratives" than to assertive acts like "arguing" and "stating facts".

Based on a focus on consensus and equality, the Ilongot culture seems to understand their speech acts that come closest to what our assertions have declarative force: they serve not to make a claim about the world, but rather to change the state of the world by articulating and manipulating social bonds instead. In other words, given their metalinguistic practices, the Ilongot do not understand their own linguistic behaviour in terms of merely talking about things, and as consisting in making and arguing about assertions. This has important implications for the normative functions of their language, which cannot be understood in isolation of articulating and manipulating social bonds.

What the description of Ilongot linguistic practices shows is that linguistic communities can have very different metalinguistic practices. These metalinguistic practices have normative implications. For example, for the Ilongot, one *should* use 'crooked speech' in convincing adults, whereas for convincing children 'invocatory speech' *should* be used. Moreover, the criteria for learning language are different from ours. We would say that learning a language involves learning words, where a child has 'learned a word' when it can reliable associate a sound with the right objects or situations. For the

Ilongot, however, the learning language is learning *tuydek*, and this cannot be disassociated form other forms of behaviour in the way we can dissociate language from behaviour.

5.3 Motivating the constitutive role of metalinguistic reflexivity

While the empirical evidence discussed so far is interesting, it does not, by itself, entail MRC. An alternative explanation of the observed linguistic diversity and corresponding metalinguistic diversity could be that linguistic practices differ, and that this difference is reflected in metalinguistic practices rather than constituted by them. In other words, for MRC to be true, it should not only be the case that there is *in actual fact* no community that has a command of linguistic normativity in the absence of metalinguistic reflexivity, we need to establish that there *could not be* linguistic normativity without metalinguistic reflexivity. In this Section, I give three arguments in favour of this modal claim.

First of all, metalinguistic reflexivity is required for determining, (re)negotiating, and showing an understanding of standards of correctness. Noë (2015, pp. 41–42, *emphasis added*) formulates this as follows:

> Language brings with it the possibility of misunderstanding, and there could be no instruction in language without the need to articulate and express a conception of how one ought to speak or what are the correct or preferred or most comprehensible ways of speaking. […] There could not be language as we know it in the absence of attitudes, values, norms, prescriptions, and ideologies about language. There never was a Garden of Eden, and so there never was a language that we carried out and made use of freely, automatically, always *without need for reflection on such basic questions* as How to go on?

> What is right and what is wrong? What does he mean when he says that? Why did he say that? and so on.

In other words, language cannot be understood as a set of context-invariant tools that are applied in a given context. One of the defining characteristics of language is that every context of use is unique, which brings with it the possibility of breakdowns in communication, breakdowns that need to be resolved in the moment. Linguistic behaviour therefore cannot be reduced to merely regular behaviour or knowledge of conventions (Davidson 1986). We constantly have to (re)negotiate our metalinguistic understanding of what is going on in conversation, and for this we need to have recourse to a meta-conversation in which we conceptualise what we are doing a we are using language.[5] This argument works both on the level of practices and on the level of the individual. Without metalinguistic practices such as *explaining the meaning of a word*, and *criticising another person's language use*, we could not have the kind of training that is essential in learning a language. Moreover, if we were to lack the resources to discuss, criticise, an (re)negotiate 'the meaning' of words, this would entail that no criteria for correct use could be agreed upon, explained, changed, or enforced (Taylor 2000, p. 489; cf. Van den Herik *under review* [Chapter 4]). To have 'correct' uses of words thus requires a way of agreeing on what the correct use *is*, and this can only be done by means of metalinguistic reflexivity. At the level of the individual, participation in metalinguistic practices is a criterion for being a competent language user. If you cannot explain what you mean when you have said something you simply do not count as a competent language user. This is the case not just for children first learning language, but continues to hold true later in life. If a student is able

5 In conversation it is very common to engage in this meta-conversation. For example, Dingemanse et al. (2015) found that across twelve different languages, *other-initiated repair* (examples in English: 'Huh?', 'What do you mean?', 'Who?') is a frequent phenomenon. It occurs on average once per 1.4 minutes of conversation, with conversations rarely going on for five minutes without such a repair. Dingemanse *et al.* (Ibid., p. 10) therefore take reflexivity to be a 'core element' that is 'essential to human language'.

to reproduce sentences he has read in a philosophy paper, but is unable to explain what these sentences mean in his own words when we ask him to do so, we are unable to ascertain whether he understands them – although we shall have our suspicions.

Second, the attribution of understanding is crucial for monitoring the coordinative function of language. For any linguistic action that has the aim to coordinate with another person beyond the current situation necessarily relies on the *possibility* of a meta-conversation in which questions of understanding and agreement *can* be addressed whenever the need arises. If you ask me to perform a task in the future, and it is important to you that the task is performed, you will want to make sure that I have understood you. If we lacked metalinguistic practices, we could not engage in this meta-conversation, and therefore could not do such things as check whether someone understood an utterance or to ask someone to explain what they mean when we did not understand them.

A third, closely related, normative function of metalinguistic reflexivity is that it enables us to hold each other accountable for our linguistic actions. The reason is the following: in order to hold someone accountable for what they have done, we have to *know* what that person has done (Enfield & Sidnell 2017). This knowledge must be agreed upon between at least the person who holds another accountable and the person who is held accountable.[6] This means that a public medium must be available by means of which it can be determined what a person has done, and the medium we use for this is language. In other words, without language there can be no social accountability. This argument holds also in the case of actions achieved by linguistic means - think for example of promises and other commitments. And thus, holding another person accountable for their linguistic actions constitutively requires metalinguistic reflexivity. As an illustration, take the following utterance: 'I'll finish the task

6 A person that does not know what he is doing is usually taken to have diminished responsibility, and therefore cannot be held accountable.

tomorrow'. If metalinguistically construed as a promise, the person can be held accountable for finishing the task tomorrow. However, the speech act can also be construed as a prediction, or as an educated guess as to when they will find the time to finish the task. *How* the utterance is metalinguistically construed thus determines the commitment the person makes. In everyday interaction, we need not always resort to explicit metalinguistic activities to construe our linguistic acts. However, for any number of reasons, such as the situation being ambiguous, or the stakes of the task being finished being very high, we *can* resort to explicit metalinguistic negotiation of the commitment. In the above example, one could respond by a proposed metalinguistic construal such as 'Is that a promise?', or 'Can I hold you to that?'. Besides this ongoing (re)negotiation of the commitment undertaken by performing a linguistic action, another important reason why metalinguistic reflexivity is crucial for social accountability is that we require a way of reporting on past speech in order to hold people accountable *now* for what they said *earlier*. We need a means of referring back to that act to justify that someone else is accountable ('But you said that you would finish the task today!').

5.4 The regress objection

There is a potentially defeating counterargument against MRC: the regress objection. In this section I introduce this regress objection based on Searle's (1995) account of institutional facts. I also discuss his enigmatic response, according to which *language does not require language because it already is language*.

Searle (1995) starts from institutional facts – facts that are dependent on human agreement in institutions, such as money, property, research grants, and marriages. These institutional facts are constituted by rules of the form

'X counts as Y (in context C)'. For example, submerging a newborn in water counts as baptising it (when done by a priest), and saying 'sorry' counts as apologising (if sincerely uttered). Searle (1995, p. 27) distinguishes between *constitutive rules* and *regulative rules*. Regulative rules *regulate a pre-existing* phenomenon. Searle gives the example of the rule 'drive on the right side of the road'. The activity of driving is not constituted by this rule; it is perfectly possible to drive a vehicle without a rule being in place that regulates on which side of the road one should drive (in the case of certain vehicles, even the road is optional). *Constitutive rules*, on the other hand, 'create the possibility of certain activities' (Ibid.). Here Searle's example is chess: without the rules of chess, one cannot play chess. Playing chess involves, as Searle says 'acting in accord with the rules'. In fact, we might make this stronger by saying that playing chess involves one's behaviour being guided by the rules. Merely acting in *accord with* the rules is not enough. If two monkeys, by accident, one after the other, move a piece on a chess board in such a way that it happens to be in accord with the rules, we would not say that they have played chess.

According to Searle (1995, pp. 69–70), the *constitutive rules* involved in institutional facts require language for their existence. The reason for this is that the Y status of Xs needs to be marked, as the Y status goes beyond the physical characteristics of Xs. For example, a ball crossing a chalk line *counts* as a goal in football. This status as a goal can only be explained in terms of people keeping track of the goals in some way. Moreover, in order to be effective in the interactions between people, the status has to be marked in a public way.[7] Searle (1995, pp. 69–70) summarises:

> Because the Y level of the shift from X to Y in the creation of institutional facts has no existence apart from its representation, we need some way of

7 Of course, in playing an informal game of football people might keep track of the score without verbally expressing the score after a goal is scored. In this case, the keeping track of the score is potentially public, as everybody could express the score they are keeping track of.

> representing it. But there is no natural prelinguistic way to represent it, because the Y element has no natural prelinguistic features in addition to the X element that would provide the means of representation. So we have to have words or other symbolic means to perform the shift from the X to the Y status.

Searle distinguishes between institutional facts and 'conditioned forms of habitual behaviour'. Animals, for example, can have a territory. They can even mark their territory by physical means, for example by leaving a scent. This, however, doesn't constitute an institutional fact, even if other animals, as a matter of fact, do not traverse the boundary of the thus marked territory. Only when the animals recognise rights and obligations that go beyond their inclinations to act, e.g., that 'they are *forbidden* to cross the line, that they are not *supposed* to cross it' (Ibid., p. 71), can we speak of institutional facts.

When Searle (1995, p. 60) gets to language, he encounters the regress objection. The reasoning is the following: 'If institutional facts require language and language is itself an institution, then it seems language must require language, and we have either infinite regress or circularity.' Note that in the case of language, the X in the constitutive rule 'X counts as Y (in C)' is linguistic. This means that the constitutive rules of language are reflexive: they are metalinguistic in nature.[8]

Based on Searle's account, we can formulate the regress objection in terms of our earlier discussion as follows:

[1] Learning language is learning to follow the constitutive rules of linguistic practices.

[2] Learning to follow the constitutive rules of linguistic practices requires being able to participate in metalinguistic practices that enable the formulation of these rules.

8 This is made explicit in Sellars' (1954, p. 1) formulation of the regress objection, but is left implicit in Searle's discussion.

[3] Therefore, learning language in linguistic practices requires *already* being able to participate in metalinguistic practices.

[4] But, being able to participate in metalinguistic practices means that you have already learned language.

[5] Therefore, learning language requires already having learned language.

Note that the regress argument thus sketched, if viable, provides a defeating argument against ontogenetic MRC. Given the fact that ontogenetic MRC entails sociogenetic MRC, it also provides a defeating argument against MRC *simpliciter*. The reason the regress objection would be a defeating argument is that the conclusion is obviously absurd, as it entails that language cannot be learned. Therefore, either [1] or [2] has to be false. Searle's solution is to accept [1] and argue against [2]: Learning to obey constitutive rules, at least in the case of language, does not require being able to participate in the metalinguistic practices that enable the formulation of these rules. In order to avoid the potential infinite regress Searle thus has to deny the force of his own argument in the case of language. Doing so requires giving language a special status, or as Searle (1995, p. 73) puts it:

> The solution to our puzzle is to see that language is precisely designed to be a self-identifying category of institutional facts. The child is brought up in a culture where she learns to treat the sounds that come out of her own and others' mouths as standing for, or meaning something, or representing something. And this is what I was driving at when I said that language doesn't require language in order to be language because it already is language.

Where most institutional facts require language for their identification, language is 'self-identifying' according to Searle. And by this Searle means that the child must have a special ability to learn language: 'the capacity to attach a sense, a symbolic function, to an object that does not have that sense intrinsically is the *precondition* not only of language but of all institutional reality'

(Ibid., p. 75, emphasis added). Searle opts to assume that children must already have the capacity to attribute a status. The explanation of institutional facts thus ultimately relies on 'primitive prelinguistic psychological states' (Ibid., p. 78) that in an important sense already contain what needs to be explained. For assuming that children can attach such a status means that they can already obey constitutive rules of the form 'X counts as Y (in C)' *before they first learn language*.

Searle's response to the regress objection is thereby exemplary of the representational theory of mind and the way this theory deals with language. On a representational approach, the key properties of language, including its normativity, are explained in terms of those very same properties, which are thought to already be in place in the cognitive system of the child in her mental representations. In the words of Shapiro (2014, p. 5), 'External symbols acquire their meaning from meaningful thoughts—how could it be otherwise?' The regress has been evaded, but only by creating a new problem. For instead of answering the question how children can learn to be sensitive to the normative properties of language, we now have to explain how mental representations can have normative properties.

