universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

Theoretical and Corpus-Based Investigation of the Relation Between Sensorimotor Processes and Linguistic Structure

verfasst von / submitted by

Tim Reinboth

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2020 / Vienna, 2020

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 013
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Joint Degree Programme MEi:CogSci Cognitive Science
Betreut von / Supervisor:	Stephanie Gross, PhD
Mitbetreut von / Co-Supervisor:	-

Abstract

This thesis offers an embodied perspective on the interface between sensorimotor cognition and linguistic structure. It is inspired by Knott (2012), who suggests that the underlying structure of a sentence like "the woman grabbed a cup" can be thought of as being based in a sequential representation of certain sensorimotor processes that make up a default order of actions involved in reaching for a cup. This thesis extends previous empirical work on the subject by conducting a new multimodal data collection that draws on corpus-based analyses of regularities in human language. In doing so, it lays the foundation for a detailed exploration of specific predictions about the sensorimotor basis of linguistic structure. To this end, the thesis makes a variety of contributions spanning i) a discussion of its conceptual basis and the theoretical frameworks on which it is based, ii) an original experimental design and data analysis pipeline, as well as a novel corpus (Reinboth, Gross, Bishop, & Krenn, 2020), iii) a software tool to visualise the data recorded for this study, plus iv) suggestions for future research and applications based on visualisations generated for this thesis, and v) a discussion of fundamental sociopolitical issues in which this work is embedded. That means the thesis draws together aspects of artificial intelligence, linguistics, philosophy, psychology, and science and technology studies into a transdisciplinary perspective on the sensorimotor basis of linguistic structure.

Keywords: corpus, embodied cognition, eye-tracking, linguistic structure, motion-capture, sensorimotor processes

General Terms: attention, computationalism, grounded cognition, language, multimodality

Acknowledgements

I'm glad to acknowledge, alphabetically by last name: Laura Bishop PhD, for her unwavering help and support, even from Norway; Stephanie Gross PhD, for her commitment, for keeping me focused, and for her tireless, clear feedback; Brigitte Krenn PhD, for her encouragement and for smilingly seeing the bigger picture beyond just the thesis itself. And always my thanks to my family, without whom everything would be especially different.

Contents

List of Figures

List of Tables

List of Boxes

GREY BOXES

Grey boxes appear in the first three chapters. Each box contains a clarification, or one or more definitions. Boxes are usually included where the corresponding concept first does some explanatory work. They are included to foreground concise definitions/clarifications, both for ease of reference and to avoid interrupting the discussion. Whenever a concepts/definition is central to the main line of argument, it is discussed in the text-body as well, though possibly at a later time.

Chapter 1

Introduction

This thesis is inspired by work about the interface between sensorimotor cognition and natural language. It continues a line of research in embodied cognitive science (see e.g. D. H. Ballard, Hayhoe, Pook, & Rao, 1997), which holds that cognition in general (see e.g. Barsalou, 2008), and language specifically is grounded in sensorimotor processes (Glenberg & Kaschak, 2002; Zwaan & Taylor, 2006; Gleitman, January, Nappa, & Trueswell, 2007; Glenberg & Gallese, 2012). Here "grounded" refers, to a varying extent, to an evolutionary, developmental, or momentary dependence of a "higher"-order cognitive processes on sensorimotor processes (c.f. Harnad, 1990; Haggard, Rossetti, & Kawato, 2008; Barsalou, 2010). It means these cognitive processes always depend on the sensory and motor capacities of an organism (for a clarification, see Box 1.1).

Meanwhile, the "embodiment" of embodied cognitive science is not a homogeneous concept (see e.g. M. Wilson, 2002; Ziemke, 2003b). Broadly, it is one kind of grounding (Barsalou, 2008), and can be understood to refer to how cognition is shaped by aspects of the body of an organism (cf. Varela, Thompson, & Rosch, 1991; R. A. Wilson & Foglia, 2017). That means the cognitive processes of any organisms are shaped by how it can act in and perceive its world. But, embodiment might be best thought of as being graded, rather than binary – as in either embodied, or not – (M. A. Arbib, Gasser, & Barrès, 2014). Then the question becomes to which extent, if any, a certain set of cognitive processes is embodied. This thesis addresses a set of cognitive processes underlying linguistic structure (see Box 1.2).