This in turn means that Searle's account is living on borrowed time: without an account of the origins of the normative properties of mental representations, the theoretical move to make mental representations constitutive of institutional facts does not solve the problem at hand. Of course, every explanation must stop somewhere, and the concept of mental representations seems sufficiently well established to use it as an *explanans*. There is, however, a growing concern that representational approaches will ultimately be unable to account for the normative properties of mental representations (Hutto & Myin 2017). The research programme that aimed to naturalise intentionality in terms of mental representations consumed a lot of energy in the 1980s and 1990s. The

most promising of the views that came out of these debates is teleosemantics (Millikan 1984), which aims to explain mental representations in terms of natural teleology, under the guise of evolved biological functions. However, the widespread optimism has subsided, and naturalising intentionality in terms of mental representations now 'bears all the hallmarks of a degenerating research program' as it 'has run up against principled obstacles it seems unable to surmount' (Kriegel 2011, pp. 3–4). The main problem is that there simply is 'a root mismatch between representational error and failure of biological function' (Burge 2010, p. 301). In the words of Putnam (1992, p. 33) 'evolution won't give you more intentionality than you pack into it'.

It is not my goal here to discuss this debate in detail. Instead, my goal is to provide an alternative to the representational solution offered by Searle by building on a non-representational approach to cognition, the ecological-enactive approach. The key move in avoiding the regress objection is to deny that language is an activity that is *constituted* by rules, and instead see linguistic normativity as being constituted by *regulative* rules that originate in reflexive uses of language. In the next Section, I show how this move enables us to explain a child's initiation into language.

5.5 An alternative response to the regress objection

There is a growing movement in the philosophy of cognition that explains cognition without invoking mental representations, known under the banner of 4E cognition (cognition is embodied, ecological, extended, enacted).[9] In this last

9 Note that not all approaches under the 4E banner eschew mental representations. As Chemero (2009) argues, there are two different types of 4E approaches to cognition: on the one hand there is an approach that builds on the representational theory of mind, and claims that sometimes non-neural bodily states or extra-organismic states can be representational vehicles (e.g. Clark and Chalmers' (1998) extended mind hypothesis), on the other hand there are approaches that aim to thoroughly reconceptualise cognition (e.g. J.J. Gibson 1979; Varela, Thompson, Rosch 1991; Thompson 2007; Chemero 2009; Noë 2012; Hutto & Myin 2017).

section I want to propose a possible response to the regress objection against MRC by building on one of the approaches that fly under the 4E banner, namely the ecological-enactive approach. The main idea is to first account for regular communicative behaviour, and show how the addition of metalinguistic reflexivity allows for the constitution of linguistic normativity by providing regulative rules.

5.5.1 The ecological-enactive approach

According to J.J. Gibson (1979, p. 215), the question an approach to cognition has to answer is how behaviour can be regular without being regulated. The approach to cognition he pioneered starts from two premises in order to account for this regularity: (1) organisms perceive affordances, which are possibilities for action, and (2) the environment of the organism is structured in such a way that these affordances are directly perceivable. The regular behaviour of organisms can be explained in terms of acting on perceived affordances. Affordances are relational: they depend both on the layout of that environment and on the skills of the organism (Chemero 2009). Organism and environment are thus mutually specifying. This means that an organism's environment is not the physical environment, but its ecological niche. In the case of humans, the ecological niche 'is shaped and sculpted by the rich variety of social practices humans engage in' (Rietveld & Kiverstein 2014, p. 326). We do not perceive affordances in isolation. Instead, we always perceive a multitude of possibilities for action, or a *field of affordances* (Rietveld & Kiverstein 2014). Not all affordances equally motivate a person to act on them. Some affordances stand out and are very motivating, others are less motivating. What affordances are motivating depends on a host of factors: what activity one is currently engaged in, one's mood, the social setting, and so on.

On this approach to cognition, communicative behaviour is in first instance understood as directly acting on the field of affordances of others, by bringing affordances to their attention (e.g., Baggs 2015). These affordances thus become more motivating, and it becomes more likely that the person acts on these affordances. When applied to language learning, a child's 'first words' do not function by standing for things, instead they are social actions aimed at directing other people's attention in order to do something together (Verbrugge 1985; Bruner 1990; Reed 1996; Roberts 2002). I have called these *attentional actions* (Van den Herik 2018 [Chapter 3]). What a child learns as she learn to say her first words is thus not referential knowledge (she does not learn that *ball* stands for balls), she rather learns a social skill, namely to indicate aspects of recurrent situations while engaging in joint activities with others. Attentional actions function like ostensive gestures: they direct attention. It is important to realise that this directing of attention is not a mute pointing. Attentional actions act as 'operators of reminiscence' (Bottineau 2010, p. 283), linking the present situation to previous situations and thereby suggesting a way of going on.

Note that this account of attentional actions is also applicable to linguistic behaviour beyond a child's first words. Our linguistic activities continue to fulfil an ostensive role, and thus remain attentional actions. Talking about some thing is to draw attention to that thing, and this can be understood on the model of attentional actions, even in the case of highly theoretical language use.[10]

It is important to realise that the learning situation from the perspective of the child is very different from the same situation from the perspective of the caregiver teaching the child these skills. As Williams (2010a) puts it, the child is 'calibrated' by her caregivers so that she responds to the world as they already do. There is thus an asymmetry in the learning situation: whereas the caregiver

10 Kukla (2017) states that Haugeland, in a seminar, said that philosophy 'was just a particularly sophisticated and elaborate form of ostension; we use philosophical discourse to direct one another's attention to how things are'.

evaluates the child's verbal behaviour in normative terms (as being (in)appropriate, (in)correct, and so on), the child lacks the skills to participate in these metalinguistic practices. Fortunately, in order to enter her linguistic community, the child only has to attune to the regularities in behaviour of her community. In other words, a properly 'calibrated' child verbally responds to her environment in ways similar to how others in her community respond *without* having the skills to entertain whether her verbal behaviour is correct or incorrect. However, because her environment is structured by the metalinguistic practices, and her caregivers sculpt her behaviour so as to be appropriate with respect to these practices, it does not matter that the child does not understand yet. All the child has to learn is to guide others' attention and have her attention be guided by the verbal behaviour of others.

One more perceivable aspect of the child's environment is the linguistic behaviour of herself and others. A child's first forays into her community's metalinguistic practices can thus be described in similar terms (see Van den Herik *under review* [Chapter 4] for an account along these lines). That is, the child's reflexive communicative skills are in the first instance reflexive attentional actions: attentional actions that direct attention to other attentional actions. This means that the child's initiation into her community's metalinguistic practices is also not to be described in normative terms. All we need are systematic links between reflexive attentional actions and verbal aspects of the environment of the child. How these systematic links come about can be described non-representationally in terms of attentional actions.

5.5.2 Normative bootstrapping

At this stage, the child's behaviour is in accordance with the rules because the caregiver uses the rules to evaluate the child's behaviour and condition her behaviour accordingly. However, the child does not yet obey the rules, nor is

she aware of those rules. In order to make the move from merely acting in accordance with the rules to obeying the rules, the child has to use the normative, metalinguistic terms she gradually learns to use to evaluate and guide her own behaviour. The crucial insight, which I take from Brown (2006), is that this regularity without regulation plays an important part in understanding how the child can make the shift from merely regular behaviour to rule-following behaviour. The key is that, as a result of her behaviour being in accordance with the rule, the child is *justified* in applying the normative metalinguistic terms to her own behaviour. This *retroactive* interpretation of her own behaviour in terms of the rules enables the child to conceive of her own behaviour as being guided by the rules *before* that was actually the case. As Brown argues, this retroactive interpretation of her own behaviour does not require backward causation, for the child's new skills consists in *interpreting* her own past behaviour as being guided by rules, not in *describing* it as such. As soon as the child can interpret her own behaviour in these normative terms, she can also conceive of herself as being subject to the rules. The child can do so because she knows she *can* behave in accordance with the rules.

Take the example of a child's 'first words'. Of course, for the child, these are not words yet, and she is also unable to provide rules for using these words. Only later will she acquire the reflexive metalinguistic skills that enable her to reflect on the regularities in her own and others verbal behaviour, and to treat these regularities as rules, i.e., as standards of correctness. The child will do so by providing criteria for correct use, for example in giving an ostensive definition (*this color* ☞■ *is black*), by telling other people *you are wrong*, by giving a verbal definition, and so on. In doing so, the child's relation to the regularities in her communicative behaviour changes. In first instance, she must, as Williams describes it (2010a), 'blindly obey' her caregiver when he corrects her. But as she learns to participate in the game of providing and discussing criteria that

separate correct from incorrect use of words, she gradually becomes sensitive to the normative dimension of language.

Note that Brown's idea of retroactive interpretation does not entail a 'light-bulb' moment when the child all of the sudden makes the leap from merely regular behaviour to truly normative behaviour. The command of meta-linguistic normative terms that enables the child to describe her own behaviour in normative terms can be acquired in a piecemeal fashion. But it does help us to explain how the child can justifiably see herself as being guided by rules in coming to grasp the normative terms of her community. By thus learning to wield metalinguistic terms, the child can gradually come to regulate her own behaviour, rather than being regulated by her caregivers.

This retroactive interpretation can not only explain how a child's sensitivity to norms can be bootstrapped, but can also explain how normativity developed in first instance. Brown (2006) gives a possible story that starts from the observation that it is evolutionarily advantageous for social animals such as our ancestors, to be able to signal a willingness to cooperate with those willing to reciprocate. He uses the example of two people who can, by working together, gather more food than they could individually. The signal that advertises willingness to cooperate need not be understood in a normative fashion as undertaking a commitment to work together; it only needs to be a reliable indicator of future behaviour. In the terms of this paper, we could call this an attentional action that directs attention to the affordance of working together. We can further imagine that it would be useful for our ancestors to keep track of these expressed 'commitments' by means of reflexive attentional actions. For example, it could be useful to make a distinction between people that can-be-trusted and people that are not-to-be-trusted when they make the willingness-to-cooperate signal. The not-to-be-trusted signal also need not be understood in a normative fashion: it could be used to indicate a person who usually does not

cooperate when making the signal, but instead takes all the gathered food at the end of the day. In a third step, we could have noticed that people that are indicated by the not-to-be-trusted signal fail to find a partner to cooperate, thereby making it advantageous to not engage in behaviour that results in being marked not-to-be-trusted. At this point we have a primitive form of normativity in play: people that behave so as not to be marked not-to-be-trusted have a reason to commit to working together when they produce the signal that indicates a willingness to cooperate. Of course, the story outlined above is a just-so story. It does show, however, how a primitive sort of normativity of communicative behaviour can arise from merely regular forms of behaviour.

5.5.3 Rules as resources

As Brown concludes rules only apply to us in as far as we view ourselves as being subject to them, that is, in as far as we have the reflexive resources for enforcing them. This idea of normativity does not mean that anything goes, for our sensitivity to the normativity of rules is grounded in regular forms of behaviour, and these regularities are grounded in a world that acts in predictable ways.

As Davidson (1986) argued, the interpretation of concrete, situated linguistic behaviour cannot be reduced to (knowledge of) conventions regarding the use of linguistic forms. Understanding what someone says is always a situated affair, and draws as much on our knowledge of the world as on our linguistic knowledge, a conclusion that blurs the boundaries between linguistic and non-linguistic knowledge. Earlier we said that metalinguistic reflexivity enables us to determine the meaning of a word, or the rules for using it, by determining and enforcing criteria for correct use – a metalinguistic activity. However, this process does not lead to an inflexible repertoire of linguistic actions, which would be wholly unusable as it would lack the possibility to

adapt to concrete situations. The formulation of rules and criteria for applying them is thus never completed, but remains a matter of ongoing determination and (re)negotiation (cf. Di Paolo, Cuffari, & De Jaegher 2018).

Viewed in this way, linguistic rules and the criteria for applying them do not determine or govern linguistic behaviour. *Pace* Searle, our linguistic activities are thus not constituted by rules in the way that playing chess is constituted by rules (cf. Love 1999). If you do not follow the rules of chess, you do not play chess. But in language, this does not hold. Instead, what comes first, both on ontogenetic and sociogenetic timescales, is shared forms of regular communicative behaviour that can be understood in terms of attentional actions. Linguistic rules then originate in a recognition of this regularity, and the subsequent use of this recognised regularity as a standard of correctness (Hacker 2014, §5). However, which rules apply is up to the participants themselves to decide. For example, by and large people's linguistic behaviour is in line with the rules of grammar. However, if under extreme stress, or when exhausted, or for comedic effect, someone might resort to single words in order to communicate. Thus 'breaking the rules' of grammar does not entail that his behaviour is no longer linguistic. Other examples include using a word in a novel way, thereby breaking with regular usage, coming up with novel words, and so on.

Instead of constituting linguistic behaviour, rules are available as *regulative resources* to people by means of which they can regulate their communication behaviour (Edwards 1997). If the situation requires it, a recourse to metalinguistic rules and criteria provides participants with ways of solving a breakdown in communication and negotiating the communicational import of utterances in an unfolding conversation. Here it is crucial to see that the metalinguistic skills that we learn in order to navigate the normativity of language are practical skills, aimed at coordinating our behaviour with others. Take the practice of making promises we discussed earlier. In everyday circum-

stances, people are not interested in a metaphysics of promises. But they *are* interested in whether someone sincerely made a promise, whether he will keep his promise, and what keeping a promise entails.

We therefore *can* compare linguistic rules with traffic rules (Wittgenstein 1976, §440). Traffic rules normatively regulate the behaviour of people as they partake in traffic and people often act in accordance with the traffic rules because they have been trained to do so. At the same time, traffic rules do not determine or govern what people do in traffic. A person might choose to ignore traffic rules. This could be because they are late, like the sound the engine makes at 150 km/h, or a situation occurs in which it is unclear to them how the rules should be interpreted. In the case of a breakdown of traffic behaviour, say an accident, the rules can be cited to ascertain who was at fault – 'You should have given right of way!' – and can thus be used to determine accountability. In the absence of traffic rules, no such judgement could be made. In his example of using the rule 'drive on the right side of the road' as a regulative rule, Searle remarked that the activity of driving does not depend on the rule. And in a way, we could have a form of traffic without having traffic rules. But this would be something very different from traffic as we know it. While it is therefore true that there could be a kind of traffic without traffic rules, the addition of regulative rules introduces a normative dimension to traffic that it otherwise would not have. In other words, while traffic is not constituted by rules, but merely regulated by it, traffic *qua* normative activity *is* constituted by rules.

The same goes for language. Linguistic behaviour *qua* communicative activity is not constituted by rules, but the normativity of language *is.* It is the addition of regulative rules, i.e., the reflexive recognition of regularities and the use of this reflexive recognition in determining criteria and standards of correctness, that makes our communicative behaviour into a normative practice.

However, where language differs from traffic is that the rules themselves can become object of (re)negotiation in the moment itself.

5.6 Conclusion

In this paper, I have defended the thesis that metalinguistic reflexivity is constitutive of linguistic normativity. I argued that without metalinguistic practices, we would be unable to determine and show an understanding of standards of correctness. The meta-conversation afforded by metalinguistic practices enables us to monitor the coordinative function of language, and (re)negotiate commitments undertaken by linguistic acts. A potential defeating counterargument against the constitutivity of metalinguistic reflexivity is the regress objection. We saw that at the root of this regress objection lies the idea that learning language requires learning to follow the constitutive rules of linguistic practices. On this view, as advocated by Searle, language is like learning chess: we *first* learn the rules and then put them in practice. This means that knowledge of the rules explains regularities in behaviour. On the ecological-enactive alternative I sketched, learning language can be understood as first learning regular communication behaviour, *and then* learning to reflexively recognise these regularities and use them as standards of correctness. On this view, it is only because a child *first* behaves in regular ways, that she can *then* interpret her own behaviour in normative metalinguistic terms, that is, as being guided by rules. Metalinguistic reflexivity thus enables regulation of communicative behaviour, and thereby constitutes linguistic normativity. The account developed in this paper thus enables us to understand how metalinguistic reflexivity constitutes linguistic normativity without falling prey to the regress objection.

6 Conclusion and directions for future research

> As philosophers, we are in the business of promoting possibilities.
>
> – Hutto & Myin, *Radicalizing Enactivism*

> The relevant sense of 'knowing the meaning of words' (which is a form of what Ryle has called *knowing how*) must be carefully distinguished from knowing the meaning of words in the sense of being able to talk about them as the lexicographer might – thus, defining them. Mastery of the language involves the latter as well as the former ability. Indeed, they are both forms of *know how*, but at different levels – the one at the 'object language' level, the other at the 'metalanguage' level.
>
> – Wilfred Sellars, *The Structure of Knowledge*

In this thesis, I have developed an ecological-enactive perspective on language. As mentioned in the introduction, the two functions of this perspective are to contribute to extending the reach of the ecological-enactive approach to typical human forms of cognition, and to throw new light on philosophical problems regarding language. In order to achieve the first goal, I have chosen to discuss aspects of language that are most straightforwardly described in representational terms: linguistic knowledge, concrete words, and content. If the ecological-enactive perspective on these aspects of language developed in this thesis is viable, this shows that the ecological-enactive approach *can* be extended to include typical human forms of cognition.

With respect to the second goal, the perspective developed in this thesis conceives of language in line with the pragmatic turn in cognitive science.

Language is conceived in action-oriented terms, as the context-bound and time-bound exercise of skills, rather than in terms of knowledge of an abstract system. In exercising these skills, we continuously make and remake language in line with a history of learning. In an important sense, language is thus what we make of it – how could it be otherwise?

In this thesis, I have provided metalinguistic rather than metaphysical solutions to philosophical problems. The ecological-enactive approach opposes the reifying tendencies of representational approaches that locate metalinguistic abstractions such as meanings and words as representations in the cognitive system of individuals. But in opposing these reifying tendencies, I have not turned a blind eye to these metalinguistic abstractions, opting instead to treat them as standing in need of explanation. As I have argued in this thesis, metalinguistic reflexivity is crucial for explaining the origins of phenomena like linguistic knowledge, semantic content, and linguistic normativity. By taking the normative structuring of metalinguistic practices seriously, these phenomena can be explained without reifying them.

Rather than understand language and cognition as reified abstractions with determinate properties, both the ecological-enactive approach and the metalinguistic approach put possibilities centre stage. For example, as discussed in Chapter 4, what separates the contentful linguistic behaviour of the competent adult from the contentless speech acts of the young child is the possibilities the adult has to engage in metalinguistic activities such as justifying, defending, and explaining what she said, when the need arises. In doing so, she does not describe the properties of her behaviour, but suggests a way of understanding them. And although an individual's history constrains the possibilities that are available to her, it does not determine what will happen. As discussed in Chapter 5, rules are not explanatory hypotheses, but resources that people can employ in order to normatively regulate social activities. According

to the perspective developed in this thesis, one of the defining characteristics of linguistic behaviour is therefore its open-endedness: the possibility for (re)negotiating its properties in the moment.

As I said, any perspective on language will be limited. Much work remains to be done. In the final paragraphs of this thesis, I identify two directions for future research.

The first direction for future research concerns the relation between linguistic and non-linguistic forms of cognition. How should we view the impact of language on non-linguistic or pre-linguistic cognitive abilities? In thinking about this question, I think it is important to remind ourselves that the distinction between the linguistic and the non-linguistic is not easy to draw on the current perspective. One reason for this is that learning language has effects that outstrip verbal behaviour. Think for example of the categorical perception discussed in Chapter 3. Here we saw that linguistic categories have category effects on putatively non-linguistic processes, in particular categorical perception. Once we have learned colour categories that are grounded in the linguistic distinction we make, these categories influence our perception even in the absence of verbal behaviour.

Of particular interest here is the idea defended in this thesis that a particular speech act is contentful if the person who performed that action is sensitive to the conditions under which that speech act would be correct, a sensitivity that is expressed in the person's possibilities for engaging in metalinguistic activities such as justifying her speech act when the need arises. I see no principled reason why this analysis cannot be extended to non-verbal acts as well. In this way, our practical dealings in everyday situations can also be subject to the same assessment. For example, a person that tries to lift an object that turns out to be heavier than she expected, can give a reason for her behaviour by saying 'I thought I could lift that'. This reason-giving behaviour could be construed as

post-hoc rationalisation, where the contentful attitude of taking this thing to be liftable is only attributed after the fact. Alternatively, we could say that the action of trying to lift the object is an expression of the fact that the person took the object to be liftable. I think that the perspective developed in this thesis aligns more closely with the second option, in particular if we think of perception in terms of the task-specific devices as discussed in Chapter 3, but of course much more needs to be said in order to substantiate this claim.

The second direction for future research concerns an ecological-enactive account of abstraction. In a sense, the problem faced by radical embodied approaches to cognition is the opposite of the symbol grounding problem faced by classical cognitivism (Harnad 1990). The symbol grounding problem is to account for the link between amodal and abstract symbols and the world. The ecological-enactive account, however, faces an 'ungrounding' problem (Rączaszek-Leonardi, Nomikou, Rohlfing, & Deacon 2018): how do embodied and situated communicative actions get symbolic properties that seem to be context-invariant? Here I shall not attempt to provide a solution to this problem, but instead propose a way of understanding it such that an ecological-enactive solution is not in principle impossible.

First of all, it is important to realise that, no matter how abstract our language becomes, it is never totally independent of any context. As I say in §3.4.1 (p. 105):

> Although it is certainly the case that we can use words in many *different* contexts, this only shows our ability to recontextualise. And this ability to *re*contextualise does not entail a decontextualised meaning. To be able to walk on many different surfaces does not entail an ability to walk *in abstracto*; similarly, being able to use a word in different or novel contexts does not entail an ability to use that word in abstracto, that is, it does not require knowledge of a decontextualised 'meaning'.

Viewed in this way, the problem is not to account for genuine context-invariant symbols, but rather for our ability to transcend contexts. In other words, what we need is an account of the skills of abstraction and generalization, rather than of the constitution of abstract objects. This comes to the fore in my discussion of content in Chapter 4. Traditionally, contents are seen as abstract objects, not bound by time or place. In contradistinction to this idea, I stress that our content-involving practices should be understood in terms of the context-dependent negotiation of correctness conditions.

In Chapter 1, I argued that there is no fact of the matter whether a linguistic unit is a repetition of an earlier unit, for no two events will ever be exactly the same. In an important sense, whether something counts as a repetition is up to the participants in a conversation to determine. This also holds in the problem of context-transcendence we are faced with here. Whether a given term can be applied to a novel context, and *counts* as being the same from the first context to another, must be determined by participants themselves. Viewed in this way, context-invariance cannot be decided by empirical means, but must rather be instituted in normative practices (see also Lance 2017, §6).

In relation to repetition and context-inpendence, written language plays an important role. It has a profound infuence on our understanding and experience of language. Love (2004, p. 542) explains:

> the very fact that what is said can be written down and, conversely, that what is written can be read aloud, fosters the idea that spoken and written counterparts instantiate some third thing, more abstract than either of them—i.e. the item(s) in the medium-neutral code itself.

Whereas spoken language consists in transient spatiotemporal patterns, written language consists of spatial patterns that remain the same from one moment to the next. It therefore stands to reason that this aspect of our sociomaterial practices proves to be an important explanatory factor in our practices that aim to

transcend context-dependence. For example, written language enabled the production of dictionaries. This in turn meant that *the meaning of a word* became something that could be 'looked up', and therefore was easily conceptualised as existing independently of the actual use of a word (Ong 1982). Although I have mentioned and briefly discussed the important role of written language in this thesis, much more can be said about this.

References

Adams, F. & Aizawa, K. (2008). *The Bounds of Cognition*. London: Wiley-Blackwell.

Adams, F. & Aizawa, K. (2010). The value of cognitivism in thinking about extended cognition, *Phenomenology and the Cognitive Sciences*, 9, 579–603. doi.org/10.1007/s11097-010-9184-9

Adolph, K.E., & Kretch, K.S. (2015). Gibson's Theory of Perceptual Learning. *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*, 127–134. doi.org/10.1016/B978-0-08-097086-8.23096-1

Agha, A. (2007). *Language and Social Relations*. Cambridge, MA: Cambridge University Press.

Anderson, M.L. (2003). Embodied cognition: A field guide. *Artificial Intelligence*, 149(1), 91–130. doi.org/10.1016/S0004-3702(03)00054-7

Anderson, M.L. (2014). *After Phrenology: Neural Reuse and the Interactive Brain.* Cambridge, MA.: The MIT Press.