Box 1.1

Sensorimotor Cognition The term "sensorimotor" (also "sensory-motor") indicates the tight coupling of action and perception. It is used to refer to anything – from the functioning of individual cells to complex behaviours – that always depends on both the sensory and motor functions of an organism (R. A. Wilson & Foglia, 2017). Sensory-motor neurons, for example, are neurons that contribute to the control of purposeful actions, and respond to both sensory information from one or more modalitie(s) (Gallese & Lakoff, 2005). This reflects a multimodality, meaning an integration of modalities (sensory, somatosensory, motor) at the level of sensory processing (for a review, see Fogassi & Gallese, 2004). In this respect, sensorimotor cognition opposes an emphasis on "supramodality", a term that stipulates that sensory integration takes in association areas that are not part of either a sensory or motor system. As a result, the sensorimotor perspective questions a modular approach to cognition, that imputes a variety of largely distinct cognitive systems (see e.g. Fodor, 1983). As opposed to this view, the present work emphasises the fundamental interdependence of sensory and motor processes, as a result of which they are referred to conjointly as "sensorimotor" processes (for some early formulations relevant to the present work, see M. Arbib, 1981; D. H. Ballard, 1991).

Specifically, this thesis concerns the ways language can be thought of as being embodied (cf. Jackendoff, 2002; Clark, 2006a; Jirak, Menz, Buccino, Borghi, & Binkofski, 2010). In particular, the claim of this thesis concerning language is that certain processes of language-related cognition are one and the same sensorimotor processes associated with humans' ability to act in the world. To be clear, the claim is that if we study certain aspects of language cognition, we are studying sensorimotor processes, too (Knott, 2012).

Box 1.2

Linguistic Structure

In this thesis, we discuss "linguistic structure" and not "syntax". The two concepts are similar but not identical. We do so, because syntax is not the focus of the present work. Instead, this thesis is concerned with the interface between language and sensorimotor processes, which operates through a shared mechanism and a particular kind of representation. With "linguistic structure" we refer to those underlying regularities in how a sentence gets in meaning that are rooted this shared mechanism. As a result of this, the term syntax is not adequate for what is investigated here. This issue is discussed in Subsect. 2.2.4.

To argue for that point, this thesis draws on a model of the sensorimotor basis on natural language syntax, developed by Knott (2012). The model proposes that human language can be characterised in terms of an underlying structure that is universal across languages, and therefore different from the surface structure of any given language (as reflected in word order, for example). That structure, Knott (2012) argues, is grounded in sensorimotor processes. To illustrate this, the entire framework is discussed in relation to action of reaching to grasp and the concrete sentence "the woman grabbed the cup". Knott (2012)'s approach is broadly introduced below (and discussed in Chapter 3). First, though, some final opening remarks follow.

Briefly, this study had three objectives. The main objective was to contribute to investigating the sensorimotor basis of language. To do so, this thesis discusses some fundamental conceptual issues (see Chapter 2) and reports on the creation of a new corpus (see Reinboth et al., 2020), based on which further empirical work can be conducted. This is the red thread that runs through the present work. It traces a theoretical and a corpus-based investigation of the relationship between sensorimotor cognition and the linguistic structure of concrete sentences like "the woman grabbed a cup".

The second objective was to develop independent lines of research, especially applications, inspired by the relationships between sensorimotor processes and language. That means this line of work does not necessarily aim to contribute to investigating Knott (2012)'s sensorimotor model. Instead, it reflects the search for ways that ideas inspired by the model can be applied (see, for example, Sect. 6.3).

The last, but not least important objective was to articulate some of the assumptions that are at work behind the main line of inquiry. This continues throughout the entire thesis. The most explicit discussion of underlying of issues with which this work is interwoven is carried out in Chapter 7. With this third objective, I point out that the articulation of implicit assumptions is an inherent and vital part of this investigation, and not an optional extra. This approach was inspired by the idea of a *critical practice* (Agre, 1997b; Boehner, David, Kaye, & Sengers, 2005). Likewise, it is motivated by critical engagement with cognitive science specifically (e.g. L. A. Suchman, 2007), and scientific practice in general (e.g. Barad, 1999; Orr, 2006; Latour & Woolgar, 2013). One challenge this poses is finding a balance between presenting relevant research one the hand, and, on the other, critically reflecting on the assumptions behind that research, and those behind this thesis, too.