Anderson, M.L. (2016). Précis of After Phrenology: Neural Reuse and the Interactive Brain. Behavioral and Brain Sciences, 39, 1–45. doi.org/10.1017/S0140525X15000631

Aizawa, K. (2015). What is this cognition that is supposed to be embodied? *Philosophical Psychology*, 28(6), 755–775. doi.org/10.1080/09515089.2013.875280

Alksnis, N. (2015). A Dilemma or a Challenge? Assessing the All-star Team in a Wider Context, *Philosophia*, *online first*. doi.org/10.1007/s11406-015-9618-2

Athanasopoulos, P., Wiggett, A., Dering, B., Kuipers, J., & Thierry, G. (2009). The Whorfian mind. Electrophysiological evidence that language shapes perception. *Journal Communicative & Integrative Biology,* 2(4), 332–334. doi.org/10.4161/cib.2.4.8400

Aristotle (1975). *Categories* and *De Interpretatione.* J.L. Ackrill (Trans.). Oxford: Clarendon Press.

Baggs, E. (2015). A radical empiricist theory of speaking: Linguistic meaning without conventions. Ecological Psychology, 27(3), 251–264. doi.org/10.1080/10407413.2015.1068655

Baggs, E. & Chemero, A. (2018). Radical embodiment in two directions, *Synthese*, *online first.* doi.org/10.1007/s11229-018-02020-9

Bakhtin, M.M. (1982). *The Dialogical Imagination* (C. Emerson & M. Holquist, Trans.). University of Texas Press.

Barber, A. (2003) Introduction. In A. Barber (Ed.) *Epistemology of Language*, 1–43. Oxford: Oxford University Press.

Barsalou, L.W. (1999). Perceptual symbol systems, *Behavioral and Brain Sciences*, 22, 577–660. doi.org/10.1017/S0140525X99002149

Berlin, B., & Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution*. California: University of California Press.

Bialystok, E. & Ryan, E.B. (1985). Toward a Definition of Metalinguistic Skill, *Merrill-Palmer Quarterly*, 31(3), 229–251.

Bingham, G.P. (1988) Task-specific devices and the perceptual bottleneck. *Human Movement Science*, 2(2–4), 225–264. doi.org/10.1016/0167-9457(88)90013-9

Block, N. (1986). Advertisement for a Semantics for Psychology, *Midwest Studies in Philosophy,* 10(1), 615–678. doi.org/10.1111/j.1475-4975.1987.tb00558.x

Bottineau, D. (2010). Language and enaction. In J. Stewart, O. Gapenne, & E. Di Paolo (Eds.) *Enaction: Towards a New Paradigm for Cognitive Science*, 267–306. Cambridge, MA: The MIT Press.

Boutonnet, B., Dering, B., Viñas-Guasch, N., & Thierry, G. (2013). Seeing Objects through the Language Glass. *Journal of Cognitive Neuroscience*, 25(10), 1702–1710. doi.org/10.1162/jocn_a_00415

Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence*, 47, 139 – 159. doi.org/10.1016/0004-3702(91)90053-M

Brown, B. (2006). Bootstrapping Norms: From Cause to Intention. In: S. Sherwin and P. Schotch (Eds.) *Engaged Philosophy: Essays in Honour of David Braybrooke*, pp. 343–364. University of Toronto Press.

Bruineberg, J., Kiverstein, J., & Rietveld, E. (2016). The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective. *Synthese*, 1–28. doi.org/10.1007/s11229-016-1239-1

Bruineberg, J. (2018). *Anticipating Affordances* (Doctoral dissertation).

Bruineberg, J., & Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances, *Frontiers in Human Neuroscience*, 8(599), 1–14. doi.org/10.3389/fnhum.2014.00599

Bruner, J. (1990). *Acts of Meaning*. Cambridge, MA: Harvard University Press.

Burge, T. (2010). *Origins of Objectivity.* Oxford: Clarendon Press.

Cairns, H. & Cairns, C. (1976) *Psycholinguistics,* New York: Holt, Rinehart & Winston.

Cappelen, H. & Lepore, E. (2007). *Language Turned On Itself The Semantics and Pragmatics of Metalinguistic Discourse*. Oxford: Oxford University Press.

Cash, M. (2008). The normativity Problem: Evolution and Naturalized Semantics, *The Journal of Mind and Behavior*, 29(1&2), 99–138.

Cazden, C.B. (1974). Play with language and metalinguistic awareness: one dimension of language experience, *International Journal of Early Childhood,* 6(12), 12–23. doi.org/10.1007/BF03176836

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

Chomsky, N. (1965). *Aspects of the Theory of Syntax.* Cambridge, MA.: The MIT Press.

Christiansen, M.H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioural and Brain Sciences*, 1–27. doi.org/10.1017/S0140525X1500031X

Clark, A. (1996). Linguistic anchors in the sea of thought? *Pragmatics & Cognition*, 4(1), 93–103. doi.org/10.1075/pc.4.1.09cla

Clark, A. (1997). *Being There*. Cambridge, Ma.: The MIT Press.

Clark, A. (2006). Language, embodiment, and the cognitive niche. *Trends in Cognitive Sciences*, 10(8), 370-374. doi.org/10.1016/j.tics.2006.06.012

Clark, A. & Chalmers, D.J. (1998). The extended mind, *Analysis,* 58(1), 7–19. doi.org/10.1093/analys/58.1.7

Clark, A., & Toribio, J. (1994). Doing without representing? *Synthese*, 101, 401–431. doi.org/10.1007/BF01063896

Clark, E.V. (1981). Acquisitional principles in lexical development. In: S.A. Gelman & J.P. Byrnes (Eds.) *Perspectives on language and thought Interrelations in development,* 31–71. Cambridge: Cambridge University Press.

Clark, E. V. (2009). *First language acquisition. Second edition*. Cambridge, MA: Cambridge University Press.

Cowley, S.J. (2011a). Distributed language. In: Cowley, S. (Ed.), *Distributed Language, (*pp. 185–210). Amsterdam/Philadelphia: John Benjamins.

Cowley, S. J. (2011b). Taking a language stance. *Ecological Psychology*, 23, 1–25. doi.org/10.1080/10407413.2011.591272

Crane, T. & French, C. (2017). The Problem of Perception. In Edward N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy (Spring 2017 Edition).* URL = <plato.stanford.edu/archives/spr2017/entries/perception-problem/>

Cruse, D.A. (2017). The Lexicon. In: M. Aronoff & J. Rees-Miller (Eds.) *The Handbook of Linguistics* 2nd Edition, 235–254. Oxford: Wiley Blackwell.

Cuffari, E.C., Di Paolo, E., & De Jaegher, H. (2015). From participatory sense-making to language: there and back again. *Phenomenology and the Cognitive Sciences,* 14(4), 1089–1125. doi.org/10.1007/s11097-014-9404-9

Davidoff, J., & Luzzatti, C. (2005). Language impairment and colour categories. Behavioral and Brain Sciences, 28(4), 494–495. doi.org/10.1017/S0140525X05280081/

Davidoff, J., & Roberson, D. (2004) Preserved thematic and impaired taxonomic categorisation: A case study. *Language and Cognitive Processes*, 19(1), 137–174. doi.org/10.1080/01690960344000125

Davidson, D., (1986). A nice derangement of epitaphs. In E. Lepore (Ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* (pp. 433–446). Oxford: Blackwell.

Davidson, D. (1992). The Second Person. *Midwest Studies in Philosophy*, 17, 255-267. doi.org/10.1111/j.1475-4975.1992.tb00154.x

Davidson, D. (1999). The Emergence of Thought. *Erkenntnis,* 51(1), 7–17. doi.org/10.1023/A:1005564223855

Davis, H.G. (2003) Introduction. In: H.G. Davis & T.J. Taylor (Eds.) *Rethinking Linguistics*, 1–16. London/New York: RoutledgeCurzon.

Davis, H.G., Taylor, T.J. (Eds.). (1990). *Redefining Linguistics.* London/New York: Routledge.

Davis, H.G., Taylor, T.J. (Eds.). (2003). *Rethinking Linguistics*. London/New York: RoutledgeCurzon.

De Jaegher, H., Di Paolo, P., & Gallagher S., (2010). Can social interaction constitute social cognition?, *Trends in Cognitive Sciences*, 14(10), 441–448. doi.org/10.1016/j.tics.2010.06.009

De Saussure, F. (2011). *Cours de linguistique generale* (W. Baskin, Trans.). New York: Columbia University Press. (Original work published in 1922).

Degenaar, J., & Myin, E. (2014). Representation hunger reconsidered. *Synthese,* 191, 3639–3648. http://doi.org/10.1007/s11229-014-0484-4

Devitt, M. (2006). *Ignorance of Language*. Oxford: Clarendon Press.

De Haan, S., Rietveld, E., Stokhof, M., & Denys, D. (2013) The phenomenology of Deep Brain Stimulation-induced changes in Obsessive-Compulsive Disorder patients: An enactive affordance-based model. *Frontiers in Human Neuroscience,* 7: 1–14. doi.org/10.3389/fnhum.2013.00653

Dingemanse, M., Roberts, S.G., Baranova, J., Blythe, J., Drew, P., Floyd, S., Gisladottir, R.S., Kendrick, K.H., Levinson, S.C., Manrique, E., Rossi, R., & Enfield, N.J. (2015). Universal Principles in the Repair of Communication Problems, *PLoS ONE,* 10(9), 1–15. doi.org/10.1371/journal.pone.0136100

Di Paolo, E. (2005) Autopoiesis, adaptivity, teleology, agency, *Phenomenology and the Cognitive Sciences*, 4, 429-452.

Di Paolo, E. (2016). Participatory Object Perception, *Journal of Consciousness Studies*, 23(5–6), 228–58.

Di Paolo, E., Cuffari, E.C., & De Jaegher, H. (2018). *Linguistic Bodies: The Continuity Between Life and Language.* Cambridge, Ma.: The Mit Press.

Di Paolo, E., Rohde, M., & De Jaegher, H. (2010). Horizons for the Enactive Mind: Values, Social Interaction, and Play. In: J. Stewart, O. Gapenne, & E. Di Paolo (Eds.), *Enaction: Towards a New Paradigm for Cognitive Science*, 33–88. Cambridge, MA.: The MIT Press.

di Pellegrino, G., Fadiga, L.., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding Motor Events: A Neurophysiological Study, *Experimental Brain Research,* 91, 176–80. doi.org/10.1007/BF00230027

Dretske, F. (1981). *Knowledge and the Flow of Information.* Cambridge, MA.: The MIT Press.

Drivonikou, G.V., Kay, P., Regier, T., Ivry, R.B., Gilbert, A.L., Franklin, A. & Davies, I.R.L., (2007). Further evidence that Whorfian effects are stronger in the right visual field than the left. *PNAS,* 104(3), 1097–1102. doi.org/10.1073/pnas.0610132104

Dummett, M. (1975). Wang's Paradox. *Synthese*, 3(4), 201–232. doi.org/10.1007/BF00485048

Dummett, M. (1978/1993). What do I know when I know a language? In: M. Dummett (Ed.) *The seas of Language,* 94–105. Oxford: Oxford University Press.

Dummett, M. (1993). *The seas of language*. Oxford: Oxford University Press.

Edwards, D. (1997). *Discourse and Cognition*. London: SAGE Publications.

Enfield, N.J. & Sidnell, J. (2017). *The Concept of Action*. Cambridge University Press.

Engel, A.K., Maye, A., Kurthen, M., & König, P. (2013). Where's the action? The pragmatic turn in cognitive science. *Trends in Cognitive Sciences*, 17(5), 202–209. doi.org/10.1016/j.tics.2013.03.006

Everett, D.L. (2009). *Don't Sleep, There are Snakes: Life and Language in the Amazonian Jungle.* New York: Vintage Departures.

Feldman, C.F. (1991). Oral Metalanguage. In: D.R. Olson & N. Torrance (Eds.), *Literacy and Orality,* 47–65. Cambridge University Press.

Fenson, L., Dale, P.S., Reznick, S.J., Bates, E., Thal, D.J., and Pethick, S.J. (1994). Variability in early communicative development. *Monographs of the Society for research in Child Development, 59*, (5, serial No. 242).

Fodor, J.A. (1975). *The Language of Thought*. Cambridge, MA: Harvard University Press.

Fodor, J.A. (1981). The Mind-Body Problem, *Scientific American*, 244, 114–125.

Fodor, J.A. (1984). Semantics, Wisconsin Style. *Synthese*, 59(3), 231–250. doi.org/10.1007/BF00869335

Fodor, J.A. (1990). *A Theory of Content and Other Essays.* Cambridge, MA.: The MIT Press.

Fodor, J. A. (2008). LOT 2: The language of thought revisited. Oxford: Oxford University Press.

Franklin, A., Drivonikou, G.V., Clifford, A., Kay, P., Regier, T., Davies, I.R.L. (2008). Lateralization of categorical perception of color changes with color term acquisition. *PNAS*, 105(47), 18221–18225. doi.org/10.1073/pnas.0809952105

Franklin, A., Drivonikou, G.V., Bevis, L., Davies, I.R.L., Kay, P., Regier, T. (2008). Categorical perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults. *PNAS*, 105(9), 3221–3225. doi.org/10.1073/pnas.0712286105

Frijda, N. H. (2007). *The laws of emotion*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Fuchs, T. (2012). The phenomenology and development of social perspectives. *Phenomenology and the Cognitive Sciences*, 1–29. doi.org/10.1007/s11097-012-9267-x

Gallagher, S. (2015). Seeing Things in the Right Way: How Social Interaction Shapes Perception. In D. Maxime and Th. Breyer (Eds.) *Normativity in Perception* (pp. 117–127). London: Palgrave Macmillan.

Gibson, E.J. (1953). Improvement in perceptual judgments as a function of controlled practice or training. *Psychological Bulletin*, 50, 401–431.

Gibson, E.J. (1963). Perceptual learning, *Annual Review of Psychology*, 14, 29–59.

Gibson, E.J. (1969). *Principles of Perceptual Learning and Development*. Appleton-Century Crofts, New York.

Gibson, E.J. (1988). Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge, *Annual Review of Psychology,* 39, 1–41. doi.org/10.1146/annurev.ps.39.020188.000245

Gibson, E.J. (2000). Perceptual Learning in Development: Some Basic Concepts, Ecological Psychology. 12(4), 295–302. doi.org/10.1207/S15326969ECO1204_04

Gibson, E.J. & Rader, N. (1979). Attention – The perceiver as performer. In G. Hale (Ed) Attention and Cognitive Development (pp. 1–21). New York: Plenum Press.

Gibson, E.J., & Spelke, E.S. (1983). The development of perception. In: P. Mussen, J.H. Flavell, E. Markman (Eds.), *Handbook of Child Psychology, Vol 3*. New York: Wiley.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston, MA: Houghton Mifflin.

Gibson, J.J. (1969). *The Senses Considered as Perceptual Systems.* London: George Allen & Unwin Ltd.

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.

Gibson, J.J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton-Mifflin.

Gibson, J.J., Gibson, E.J. (1955). Perceptual learning: differentiation or enrichment? *Psychological Review,* 62, 32–41. doi.org/10.1037/h0048826

Gilbert, A.L., Regier, T., Kay, P., & Ivry, R.B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *PNAS*, 103(2), 489–494. doi.org/10.1073/pnas.0509868103

Gilbert, A.L., Regier, T., Kay, P., & Ivry, R.B. (2008). Support for lateralization of the Whorf effect beyond the realm of color discrimination. *Brain and Language*, 105(2), 91–98. doi.org/10.1016/j.bandl.2007.06.001

Goldstein, L. (2004) Philosophical Integrations, *Language Sciences*, 26, 545–563. doi.org/10.1016/j.langsci.2004.09.004

Goldstein, M.H., King, A.P., & West, M.J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *PNAS*, 100(12), 8030–8035. doi.org/10.1073/pnas.1332441100

Gombert, J.E. (1992). Metalinguistic Development. University of Chicago Press.

Goody, J. (1977). *The domestication of the savage mind*. Cambridge University Press.

Hacker, P.M.S. (2014). Two Conceptions of Language. *Erkenntnis*, 79, 1271–1288. doi.org/10.1007/s10670-013-9558-9

Hakes, D.T. (1980). *The Development of Metalinguistic Abilities in Children.* Berlin: Springer-Verlag.

Hanna, P, (2006). Swimming and Speaking Spanish. *Philosophia,* 34, 267–285. doi.org/10.1007/s11406-006-9031-y

Harman, G. (1973). *Thought.* Princeton: Princeton University Press.

Harnad, S. (1990). The Symbol Grounding Problem, *Physica D: Nonlinear Phenomena,* 42(1–3), 335–346. doi.org/10.1016/0167-2789(90)90087-6

Harris, R. (1980). *The Language-Makers*. London: Duckworth.

Harris, R. (1990). On Redefining Linguistics. In H. Davis & T.J. Taylor (Eds.), *Redefining Linguistics.* London/New York: Routledge.

Harris, R. (1996) *The Language Connection.* Bristol: Thoemmes Press.

Harris, R. (1998). *Introduction to Integrational Linguistics*. Pergamon, Oxford.

Harris, R. (2000). *Rethinking Writing*. London/New York: Continuum.

Harris, R. (2001) *The Language Myth in Western Europe*. Richmond: Curzon.

Harris, R. (2003) On redefining linguistics. In: H.G. Davis & T.J. Taylor (Eds.) *Rethinking Linguistics*, 17–68. London/New York: RoutledgeCurzon.

Harris, R. (2004) Integrationism, language, mind and world, *Language Sciences,* 26, 727–739. doi.org/10.1016/j.langsci.2004.09.013

Harris, R. (2009) *Integrationist Notes and Papers. 2006–2008*. Gamlingay: Bright Pen.

Harvey, M.I. (2015). Content in languaging: why radical enactivism is incompatible with representational theories of language. *Language Sciences,* 48, 90–129. doi.org/10.1016/j.langsci.2014.12.004

Haugeland, J. (1990). The intentionality all-stars. *Philosophical Perspectives*, 4, 383–427. doi.org/10.2307/2214199

Haugeland, J. (2013). *Dasein Disclosed*. Cambridge, MA.: The Harvard University Press.

Heras-Escribano, M. (2019). Pragmatism, enactivism, and ecological psychology: towards a unified approach to post-cognitivism, *Synthese*, online first. doi.org/10.1007/s11229-019-02111-1

Heras-Escribano, M., Noble, J., & De Pinedo, M. (2013) The only wrong cell is the dead one: On the enactive approach to normativity. In: P. Liò et al. (Eds.) *Advances in Artificial Life*, 665–670. Cambridge, MA.: The MIT Press.

Heras-Escribano, M. & De Pinedo, M. (2015) Are affordances normative?, *Phenomenology and the Cognitive Sciences*, online first, 1–25. doi.org/10.1007/s11097-015-9440-0

Heras-Escribano, M. (2019). Pragmatism, enactivism, and ecological psychology: towards a unified approach to post-cognitivism. *Synthese*, online first, 1–27. doi.org/10.1007/s11229-019-02111-1

Hobbes, T. (1989). *Leviathan*, M. Oakeshott (Ed.). Oxford: Blackwell. (original published in 1651).

Hurley, S. (1998). *Consciousness in Action*. Cambridge, MA.:Harvard University Press.

Hurley, S. (2001). Perception and Action: Alternative Views, *Synthese*, 129, 3–40. doi.org/10.1023/A:1012643006930

Hutto, D. D. (2006). Both Bradley and biology: Reply to Rudd. In R. Menary (Ed.) *Radical Enactivism: Intentionality, Phenomenology and Narrative*, 81–105. Amsterdam/Philadelphia: John Benjamins.

Hutto, D. D. (2008). *Folk Psychological Narratives: The Sociocultural Basis of Understanding Reasons*. Cambridge, Ma.: The MIT Press.

Hutto, D.D. (2015). REC: Revolution Effected by Clarification, *Topoi,* online first. doi.org/10.1007/s11245-015-9358-8

Hutto, D. D. & Myin, E. (2013). *Radicalizing Enactivism. Basic Minds without Content.* Cambridge, Ma.: The MIT Press.

Hutto, D. D. & Myin, E. (2017). *Evolving Enactivism. Basic Minds meet Content.* Cambridge, Ma.: The MIT Press.

Hutto, D.D. & Satne, G. (2015). The Natural Origins of Content. *Philosophia,* 43(3), 521–536. doi.org/10.1007/s11406-015-9644-0

Ingold, T. (2001). From the transmission of representation to the education of attention. in: H. Whitehouse (Ed.) *The debated mind: Evolutionary psychology versus ethnography,* 113–153. New York, NY.: Berg.

Jackendoff, R. (1996). How language helps us thinks, *Pragmatics & Cognition,* 4(1), 1-34. doi.org/10.1075/pc.4.1.03jac

Jakobson, R. (1956/1985). Metalanguage as Linguistic Problem. In: S. Rudy (Ed.), *Roman Jakobson Selected Writings VII, Contributions to Comparative Mythology. Studies in Linguistics and Philology, 1972–1982,* 113–121. Berlin/New York/Amsterdam: Mouton Publishers.

Ježek, E. (2016). *The Lexicon: An Introduction*. Oxford: Oxford University Press.

Karmiloff-Smith, A. (1986). From meta-processes to conscious access: evidence from children's metalinguistic and repair data, *Cognition*, 23, 95–147.

Keijzer, F. (2001). *Representation and Behaviour.* The Mit Press.

Kiverstein, J. & Clark, A. (2009) Introduction: Mind Embodied, Embedded, Enacted: One Church or Many? *Topoi*, 28, 1–7. doi.org/10.1007/s11245-008-9041-4

Klaassen, P., Rietveld, E., & Topal, J. (2010). Inviting complementary perspectives on situated normativity in everyday life. *Phenomenology and the Cognitive Sciences,* 9, 53–73. doi.org/10.1007/s11097-009-9133-7

Kravchenko, A. V. (2007). Essential properties of language, or, why language is not a code. *Language Sciences*, 29 , 650–671. doi.org/10.1016/j.langsci.2007.01.004

Kravchenko, A.V. (2009). The Experiental Basis of Speech and Writing as Different Cognitive Domains. *Pragmatics & Cognition*, 17(3), 527–548. doi.org/10.1075/p&c.17.3.03kra

Kriegel, U. (2011). *The sources of intentionality*. Oxford: Oxford University Press.

Kukla, R. (2017). Ostension and Assertion. In: Z. Adams and J. Browning (Eds.) *Giving a Damn – Essays in Dialogue with John Haugeland,* pp. 103–130. Cambridge, MA.: The MIT Press.

Kukla, R. & Lance, M. (2014). Intersubjectivity and Receptive Experience. *The Southern Journal of Philosophy*, 52(1), 22–42. doi.org/10.1111/sjp.12047

Lakoff, G. & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.

Lance, M. (2017). Language Embodied and Embedded: Walking the Talk. In: Z. Adams & J. Browning (Eds.) *Giving a Damn,* 161–186. Cambridge, MA: The MIT Press.

Linell, P. (2005). *The written language bias in linguistics: Its nature, origins and transformations*. London: Routledge.

Lobo, L., Heras-Escribano, M., & Travieso, D. (2018). The History and Philosophy of Ecological Psychology, *Frontiers in Psychology*, 9, 1–15. doi.org/10.3389/fpsyg.2018.02228

Locke, J. (2004). *An Essay Concerning Human Understanding*. R. Woolhouse (Ed.). London: Penguin Books. (Original work published in 1690)

Love, N. (1990). The locus of languages in a redefined linguistics. In: H.G. Davis & T.J. Taylor (Eds.), *Redefining Linguistics* (pp. 53–118). London/New York: Routledge.

Love, N. (1995). On construing the world of language. In J. R. Taylor (Ed.) *Language and the Cognitive Construal of the World,* 377–390. Berlin: Mouton de Gruyter.

Love, N. (1999). Searle on language, *Language & Communication,* 19, 9–25. doi.org/10.1016/S0271-5309(98)00015-9

Love, N. (2003). Rethinking the fundamental assumption of linguistics. In: H.G. Davis & T.J. Taylor (Eds.) *Rethinking Linguistics*, 69–94. London/New York: RoutledgeCurzon.

Love, N. (2004). Cognition and the language myth. *Language Sciences*, 26 , 525–544. doi.org/10.1016/j.langsci.2004.09.003

Love, N. (2007). Are languages digital codes? *Language Sciences,* 29, 690–709. doi.org/10.1016/j.langsci.2007.01.008

Love, N. (2017). On languaging and languages. *Language Sciences*, 61, 113–147. doi.org/10.1016/j.langsci.2017.04.001

Lucy, J.A. (1993). General Introduction. In: J.A. Lucy (Ed.), *Reflexive language: Reported speech and metapragmatics*, 1–4. Cambridge University Press.