1.1 Knott (2012)'s Sensorimotor Model

At this point, I will introduce the framework that inspired this thesis as briefly as possible. It is discussed in detail in Chapter. 3. A few opening comments are necessary, though, to explain what this thesis is about. Starting from the beginning, Knott (2012)'s framework is built on research in both neuroscience (e.g. M. Arbib, 1981), and artificial intelligence (AI) (e.g. Agre & Chapman, 1987). The neuroscientific line of research investigated the dynamics of routine activities, searching for regularities in the interaction between a human and her environment. On the AI side, Agre and Chapman (1987) criticised traditional approaches to planning. In contrast to these, they developed a theory of activity that also emphasised the role of regularities in the interaction between agents and their environment.

In a nutshell, Knott (2012) then builds on the idea shared by both sets of research, namely that many routine behaviours are, in fact, constituted by a canonical sequence of actions. "Canonical" means such a sequence would have a default order in which the cognitive operations that make up the corresponding action have to (or usually do) occur. The kind of action this is about is captured by the example of a woman grabbing a cup. In fact, the action of reaching to grasp appears to follow a regular pattern of hand movements (Jeannerod, 1981), eye-movements (Webb, Knott, & MacAskill, 2010), attention (D. H. Ballard, 1986) and information processing (Jeannerod, 1996). In other words, for reaching to grasp, a certain sequence of action is canonical. That canonical sequence is the conceptual keystone of this thesis. Because, using the example of reaching to grasp, Knott (2012) develops an account of the relationship between language and sensorimotor cognition that depends on the action having such a regular, sequential structure.

Making this argument requires an interface between sensorimotor cognition and language. Abstractly, this can be thought of as a shared mechanism, a set of cognitive processes that contributes to both the woman reaching to grasp a cup and to her understanding the sentence "the woman grabbed a cup". For Knott (2012), this overlap occurs because the syntatic structure of the sentence "the woman grabbed a cup" can be understood as a description of the sequence of actions involved in experiencing that episode, either in the first or third-person. This statement will be explored in detail in Sect. 3. For now, it is most important that i) this thesis builds on the idea of a set of shared mechanisms that subserve both sensorimotor and linguistic cognition, and ii) that linguistic structures selectively reproduces the order (or sequence / temporal structure) of sensorimotor processes.

Broadly, then, Knott (2012) investigated whether there is a direct structural relationship between (i) sensorimotor processes, (ii) the cognitive representation of episodes (e.g. reaching to grasp a cup), and (iii) the sentences describing those episodes (see also Webb, 2015). I will likewise argue that part of the structure we observe in language is derived from the canonical structure of sensorimotor processes and representations that underlie action performance and action observation. This is made possible by a specific interface between sensorimotor process and linguistic structure, namely working memory (for a specific model, see A. D. Baddeley & Hitch, 1974; A. Baddeley, 2003). Here, a sequential representation of actions is maintained that captures some of the (temporal) dynamics of the action (A. Baddeley, 2000). Significantly, the working memory representation is not static. Instead, it is thought of as an unfolding process of remembering that simulates some of the dynamics of the original action (for review discussion of simulation and embodied memory, see e.g. Dijkstra & Post, 2015; Soylu, 2016).

Specifically, the working memory representation of an episode reflects a pattern of attentional actions that guides the actions as it is performed. This pattern (sequence) Knott (2012) argues, has all the properties that would allow it to acts as the basis of linguistic structure. Namely, it contains repeated references to the agent, patient and action that exist as thematic roles in the sentence describing the action. These repeated references are instrumental, because they allow a canonical sensorimotor sequence, translated into a universal linguistic structure, to give rise to entire variety of word orders evident across different human languages (see Sect. 2.2.4). The idea of universal, underlying linguistic structure draws on a version of Minimalist theory (Chomsky, 2014) that Knott (2012) reinterprets in sensorimotor terms.

Whereas, on the sensorimotor side, the basic idea is that a simple action, such as grabbing a cup, is achieved through well-defined sensorimotor operations that occur in (more or less) strict sequence. On the linguistic side, the framework envisions that all natural language syntax can be characterised by a similarly strict sequence. Knott (2012)'s is an appealing approach, because it outlines a single mechanism that accounts for both the surface variability and the underlying similarity of human language. This is possible because certain kinds of actions reoccur throughout the sequence that constitutes reaching to grasp. This underlying sequence reflects the universal, linguistic structure, and the basis for every human language (for a discussion, see Sect. 2.2.4). [1]

Meanwhile, it is the regularities in the sensorimotor processes associated with a certain action that are the object of this thesis. In this case, as in Knott (2012) that action is reaching to grasp. What I will argue is, that these regularities are exploited as the basis for the underlying structure of language, via an interface of sensorimotor and linguistic processes through working memory. [2] To investigate this, the present work contributes (among other things) valuable new empirical data. Since, instead, much previous work has focused on computationally modelling the working memory mechanisms involved (see also Takac & Knott, 2015; Knott & Takac, 2016). As a result, this thesis promises several novel contributions to the original framework.

1.2 Going Beyond Knott (2012)'s Sensorimotor Model

As it stands, the precise relation between linguistic structure and sensorimotor operations remains an open question – in fact, there has been little empirical work besides that by Webb et al. (2010). Two possible explanations for why there is no decisive evidence for the particular relation suggested by Knott (2012) are that either (i) the hypothesised relationship does not exist, or (ii) that previous experimental methods were not well suited to explore it (Webb, 2015).

Concerning (i), one issue is that the theory could be incorrect in any number of ways. If that is the case, the data we collected here could still help revise the current proposal. In this respect, two further points should be kept in mind. Namely, (a) it is possible both that the link between sensorimotor representation and linguistic structure may not be as close as it stipulated here, or (b) that this link is simply not apparent from the verbal descriptions of the task the agent is currently performing. Why this might be the case would be another questions, and is addressed explicitly in Sect. 4.5 and Sect. 5.4. [3]

In this situation, the contributions of this thesis to the empirical work being conducted to test Knott (2012)'s model include a novel corpus, a software tool for generating visualisation of the data in the corpus, and suggestions for specific future work, designed to test individual predictions of the model. With these, this thesis considerably extends the current state of research on the sensorimotor model.

[1] This idea is reinforced by biologically-inspired computational models that are able to produce different word orders corresponding to different languages based on the same sequential representations plus input matching the corresponding language (e.g. Takac, Benuskova, & Knott, 2012; Takac & Knott, 2018).

[2] To be clear, the conceptual differentiation into sensorimotor and linguistic processes is not intended to reflect any such distinction in neural terms. On the contrary, the argument here, it should become clear, is that a set of cognitive processes is shared across linguistic and sensorimotor cognition.

[3] Granted, one possible reason is that the structure of the verbal description does not depend on the structure of the sensorimotor representation that this thesis argues for.

At the same time, though, this thesis is not completely in line with Knott (2012)'s proposal. Most importantly, the original focus on syntax is replaced by a discussion in terms of linguistic structure (see Box 1.2). Basically, this reflects a move away from Chomskyan linguistics, in favour of an approach that is situated, embodied and consequently more empiricist. This will be explained further in Sect. 2.2.4. Importantly, the merits of the original proposal survive the change in focus. Indeed, many of Knott (2012)'s ideas play a central role in this thesis.

1.3 Structure of the Thesis

In some ways, then, this thesis is quite conventional. It features computationalism, "plans", representations, neural networks, and functional neuroscience. Indeed, the empirical, corpus-based work reported here captures more of this aspect. In other parts, though, this thesis reflects an attempt to cross the disciplinary boundaries within which these topics are usually discussed. This aspect of the thesis is reflected more in the theoretical parts. As a whole, this thesis captures both aspects, and brings them into conversation.

On the corpus-based side, the empirical research reported here is motivated by one general, and several more specific questions. The general question can be summarised as: "what role, if any, do sensorimotor processes play in grounding linguistic structure"? Building on this, the specific questions query concerns from the dynamics of sensorimotor processes involved, their relationship to word order, and how a systematic relationship between sensorimotor processes and language could be applied. In terms of the corpus, novel contributions include the collection of data from the perspective of the acting agent, rather than on observation of other's actions answer, which is a open question identified by Webb (2015).

First though, this thesis begins with theoretical aspects. So, Chapter 2 offers a theoretical introduction, separated into two sections. In Sect. 2.1 I make a number of conceptual clarifications that are intended to establish definitions. This should also deal with some ambiguities about how certain concepts will used during this thesis. Next, I discuss a few of the frameworks in which this thesis is embedded (Sect. 2.2). In doing so, I draw together the positions this thesis takes on embodied cognition (see Sect. 2.2.1), computationalism (see Sect. 2.2.3), and Knott (2012)'s sensorimotor interpretation of Minimalism (see Sect. 2.2.4).

Next, Chapter 3 summarises Knott (2012)'s sensorimotor model. I first introduce some basic ideas that the model is built on, then discuss the role of attention in embodiment (Sect. 3.1). This introduces D. H. Ballard et al. (1997) crucial work on "deictic routines". Then, Sect. 3.2 outlines the sequence of actions captured by (Knott, 2012)'s model. The role of working memory in the model is addressed in Sect. 3.3 and Sect. 3.3.1. Finally, Sect. 3.4 adds an evaluation of the senorimotor model.