Lupyan, G. (2012). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Frontiers in Psychology*. doi.org/10.3389/fpsyg.2012.00054

Lupyan, G. (2016). The Centrality of Language in Human Cognition. *Language Learning*, 66(3), 516–553. doi.org/10.1111/lang.12155

LupyanG., & Spivey, M.J. (2010). Redundant spoken labels facilitate perception of multiple items. *Attention, Perception, & Psychophysics,* 72(8), 2236–2253. doi.org/10.3758/BF03196698

Lupyan, G. & Swingley, D. (2012). Self-directed speech affects visual search performance. *The Quarterly Journal of Experimental Psychology*, 65(6), 1068–1085. doi.org/10.1080/17470218.2011.647039

MacDonald, G. & Papineau, D. (2006). Introduction: Prospects and Problems for Teleosemantics. In: G. MacDonald & D. Papineau (Eds.), *Teleosemantics – New Philosophical Essays,* 1–22. Oxford: Clarendon Press.

McDaniels, D. and Smith Cairns, H. (1996). Eliciting Judgments of Grammaticality and Reference. In: D. McDaniels, McKee, & H. Smith Cairns (Eds.), *Methods for Assessing Children's Syntax Language, Speech, and Communication*, 233–254. Cambridge, Ma.: The MIT Press.

Maravita, A. & Iriki, A. (2004). Tools for the body (schema). *TRENDS in Cognitive Sciences*, 8(2), 79–86. doi.org/10.1016/j.tics.2003.12.008

Markman, E.M. (1981). The whole-object, taxonomic, and mutual exclusivity assumptions as initial constraints on word meanings. In: S. A. Gelman & J. P. Byrnes (Eds.). *Perspectives on language and thought Interrelations in development*, 72–107. Cambridge, MA: Cambridge University Press.

Marr, D. (2010). *Vision*. Cambridge, Ma.: The MIT Press. (original published in 1982).

Matthews, R., (2003). Does linguistic competence require knowledge of language? In: Barber, A. (Ed.), *Epistemology of Language*. Oxford University Press, Oxford, pp. 187–213.

Maturana, H.R. (1978). Biology of language: The epistemology of reality. In: Miller, G.A. and Lenneberg, E. (Eds.). *Psychology and Biology of Language and Thought,* 27–63. New York, NY: Academic Press.

Maturana, H.R. & Varela, F.J. (1980). *Autopoiesis and Cognition: The Realization of the Living*. Dordrecht/Boston/London: D. Reidel Publishing Company.

Maturana, H.R. & Varela. F.J. (1987). *The Tree of Knowledge: The Biological Roots of Human Understanding*. Boston: New Science Library.

Miłkowski, M. (2015). The Hard Problem Of Content: Solved (Long Ago). *Studies in Logic, Grammar and Rhetoric*, 41(54), 73--88. doi.org/10.1515/slgr-2015-0021

Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories*. Cambridge, Ma.: The MIT Press.

Millikan, R.G. (1998) Language Conventions Made Simple, *The Journal of Philosophy*, XCV(4), 161-180. doi.org/10.2307/2564683

Millikan, R. G. (2005). *Language: A Biological Model*. Oxford: Oxford University Press.

Mollo, D.M. (2015). Being Clear on Content - Commentary on Hutto and Satne, *Philosophia*, online first. doi.org/10.1007/s11406-015-9622-6

Myhill, D. & Jones, S. (2015). Conceptualizing metalinguistic understanding in writing, *Cultura y Educación*, 27(4), 839–867. doi.org/10.1080/11356405.2015.1089387

Myin, E. (2003). An account of color without a subject? *Behavioral and Brain Sciences*, 26, 42– 3. doi.org/10.1017/ S0140525X03440016

Myin, E. (2016). Perception as Something We Do, *Journal of Consciousness Studies*, 23(5-6) 80-104.

Neander, K. (2017). *A Mark of the Mental: In Defense of Informational Teleosemantics.* Cambridge: MIT Press.

Noë, A. (2004). *Action in Perception*. Cambrigde, Ma.: The MIT Press.

Noë, A. (2005) Against intellectualism, *Analysis*, 65(4), 278–290. doi.org/10.1093/analys/65.4.278

Noë, A. (2009) *Out of our Heads: Why You Are Not Your Brain, and Other Lessons from the Biology of Consciousness.* New York: Hill and Wang.

Noë, A. (2012). *Varieties of Presence*. Cambridge, MA.:Harvard University Press.

Noë, A. (2015). *Strange Tools*: Art and Human Nature. New York: Hill & Wang.

Ong, W.J. (1982). *Orality and Literacy.* London/New York: Routledge.

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

O'Regan, J.K. & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5), 883–917. doi.org/10.1017/S0140525X01000115

Pablé, A. & Hutton, C. (2015) *Semiotics, Communication and Cognition*. Boston/Berlin: De Gruyter Mouton

Papafragou, A. (2005). Relations between language and thought: Individuation and the count/mass distinction. In H. Cohen & C. Lefebvre (Eds.) *Handbook of Categorization in Cognitive Science* (pp. 256–272). Oxford: Elsevier.

Papineau, D. (1984). Representation and Explanation, *Philosophy of Science*, 51, 550–72.

Pearson, J. & Kosslyn, S.M. (2015). The heterogeneity of mental representation: Ending the imagery debate, *PNAS*, 112(33), 10089–10092. doi.org/10.1073/pnas.1504933112

Piccinini, G. (2008). Computation without Representation, *Philosophical Studies*, 137(2), 205–241. doi.org/10.1007/s11098-005-5385-4

Pilling, M., Wiggett, A., Özgen, E., & Davies, I.R.L. (2003). Is color "categorical perception" really perceptual? *Memory & Cognition,* 31(4), 538–551. doi.org/10.3758/BF03196095

Pinker, S. (1994). *The Language Instinct: How the Mind Creates Language*. New York: William Morrow and Company.

Pinker, S. (2007). *The Stuff of Thought: Language As a Window Into Human Nature*. London: Penguin.

Pitchford, N.J. & Mullen, K.T. (2002). Is the acquisition of basic-colour terms in young children constrained? *Perception*, 31, 1349–1370. doi.org/10.1068/p3405

Pitt, D. (2018). Mental Representation, *The Stanford Encyclopedia of Philosophy (Winter 2018 Edition)*. URL = <https://plato.stanford.edu/archives/win2018/entries/mental-representation/>.

Proffitt , D. R., & Linkenauger, S. A. (2013). Perception viewed as a phenotypic expression. In: W. Prinz, M. Beisert, & A. Herwig (Eds.) *Action science: Foundations of an Emerging Discipline,* 171 – 197. Cambridge, MA : MIT Press.

Putnam, H. (1975) The meaning of 'meaning', *Minnesota Studies in the Philosophy of Science,* 7, 131–193.

Putnam, H. (1992). *Renewing Philosophy*. Harvard University Press.

Quine, W. V. O. (1960). *Word and Object*. Cambridge, MA.: The MIT Press.

Rączaszek-Leonardi, J. (2009). Symbols as constraints: The structuring role of dynamics and self-organization in natural language. *Pragmatics & Cognition*, 17 (3), 653–676. doi.org/10.1075/pc.17.3.09ras

Rączaszek-Leonardi, J. (2011). Language as a system of replicable constraints. In H. Pattee (Ed.), Laws, Language and Life: Howard Pattee's classic papers on the physics of symbols, (pp. 295–333). Dordrecht: Springer.

Rączaszek-Leonardi, J. (2016) How does a word become a message? An illustration on a developmental time-scale. *New Ideas in Psychology*, 42, 46–55. doi.org/10.1016/j.newideapsych.2015.08.001

Rączaszek-Leonardi, J., & Kelso, J. A. S. (2008). Reconciling symbolic and dynamic aspects of language: Toward a dynamic psycholinguistics. *New Ideas in Psychology*, 26 (2), 193–207. doi.org/10.1016/j.newideapsych.2007.07.003

Rączaszek-Leonardi, J., Nomikou, I., & Rohlfing, K. J. (2013). Young children's dialogical actions: the beginnings of purposeful intersubjectivity. *IEEE Transactions on Autonomous Mental Development*, 5(3), 210–221. doi.org/10.1109/TAMD.2013.2273258.

Rączaszek-Leonardi, J., Nomikou, I., Rohlfing, K.J., & Deacon, T.W. (2018). Language Development From an Ecological Perspective: Ecologically Valid Ways to Abstract Symbols, *Ecological Psychology*, 30(1), 39–73. doi.org/10.1080/10407413.2017.1410387

Ramsey, W. M. (2007). *Representation Reconsidered*. Cambridge University Press.

Reddy, M.J. (1993). The conduit metaphor: A case of frame conflict in our language about language. In A. Ortony (Ed.), *Metaphor and Thought (2nd Edition)* (pp. 164–201). Cambridge, MA: Cambridge University Press.

Reed, E. S. (1995). The ecological approach to language development: a radical solution to Chomsky's and Quine's problems. *Language & Communication*, 15(1), 1–29. doi.org/10.1016/0271-5309(94)E0010-9

Reed, E. S. (1996). *Encountering the world: Toward an ecological psychology (1st ed.)*. New York: Oxford University Press.

Regier, T. & Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in Cognitive Science*, 13(10), 439–446. doi.org/10.1016/j.tics.2009.07.001

Rescorla, M. (2016). Bayesian Sensorimotor Psychology. *Mind and Language,* 31(1), 3–36. doi.org/10.1111/mila.12093

Rietveld, E. & Kiverstein, J. (2014). A rich landscape of affordances. *Ecological Psychology*, 26(4), 325–352. doi.org/10.1080/10407413.2014.958035

Rietveld, E., Denys, D., & van Westen, M. (2018). Ecological-Enactive Cognition as engaging with a field of relevant affordances: The Skilled Intentionality Framework (SIF). In: A. Newen, L. de Bruin, & S. Gallagher (Eds.) *The Oxford Handbook of 4E Cognition*. doi.org/10.1093/oxfordhb/9780198735410.013.3

Roberson, D. & Davidoff, J. (2000). The categorical perception of colors and facial expressions- the effect of verbal interference. *Memory & Cognition*, 28(6), 977–986. doi.org/10.3758/BF03209345

Roberson, D., Davidoff, J., Davies, I.R.L., & Shapiro, L.R. (2005). The Development of Color Categories in Two Languages: A Longitudinal Study. *Journal of Experimental Psychology: General*, 133(4), 554–571. doi.org/10.1037/0096-3445.133.4.554

Roberson, D., Davies, I.R.L., & Davidoff, J., (2000). Color Categories Are Not Universal: Replications and New Evidence From a Stone-Age Culture. *Journal of Experimental Psychology: General,* 129(3), 369–398. doi.org/10.1037/0096-3445.129.3.369

Roberts, L.D. (2002). How words have content: an explanatory hypothesis based on developmental psychology, *Language and Communication,* 22, 83–106.

Rosaldo, M. (1973). I have nothing to hide: The language of Ilongot oratory, *Language in Society*, 2, 193–223.

Rosaldo, M. (1982). The things we do with words: Ilongot speech acts and speech act theory in philosophy, *Language in Society*, 2, 203–237.

Rosch Heider, E., & Olivier, D. C. (1972). The structure of the color space in naming and memory for two languages. *Cognitive Psychology*, 3, 337–354.

Rosenberg, A. (2015). The Genealogy of Content or the Future of an Illusion, *Philosophia*, online first. doi.org/10.1007/s11406-015-9624-4

Rowlands, M. (2015). Hard Problems of Intentionality, *Philosophia,* online first. doi.org/10.1007/s11406-015-9626-2

Rowlands, M. (2017). Arguing about representation, *Synthese*, 194, 4215–4232. doi.org/10.1007/s11229-014-0646-4.

Runeson, S. (1977).On the possibility of "smart" perceptual mechanisms. *Scandinavian Journal of Psychology*, 18, 172–179.

Ryle, G. (1945a). Knowing how and knowing that. *Proceedings of the Aristotelian Society, New Series*, 46, 1–16.

Ryle, G. (1945b) Philosophical Arguments. In G. Ryle (2009) *Collected Essays 1929-1968*, 203-221. New York, NY: Routledge.

Ryle, G. (1949) *The Concept of Mind*. University of Chicago Press.