After this, the next three sections describe the LKG-corpus: how the data was collected (see Chapter 4), how it was processed (see Chapter 5), and how it can be analysed and interpreted (see Chapter 6). However, while data collection and processing are reported as usual in Chapter 4 and Chapter 5, the emphasis in these sections is ultimately on lessons learned from the study design and analysis methods. Likewise, Chapter 6 mainly discusses how the data reported here can be analysed, and what research might be pursued on this basis. It also introduces a tool to visualise the data collected here (Sect. 6.1) and a discussion of applications of this data (Sect. 6.3).

Finally, Chapter 7 delivers on the promise of a cross-disciplinary discussion of some underlying questions. Due to editorial constraints, these issues are not tackled comprehensively. Instead, the aim was to bring them to the reader's attention, either to sensitise to or to remind her of two points, namely (i) that these concerns exist, and (ii) that they are fundamental. For example, Sect. 7.1 addresses the tension between objectivity in science and the subjectivity of scientific knowledge production (see e.g. E. Anderson, 2020). Since, knowledge production, like cognition, is situated (D. Haraway, 1988). Relating these questions back to the study design reported here, Sect. 7.2 asks what kind of knowledge is produced when bodies are turned into data (see e.g. Viseu & Suchman, 2010). I also consider what happens to our understanding of bodies and knowledge, as a result of abstraction, quantification and the reduction to data. In doing this, in particular, I deliver on the promise of a critical practice.

Then, to conclude, Chapter 8 adds a summary and an inventory of the novel contributions of this thesis, in relation to the work it is based on. In advance, previous work most notably includes the sensorimotor model itself (Knott, 2012), and the seminal empirical efforts of Webb et al. (2010); Webb (2015). Chapter 3 is heavily indebted to these authors. In contrast, the study design reported here, several aspects of the analysis, the corpus itself, data visualisation tool, and discussion of underlying framework and sociopolitical issues all reflect novel contributions. Indeed, the first original contribution follows in the coming chapter, which begins the thesis proper by discussing and clarifying some core concepts and frameworks on which the rest of this effort depends.

Note on the LKG-Corpus

The corpus has been published as the Linguistic, Kinematic and Gaze Data in Task Descriptions Corpus by Reinboth et al. (2020). [4] This thesis picks up from that effort, but should be seen as a new contribution.

[4] to be available online at http://www.ofai.at/research/interact/resources.html

Chapter 2

Situating the Present Work

Given the heteroglossia of cognitive science, this chapter attempts to clarify how a number of concepts are used and understood in the present work. In doing so, I hope to avoid the misunderstandings that are characteristic of debates with inadequately defined concepts. The issue of the parallel languages that can emerge in academia (as well), and the almost complete inability of users of different ones to relate to one another has perhaps been best captured by the inimical Bill Watterson (2005) – see Fig 2.1. Because, at least two challenges arise when different frameworks bring distinct nomenclatures. On the one hand, two frameworks may address the same phenomenon in terms that are so different that it is not clear that the same phenomenon is being discussed at all. On the other hand, two different approaches may have perspectives and nomenclatures that are so different, that they cannot address one and the same phenomenon in the first place. Both are challenges we should be sensitive to.

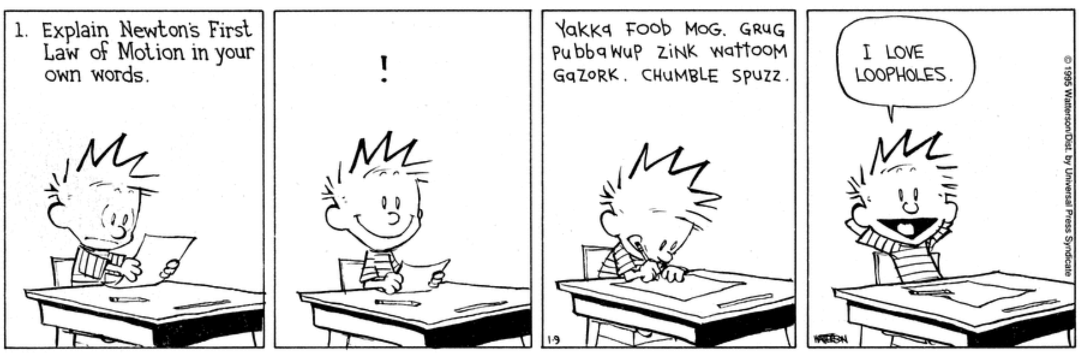

Figure 2.1: What is a loophole in this cartoon, is an ever-present challenge in scientific communication. As the cartoon indicates, we may be referring to the same phenomenon, even saying the same thing about it, but be unable to understand one another at the same time.

In light of this, the following sections outlines a number concepts and frameworks that the present work depends on. As a starting point, I take an explicit position on a number of general conceptual issues most relevant to the present work (see Sect. 2.1). Then, in Sect. 2.2, these positions are deepened as I introduce two frameworks in which the current work is embedded: embodied cognition and computationalism; their introduction is followed by a brief discussion of the minimalist program (Chomsky, 2014) in Sect 2.2.4. Although that framework is not part the work proposed by this thesis, it does play a central role in (Knott, 2012), which is why it bears discussion. I conclude this chapter with a few notes relating this work to the development of cognitive science as a scientific discipline.

2.1 Conceptual Clarifications

This section reviews a number of fundamental concepts in cognitive science. The purpose is not to discuss these at length, but to sensitise the reader to, or remind her of certain issues. Given editorial constraints that preclude a thorough exploration, this short introduction should also indicate an awareness of these issues on the author's part. I raise these issues within the scope of the present work out of the conviction that a critical empirical practice requires it. While a number of considerations have been relegated to later sections (e.g. Sect. 7.1, others must be discussed before this work can proceed.

2.1.1 Frameworks, Theories, and Models

Proceeding from most specific to most general, a model is constituted by a theory and (situation-specific) assumptions about how that theory can be applied in a given situation. The theory itself makes testable general predictions deriving from an elaboration on one aspect of a framework. The framework, in turn, is a set of general claims about a phenomenon, including specific constructs and an emphasis on a given aspect of that phenomenon (for a discussion of this terminology, see e.g. J. R. Anderson, 1983). Note that when we test and evaluate a model, what we are evaluating is always a specific instance of that model. Knott (2012)'s proposal can be read as each of these. As a result, the present work discusses frameworks, theories and models at different times. The author endeavours to make it clear which is being discussed when. The most frequent reference is to Knott (2012)'s sensorimotor *model*, meaning some specific assumptions are always being referred to as well.

Meanwhile, one point about theories is that making testable predictions does not require a formulation in terms of postulates and laws (cf. A. Baddeley, 2012). In contrast, we can think of a theory like a map, an artefact that organises knowledge and provides a tool for interacting with it (Toulmin, 1953). Thus, while the accuracy of predictions is important, this thesis asserts it is more important that the theory is productive and that it generates fruitful questions (Lakatos, 1976). This issue is especially important here, because I do not intend to arbitrate opposing sets of beliefs in this work (recall Agre, 1997a). Part of the reason for that is that there is good reason to be sceptical about the idea that there is a unitary phenomenon "out there" to be theorised about (Latour & Woolgar, 2013). This thesis is all for generating clear, testable predictions, but it departs from (Popper, 1959)'s logic by being opposed to the idea that an "objective" truth of its subject can be settled in favour of one or another view (for more discussion of this issue, see Sect. 7.1).

Concerning models, it is important that the instances discussed in the present work are of different types. There are examples of *computational* modelling, e.g. cognitive architectures (cf. Takac et al., 2012), and examples of *theoretical* modelling, e.g. the model of action described by (Knott, 2012). These stand in a particular relation to each other and to empirical testing. In general, the usefulness of the model is derived from the fact that models are abstractions that should allow us to investigate phenomena we cannot deal with directly. As such they must be simpler than the modelled phenomenon but preserve its essential features (think also of Bonini's paradox that models become less and less useful abstractions the more specific they become Bonini, 1963). Related challenges are known as i) the frame problem: how to identify the necessary axioms on which to base the description of a phenomenon (McCarthy & Hayes, 1969), and ii) the irrelevant specification problem: how to avoid including unnecessary aspects in the model by mistake.

A successful model should address each of these points. In addition, a number of utility criteria including parsimony, effectiveness, generality, accuracy and surprise offer ways to assess models (Fum, Missier, & Stocco, 2007). These criteria will be used to evaluate Knott (2012)'s claims about the sensorimotor basis of natural language structure in Sect. 3.4.

To understand what a model is, it is worth keeping in mind how models relate to their target phenomena. The perspective of this thesis is constructivist, meaning no model picks out a real and definite target. Instead, the process of developing a model should be understood as one aspect that contributes to the way we come to understand the phenomenon. Simply put, modelling the phenomenon constructs our understanding of it. In light of this, we should bear in mind that models are built using (and limited by) specific representational means. These means facilitate the study of certain scientific questions, and impede the study of others (Knuuttila, 2011).