Ryle, G. (1962) Abstractions. In: G. Ryle (2009) *Collected Essays 1929-1968*, 448–458. New York, NY: Routledge.

Ryle, G. (2009). *The Concept of Mind* 60th anniversary edition. J. Tanney (Ed.). London/New York: Routledge. (Original work published in 1949)

Searle, J.R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.

Searle, J.R. (1995). T*he Construction of Social Reality*. London: Allen Lane (The Penguin Press).

Sellars, W. (1954). Some Reflections on Language Games, *Philosophy of Science*, 21(3), 204–228.

Sellars, W. (1971/1975). The Structure of Knowledge. In H. Castañeda (Ed.) *Action, Knowledge, and Reality. Critical Studies in Honor of Wilfrid Sellars*, 295–348. Indianapolis, IN: The Bobbs-Merrill Company.

Shanker, S.G. & Taylor, T.J. (2001) The House that Bruner Built. In: Bakhurst, D., Shanker, S.G. (Eds.) *Jerome Bruner: Language, Culture, Self,* 50–70. Sage: London.

Shapiro, L.A. (2014). Radicalizing Enactivism: Basic Minds without Content, by Daniel D. Hutto and Erik Myin, *Mind*, 123(489), 213–220. doi.org/10.1093/mind/fzu033

Shea, N. (2013). Naturalising Representational Content. *Philosophy Compass*, 8: 496–509. doi.org/10.1111/phc3.12033

Simpson, D. (2010). Language and know-how. *Phenomenology and Cognitive Science*, 9, 629–643. doi.org/10.1007/s11097-010-9183-x

Smith Cairns, H. (2015). Metalinguistic skills of children. In: M. Becker, J. Grinstead, & J. Rothman (Eds.) *Generative Linguistics and Acquisition*, pp. 271–290. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Snow, J.C., Pettypiece, C.E., McAdam, T.D., McLean, A.D. Stroman, P.W., Goodale, M.A., & Culham, J.C. (2011). Bringing the real world into the fMRI scanner: Repetition effects for pictures versus real objects, *Scientific Reports*, 1(130), 1–10. doi.org/10.1038/srep00130

Stanley, J. & Williamson, T. (2001) Knowing how, *The Journal of Philosophy,* 97, 411–444.

Stewart, J. (2010). Foundational issues in enaction as a paradigm for cognitive science: From the origin of life to consciousness and writing. In J. Stewart, O. Gapenne, & E. Di Paolo (Eds.), *Enaction: Towards a New Paradigm for Cognitive Science* (pp. 1–32). Cambridge, MA: The MIT Press.

Stewart, J., Gapenne, O. & Di Paolo, E. (Eds.). *Enaction: Towards a New Paradigm for Cognitive Science*. Cambridge, MA.: The MIT Press.

Suegami, T., Aminihajibashi, S., & Laeng, B. (2014). Another look at category effects on colour perception and lateralisation. *Cognitive Processing*, 15(2), 217–226. doi.org/10.1007/s10339-013-0595-8

Swenson, R. (1992). Autocatakinetics, yes – autopoiesis, no: Steps towards a unified theory of evolutionary ordering. *International Journal of General Systems,* 21(2), 207–228. doi.org/10.1080/03081079208945072

Tanney, J. (2009). Foreword. In: G. Ryle *Collected Papers Volume 2. Collected Essays 1929–1968*. London/New York: Routledge.

Tanney, J. (2015) Gilbert Ryle. *The Stanford Encyclopedia of Philosophy (Spring 2015 Edition)*, Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2015/entries/ryle/>.

Taylor, C. (2016). *The Language Animal. The Full Shape of the Human Linguistic Capacity*. Cambridge, Ma./London: The Belknap Press of Harvard University Press.

Taylor, T. J. (1986) Do you understand? Criteria of Understanding in Verbal Interaction, Language and Communication, 6(3), 171-180.

Taylor, T.J. (1990). Normativity and linguistic form. In H.G. Davis & T.J. Taylor (Eds.), *Redefining Linguistics (*pp. 118–148). London/New York: Routledge.

Taylor, T.J. (1992). *Mutual Misunderstanding: Scepticism and the Theorizing of Language and Interpretation*. Durham, NC/London: Duke University Press/Routledge.

Taylor, T.J. (2000). Language constructing language: the implications of reflexivity for linguistic theory. *Language Sciences*, 22, 483–499. doi.org/10.1016/S0388-0001(00)00016-4

Taylor, T.J. (2010). Where does language come from? The role of reflexive enculturation in language development. *Language Sciences,* 32, 14–27. doi.org/10.1016/j.langsci.2008.12.014

Taylor, T.J. (2011) Language development and the integrationist, *Language Sciences*, 33, 579–583. doi.org/10.1016/j.langsci.2011.04.029

Taylor, T.J. (2012) Understanding others and understanding language: how do children do it?, *Language Sciences,* 34, 1-12. doi.org/10.1016/j.langsci.2011.07.001

Taylor, T.J. (2013). Calibrating the child for language: Meredith Williams on a Wittgensteinian approach to language socialization. *Language Sciences*, 40, 308–320. doi.org/10.1016/j.langsci.2013.07.002

Taylor, T.J., & Shanker, S. (2003) Rethinking language acquisition: what the child learns. In: H.G. Davis & T.J. Taylor (Eds.) *Rethinking Linguistics*, 151–170. London/New York: RoutledgeCurzon.

Thelen, E., & Smith, L.B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press.

Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Harvard University Press.

Thierry, G., Athanasopoulos, P., Wigget, A., Dering, B., and Kuipers, J. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *PNAS*, 106(11), 4567–4570. doi.org/10.1073/pnas.0811155106

Toolan, M. (1996) *Total Speech: An Integrational Approach to Language*. Durham/London: Duke University Press.

Turvey, M. T., Shaw, R., Reed, E., & Mace, W. (1981). Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn. *Cognition*, 9, 237–304.

Van den Herik, J.C. (2013). Is Information Out There? *Erasmus Student Journal of Philosophy,* 5, 20–33.

van den Herik, J.C. (2014). *Why Radical Enactivism is not radical enough: A case for really Radical Enactivism*. Unpublished master's thesis. Faculty of Philosophy, University of Rotterdam.

Van den Herik, J.C. (2017). Linguistic know-how and the orders of language. *Language Sciences*, 61, 17–27. doi.org/10.1016/j.langsci.2016.09.009

Van den Herik, J.C. (2018). Attentional Actions – An Ecological-Enactive Account of Utterances of Concrete Words, *Psychology of Language and Communication*, 22(1), 90–123. doi.org/10.2478/plc-2018-0005

Van den Herik, J.C. (under review). *On the Way to Solving the Hard Problem of Content.*

van Dijk, L. (2016). Laying down a path in talking, *Philosophical Psychology*, 29(7), 993–1003. doi.org/10.1080/09515089.2016.1213379

Van Dijk, L., Withagen, R., & Bongers, R.M. (2015). Information without content: A Gibsonian reply to enactivists' worries. *Cognition*, 134, 210–214. doi.org/10.1016/j.cognition.2014.10.012

Van Elk, M., Slors, M., & Bekkering, H. (2010). Embodied language comprehension requires an enactivist paradigm of cognition. *Frontiers in Psychology*, 1 (234), 1–9. doi.org/10.3389/fpsyg.2010.00234

Van Gelder, T. (1998). The dynamical hypothesis in cognitive science. Behavioral and Brain Sciences, 21, 615–628.

Van Kleeck, A. (1982). The Emergence of Linguistic Awareness: A Cognitive Framework, *Merrill-Palmer Quarterly*, 28(2), 237–265.

Varela, F.J., Thompson, E., Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA.: The MIT Press.

Verbrugge, R. R. (1985). Language and event perception: steps toward a synthesis. In W. H. Warren, & R. E. Shaw (Eds.), *Persistence and change. Proceedings of the first international conference on event perception* (pp. 157–194). Hillsdale, NJ: Lawrence Earlbaum Associates.

Verschueren, J. (2000). Notes on the Role of Metapragmatic Awareness in Language Use, *Pragmatics,* 10(4), 439–456. doi.org/10.1075/prag.10.4.02ver

Villalobos, M. & Dewhurst, J. (2017). Why post-cognitivism does not (necessarily) entail anti-computationalism, *Adaptive Behavior*, 23(3), 117–128. doi.org/10.1177/1059712317710496

Voloshinov, V.N. (1930/1973). *Marxism and the Philosophy of Language* (L. Matejka & I.R. Titunik, Trans.). New York/London: Seminar Press.

Von Foerster, H. & Poerksen, B. (2002). *Understanding Systems: Conversations on Epistemology and Ethics* (K. Leube (Trans.)). New York: Kluwer Academic/Plenum Publishers.

Ward, D., Silverman, D., & Villalobos, M. (2017). Introduction: The Varieties of Enactivism, *Topoi*, 36, 365–375. doi.org/10.1007/s11245-017-9484-6

Whitehead, A.N. (1929). *Process and reality.* New York: Harper.

Whorf, B.L. (1956). *Language, Thought and Reality: Selected Writings of Benjamin Lee Whorf,* J. Carroll (Ed). Cambridge, Ma.: The M.I.T. Press.

Williams, M. (1994). The significance of learning in Wittgenstein's later philosophy. *Canadian Journal of Philosophy*, 24, 173–203.

Williams, M. (1999). *Wittgenstein, Mind and Meaning: Towards a Social Conception of Mind*. London and New York: Routledge.

Williams, M., (2010a). *Blind Obedience: The Structure and Content of Wittgenstein's Later Philosophy*. London and New York: Routledge.

Williams, M. (2010b). Normative Naturalism. *International Journal of Philosophical Studies*, 18(3), 355–375. doi.org/10.1080/09672559.2010.492115

Wilson, A.D., & Golonka, S. (2013). Embodied cognition is not what you think it is. *Frontiers in Psychology*, 4, 1–13. doi.org/10.3389/fpsyg.2013.00058

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4), 625–636. doi.org/10.3758/BF03196322

Winawer, J., Witthoft, N., Frank, M.C., Wu, L., Wade, A.R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *PNAS*, 104(19), 7780–7785. doi.org/10.1073/pnas.0701644104

Withagen, R., de Poel, H. J., Araújo, D., & Pepping, G. J. (2012). Affordances can invite behavior: Reconsidering the relationship between affordances and agency. *New Ideas in Psychology*, 30, 250–258. doi.org/10.1016/j.newideapsych.2011.12.003

Wittgenstein, L. (1967). *Zettel*. Oxford: Blackwell.

Wittgenstein, L. (1983) *Remarks on the Foundations of Mathematics*. Oxford: Blackwell.

Wittgenstein, L. (1953/2009). *Philosophical Investigations* (Revised 4th ed.). Oxford: Wiley-Blackwell.

Wolf, G. & Love, N. (1993) Integrational linguistics: an introductory survey. In: A. Crochetiere, J-C. Boulanger and C. Ouellon (Eds.) *Actes du XVe Congres International des Linguistes (Vol. 1)*, 313–320. Sainte Foy: Presses de l'Universite Laval.

Zhu, Q., & Bingham, G.P. (2010). Learning to perceive the affordance for long-distance throwing: smart mechanism or function learning? *Journal of Experimental Psychology: Human Perception and Performance*, 36(4), 862–875. doi.org/10.1037/a0018738

Zukow-Goldring, P. (2012). Assisted imitation: first steps in the seed model of language development. *Language Sciences,* 34, 569–582. doi.org/10.1016/j.langsci.2012.03.012

Zukow-Goldring, P. & Ferko, K.R. (1994). An ecological approach to the emergence of the lexicon: Socializing attention. In V. John-Steiner, C. Panofsky, & L. Smith (Eds.), *Sociocultural Approaches to Language and Literacy: An Interactionist Perspective* (pp. 170–194). Cambridge, MA: Cambridge University Press.

Samenvatting

Dit proefschrift ontwikkelt een perspectief op taal en haar ontwikkeling door voort te bouwen op twee benaderingen. De eerste is de *ecologisch-enactieve benadering* van cognitie. In tegenstelling tot het wijdverspreide idee dat cognitie bestaat uit informatieverwerking in het brein, verklaart de ecologisch-enactieve benadering menselijke cognitie in relationele termen, als vaardige interacties met een sociomateriële omgeving die is gevormd door praktijken. De tweede is de *metatalige benadering* van taal. Volgens deze aanpak is metatalig taalgebruik – praten over praten – cruciaal om taal en haar ontwikkeling te begrijpen. Ik verdedig twee thesen:

1. Het eerste communicatiegedrag van kinderen kan verklaard worden in termen van aandachtrichtende handelingen: sociale handelingen die werken door het richten van iemand anders aandacht.
2. Om een sensitiviteit voor belangrijke eigenschappen van taal te ontwikkelen, zoals semantische inhoud en normativiteit, moet een kind metatalige vaardigheden leren.