So, Knott (2012)'s work, and this effort, too, are expressions of the means by which they are conducted. Models are part of the epistemic regime (see Böschen, 2019), embedded in which the present work takes place. In that sense, they are one of the stabilised practices through which problems are articulated, debated, and considered solved. This point will be picked up again in Sect. 7.1. Briefly considering these issues at this point serves as a reminder of the promises and pitfalls of any form of modelling. Aware of these points, the present work is careful not to make totalising claims, or imagine that it is depicting an objectively knowable reality "out there".

2.1.2 Levels of Analysis

Knott (2012)'s attempt to model action faces an *identification problem*. Indeed, any attempt at modelling cognition, or any other process with hidden mechanisms that cannot be determined directly, does so. This is because many different mechanisms may produce the same behaviour. For a given cognitive process, Marr (1982) proposes three levels of representation. These are the levels of i) computational theory: a description of the function of that process in terms of goal and strategy, ii) representation and algorithm: a description of the form information takes and how it is manipulated, and iii) implementation: the physical instantiation of the process. Crucially a given competence or computational theory can be implemented in different sets of representation and algorithm, which in turn afford different physical implementations. The levels are useful because they allow us to specify different kinds of mechanisms, different types of effects on observed behaviour.

At the same time, we should bear in mind that these levels do not neatly carve nature at 'her' gendered joints. [1] Marr was explicit that the levels of analysis he proposed can never be thought apart. Each depends on and is inseparable from the other two. They are useful, not to carve up phenomena, but because they help us to state from which perspective we are analysing a phenomenon. When drawn on in this sense, a backseat is assigned to the contested aspects of Marr's vision of cognitive processing, like his emphasis on cognition detached from the agents situation (for a critique of this aspects of Marr's levels, see McClamrock, 1991). In the present work, for example, we might distinguish different perspectives on reaching to grasp, namely a description of i) what it entails for an agent to be reaching to grasp an object, ii) an abstract

[1] This is a comment concerning the discursive practice of classical, male-dominated science-and-technology (for a discussion of male-dominance and the feminist questions it raises, see e.g. Irigaray, 1989). Specifically, this concerns the gendering of nature as female. From a feminist perspective, questions have been raised both about the use of gendered descriptions in scientific discourse, such as, for example, this consistent gendering of nature as female (Ortner, 1972). This is a practice that is continuous with the subjugation of woman (and nature) (see also D. J. Haraway, 1989). In fact, feminist perspectives have identified that same subjugation in the habit of creating strict taxonomies, which often fail to take into account any phenomena, perspectives and people that do not fit the expected 'normal' (D. Haraway, 1988). Wherever this practice is performed, we should be aware of it at least. Here, I take the opportunity to question such constructions as "nature = female" by referring to seminal objections to it (cited above).

definition of how it is possible for the agent to perform the action, and iii) a description of the processes that physically occur while the agent performs the action we name reaching-to-grasp.

As early as this, the reader may note that the computational language used in this effort should be read as being strictly metaphorical. Though, perhaps 'metaphorical' is misleading, since calling something a metaphor implies that it cannot really be tested and rejected, which computational models can be (Miłkowski, 2018). Indeed, the proposed work is a description of information processing by the agent. The image of metaphor is brought up here solely to acknowledge that focusing on the computational level means we run the risk of dealing with an idealisation of cognition, instead of a phenomenon that is inseparable from implementational aspects (Miłkowski, 2013), which in the case of human cognition not-so-small factors like culture. Again we encounter the challenge of levels, in this case as the problem of how to emphasise important processes against a rich background of other aspects and copious implementational detail (Kitcher, 1984). This issues will be taken up in the following section, specifically in Sect. 2.2.3.

2.1.3 Symbols, Signs, Representations

Symbols have a prominent place in cognitive science. At the same time, the term is notoriously ambiguous (see e.g. Chapter 4 in Miłkowski, 2013). Note that symbol, here, does not necessarily mean a sign that refers to its object in virtue of some observed general or conventional connection (Peirce, 1995). Instead, symbolic representations should be thought of as a kind of cognitive code (L. A. Suchman, 2007). This is usually understood as being instantiated physically in the brain, allowing operations to be performed on it to produce mental states, such as a belief that the woman grabbed a cup. The notion of mental representation used here is indebted to Craik (1943), and builds on the idea that the brain implements processes that share an abstract structure with target processes in the body and world – more on this in Sect. 2.2.1.1.