Het in dit proefschrift ontwikkelde ecologisch-enactief perspectief op taal dient twee functies. Allereerst is de ecologisch-enactieve benadering begonnen met het verklaren van basaal gedrag, zoals bewegen en grijpen. Een benadering in de cognitieve wetenschappen moet echter het volledige spectrum van menselijke cognitie kunnen verklaren. Als het perspectief dat in dit proefschrift is ontwikkeld houdbaar is, vormt dit een bijdrage aan de uitbreiding van de ecologisch-enactieve benadering naar typische menselijke vormen van cognitie. Ten tweede werpt dit perspectief nieuw licht op filosofische problemen met

betrekking tot taal. In de verschillende hoofdstukken behandel ik vragen met betrekking tot de aard van talige kennis, verklaringen van communicatief gedrag, en de oorsprong van semantische inhoud en talige normativiteit.

Dit proefschrift bestaat uit een inleiding, vier hoofdstukken die zijn geschreven als onafhankelijke onderzoeksartikelen, en derhalve afzonderlijk gelezen kunnen worden, en een algemene conclusie waarin richtingen voor toekomstig onderzoek worden beschreven.

In de inleiding wordt het perspectief dat in dit proefschrift wordt ontwikkeld gecontextualiseerd. In het eerste deel van de inleiding introduceer ik de *pragmatische wende* in de studie van cognitie. Deze wende markeert de hedendaagse verschuiving van het *klassiek cognitivisme*, dat stelt dat alle cognitie bestaat uit breingebonden berekeningen over mentale representaties, naar een handelingsgeoriënteerd paradigma, dat cognitie definieert als perceptueel geleide handelingen. In het tweede deel van de inleiding introduceer ik een breed gedeelde aanname die ten grondslag ligt aan veel theoretische reflecties over taal, namelijk dat taal gezien moet worden als een abstract systeem van codes. Volgens deze aanname is taal een code voor gedachten, zodat door het uitwisselen van deze codes gedachten van de ene persoon naar de andere kunnen worden overgedragen. Ik laat zien waarom deze aanname hand in hand gaat het klassiek cognitivisme. Daarna introduceer ik de metatalige benadering, die belangrijke eigenschappen van taal verklaart in termen van normatieve metatalige praktijken.

In hoofdstuk twee stel ik een verklaring van talige kennis als Ryleaanse praktische vaardigheden voor. Ik behandel Nigel Loves invloedrijke onderscheid tussen eerste orde talig gedrag en tweede orde metatalige constructen die we gebruiken om dit eerste orde talige gedrag te beschrijven, zoals *woord, zin, betekenis,* en *begrip*. Gebaseerd op Taylor's idee van criteriumsrelaties, definieer ik talige kennis als de praktische vaardigheid

criteria te geven en te herkennen. Om talig te kunnen communiceren moet men een gevoeligheid hebben voor de criteriumsrelaties die voortkomen uit metatalige praktijken. Ik concludeer dat alle eerste orde talige activiteit een tweede orde dimensie heeft, in de zin dat wij talige activiteit direct waarnemen in termen van metatalige praktijken.

In hoofdstuk drie ontwikkel ik een ecologisch-enactieve verklaring van het gebruik van concrete woorden – woorden die worden gebruikt om observeerbare situaties, gebeurtenissen, objecten of kenmerken aan te duiden. Voortbouwend op het scholing van aandachtsmodel van leren, definieer ik het gebruik van concrete woorden als aandachtsrichtende handelingen *(attentional actions)*: een herhaalbare vorm van gedrag die door een persoon wordt uitgevoerd om een aspect van de huidige situatie aan te wijzen aan iemand anders om iets te bereiken. Recent empirisch onderzoek laat zien dat de aandachtsrichtende categorie-effecten van talige categorieën cognitieve processen beïnvloeden in de afwezigheid van verbaal gedrag. Ik betoog dat dit begrepen kan worden op basis van het idee dat mensen continue zelforganiseren in taak-specifieke manieren. Dit is een vorm van fenotypische herorganisatie. Dit proces van continue zelforganisatie wordt begrensd door de leergeschiedenis, de handelingen van de persoon zelf en van anderen, en door aspecten van de omgeving. Op deze manier stelt dit hoofdstuk een nieuwe verklaring voor voor het fenomeen van categorie-effecten die voortkomen uit talige categorieën.

In hoofdstuk vier stel ik een manier voor om het probleem van de oorsprong van semantische inhoud op te lossen door te laten zien hoe kinderen leren deel te nemen in praktijken waarin inhoud een rol speelt. Ik beschrijf de vaardigheden die een kind nodig heeft om een gevoeligheid te ontwikkelen voor de semantische inhoud van taalhandelingen. In het bijzonder betoog ik dat een kind ‘gecalibreerd’ moet worden om haar wereld te ‘meten’ in termen van de

normatieve gelijkheidsoordelen van haar gemeenschap. Dit proces kan verklaard worden middels het scholing van aandachtsmodel voor leren. Het gecalibreerde kind moet haar ‘metingen’ nog leren te gebruiken in het produceren en begrijpen van inhoudsvolle taalhandelingen. Het kind leert dit door reflexieve talige vaardigheden te verwerven die haar in staat stellen deel te nemen aan het bepalen van correctheidsvoorwaarden.

In hoofdstuk vijf verdedig ik de these dat metatalige reflexiviteit constitutief is voor talige normativiteit. Een potentieel vernietigend tegenargument tegen de constitutiviteit van metatalige reflexiviteit is het regressieargument. Middels een bespreking van het werk van Searle laat ik zien dat dit regressieargument voortkomt uit een visie op taalverwerving als het leren volgen van de constitutieve regels van talige praktijken. Volgens deze visie is taal leren vergelijkbaar met leren schaken: we leren *eerst* de regels, om die vervolgens toe te passen. Ik stel een ecologisch-enactief alternatief voor, waarin taalverwerving begrepen wordt als het eerst aanleren van regelmatig communicatiegedrag, dat verklaard kan worden in termen van aandachtsrichtende handelingen. Volgens deze alternatieve visie moet een kind eerst regelmatig gedrag vertonen, zodat ze dit regelmatige gedrag retrospectief in normatieve metatalige termen kan interpreteren. Metatalige reflexiviteit maakt het mogelijk om communicatiegedrag te reguleren, en is derhalve noodzakelijk voor talige normativiteit. Ik betoog dat talige regels beschikbaar zijn voor mensen als middelen om de eigenschappen van taal en talig gedrag in onderling overleg te bepalen. De benadering in dit hoofdstuk maakt het mogelijk de constitutieve rol van metatalige reflexiviteit te begrijpen zonder hierbij ten prooi te vallen aan het regressieargument.

Summary

This thesis proposes a perspective on language and its development by starting from two approaches. The first is the *ecological-enactive approach* to cognition. In opposition to the widespread idea that cognition is information-processing in the brain, the ecological-enactive approach explains human cognition in relational terms, as skilful interactions with a sociomaterial environment shaped by practices. The second is the *metalinguistic approach* to language, which holds that reflexive or metalinguistic language use – talking about talking – is crucial for understanding language and its development. In particular, I defend two theses:

1. A child's initial communicative behaviour can be explained in terms of attentional actions: social actions that function by directing someone else's attention.
2. In order for the child's communicative behaviour to be sensitive to key properties of language, such as semantic content and normativity, she needs to learn metalinguistic skills.

The development of this ecological-enactive perspective on language serves two functions. *First*, the ecological-enactive approach started by considering basic behaviour, such as locomotion and grasping. An approach in the cognitive sciences, however, should be able to account for the full gamut of human cognition. If the perspective developed in this thesis is viable, this is a contribution to extending the ecological-enactive approach to typically human forms of cognition. *Second*, this perspective throws new light on philosophical problems concerning language. In the different chapters, I deal with questions

concerning the nature of linguistic knowledge, explanations of communicative behaviour, and the origins of semantic content and linguistic normativity.

This thesis consists of an introduction, four chapters that are written as independent research papers, and thus can be read separately, and a general conclusion that identifies directions for future research.

The introduction serves to contextualise the perspective developed in this thesis. In the first part of this introduction, I introduce the *pragmatic turn* in the study of cognition. This turn marks the current shift from *classical cognitivism*, according to which all cognition consists essentially in brain-bound computations of mental representations, to an action-oriented paradigm, according to which cognition is redefined as perceptually guided action. In the second part of the introduction, I introduce a dominant thread in theoretical reflections on language that conceive of language as an abstract system of codes. On this *code-view*, a language encodes thought and thereby allows for the transference of thoughts from one person to another. I show why this view of language goes hand in hand with a classical cognitivist account of cognition. I then introduce the metalinguistic approach, according to which key properties of language should be explained in terms of normative metalinguistic practices.

In Chapter 2, I propose an account of linguistic knowledge in terms of Rylean know-how. I discuss Nigel Love's seminal distinction between first-order linguistic behaviour and second-order or metalinguistic constructs that we use to describe this first-order linguistic behaviour, such as *word, sentence, meaning,* and *understanding*. Based on Taylor's notion of criterial relations, linguistic knowledge is defined as knowing-how to provide and recognise criterial support. To be a competent linguistic communicator one has to be sensitive to the criterial relations that originate in metalinguistic practices. I conclude that all first-order linguistic activity has a second-order dimension to

it, in the sense that it is directly perceived in terms of the metalinguistic practices that constitute it.

In Chapter 3, I develop an ecological-enactive account of utterances of concrete words – words used to indicate observable situations, events, objects, or characteristics. Building on the education of attention model of learning, utterances of concrete words are defined as *attentional actions*: a repeatable form of behaviour performed by a person to indicate (i.e. point out) a particular aspect of the current situation *to* someone *in order to* achieve something. Recent empirical studies show that the attention-directing effects of linguistically constituted categories affect cognitive processes in the absence of overt language use. I argue that this can be understood by relying on the idea that humans continuously self-organise into task specific devices, a form of phenotypic reorganisation. The unfolding of this process is constrained by a history of learning, the actions of oneself, others, as well as aspects of the environment. In this way this chapter proposes a novel explanation for the empirical phenomenon of category effects.

In Chapter 4, I propose a way of solving the hard problem of content by showing how children can become competent participants in content-involving practices. I develop an account of the skills a child has to learn in order to develop a sensitivity to the contents of speech acts. In particular, I argue that a child has to be 'calibrated' to measure her world in terms of the normative similarity judgments of her community, a process that can be explained in terms of the education of attention. The calibrated child still has to learn to put her 'measurements' to use in producing and understanding contentful speech acts. The child learns to do so by acquiring reflexive linguistic skills that enable her to participate in the negotiation of correctness conditions.

In Chapter 5, I defend the thesis that metalinguistic reflexivity is constitutive of linguistic normativity. A potential defeating counterargument

against the constitutivity of metalinguistic reflexivity is the regress objection. By discussing the work of Searle, I show that this regress objection originates in the idea that learning language requires learning to follow the constitutive rules of linguistic practices. On this view, learning language is like learning chess: we *first* learn the rules and then put them in practice. I propose an ecological-enactive alternative, according to which learning language can be understood as first learning regular communication behaviour, which can be explained in terms of attentional actions. On this view, it is only because a child *first* behaves in regular ways, that she can *then* retrospectively interpret her own behaviour in normative metalinguistic terms, that is, as being guided by rules. Metalinguistic reflexivity thus enables regulation of communicative behaviour, and thereby constitutes linguistic normativity. I argue that linguistic rules are resources: they are available to participants in order to (re)negotiate properties of language and linguistic behaviour. The account developed in this chapter thus enables us to understand the constitutive role of metalinguistic reflexivity for linguistic normativity without falling prey to the regress objection.

About the author

Jasper van den Herik (1986) holds an MA (*cum laude*) and BA in Philosophy and a BSc in Psychology from the Erasmus University Rotterdam. He has presented work at many workshop and conferences, including the European Workshop of Ecological Psychology, The International Conference on Language and Enaction, The International Conference on Interactivity, Language and Cognition, and the European Philosophy of Science Association Conference.

Personal website
www.jaspervandenherik.nl