Now, the role of symbols has been questioned in various ways for many decades (for a discussion, see e.g. Boden, 2008). The controversy concerning situated action (L. A. Suchman, 1987; Vera & Simon, 1993; L. Suchman, 1993) exemplifies the challenge of reconciling what might be seen as two different sets of priorities. Proponents on both sides insisted their perspective was the more important and that their work was being misrepresented. Here, I make no effort to arbitrate that debate. Instead, I want to draw out the point that symbolic structures may be more useful for the description of models of cognition, rather than for constructing formal models (Smolensky, 1987). Indeed, the symbolic is only an approximate and selective account of cognition, and a computational theory of mind does not depend upon it (Miłkowski, 2018).

However, it is important to be careful about what this entails, lest we reach the same impasse as the discussants of situated action. After all, all frameworks, theories and models are approximate and selective. Then, to understand the role of symbols, we can take points from each side of the situated action debate. From the symbolic perspective, we might keep in mind that any appeal to some direct awareness and the organism-environment interaction faces a challenge in accounting for anticipatory behaviour, and the subsequent performing of those behaviours (for a non-representationalist view though, see Chemero, 2009). That suggests mental representations of some form, though not necessarily in terms of symbols, remains important. On the other hand, situated actions need not be nonrepresentational (despite appropriations of the term by Brooks, 1991; Vera & Simon, 1993). What L. A. Suchman (1987) denies (and dynamicists likewise, see e.g. Thelen & Smith, 1996), is that the internal models posited by symbolic accounts can guide behaviour in relative independence from the world. Consistent with this denial, the view of representations adopted here, eschews the concept of a symbol, and finds representations always indebted to body and environment. This is a central tenet of embodied cognition.

2.2 Theoretical Frameworks

Consider one issue that permeates the current work: the question of computationalism, and the controversy around a computational theory of mind (for example, contrast Varela et al., 1991; Clark, 2014). A number of challenges have been posed over the years that demand an answer. Once again though, this is not the place to argue these points. However, I will attempt to clearly position myself with respect to a few long-standing antagonisms.

This chapter contains discussion of two core frameworks: embodied cognition and computationalism; in addition to this, I introduce the minimalist program (Chomsky, 2014) and outline why it plays a smaller role here than in Knott (2012). Concerning embodied cognition, I give a brief overview of the various frameworks that could be referred to by the term. In addition, some data relevant to the current work are introduced. For computationalism, I confront some of the conceptual challenges alluded to above (see Sect. 2.1.2). Finally, I introduce the minimalist program, and explain its significance to Knott (2012), as well as why its plays a much smaller role in the present work. This chapter ends with some comments on the significance of the computational study of language for cognitive science, and in its history of ideas.

2.2.1 Embodied Cognition

At its simplest, embodied cognition proposes that the cognitive systems of perception and action also ground other cognitive processes, including language and memory (Borghi & Pecher, 2011). Embodied cognition is best understood as a framework, meaning a set of general claims about a phenomenon, including specific constructs and an emphasis on a given aspect of the phenomenon (J. R. Anderson, 1983). Typically, research focuses on what have been called 'higher-order' cognitive capacities (e.g. language). The emphasis, rather than being on a specific capacity, is on how a variety of cognitive processes are related to sensory and motor processes. Throughout, the general claim is that the latter type of processes (sensory and motor) are (partially) constitutive of the so-called 'higher-order' cognitive functions (R. A. Wilson & Foglia, 2017).

Within embodied cognition, it is possible to distinguish a number of weak (e.g. M. A. Arbib et al., 2014) and strong alternatives (e.g. Varela et al., 1991). Especially the strong variants are imbued with a heavy load of conceptual baggage. Both the ontological strategy of the ecological and the epistemological approaches of the diverse strong embodied approaches make some valuable contributions (Baggs & Chemero, 2018). The present work takes place within the weak embodied tradition though. So, for example, while supporting an embodied account, this work holds on to both a computational theory of mind and to mental representations (see Sect. 2.2.3).

Even within the domain demarcated by the term embodied cognition, there is a considerable heterogeneity in the understandings of what the embodiment refers to. For example, M. Wilson (2002) identified various types of embodiment in the cognitive science literature, while (Ziemke, 2001) contributes a number of different perspectives on the concept itself from a more philosophical point of view. So, what, in the weak view, is embodied cognition? This can be said to hinge on three claims about the body, namely that bodily processes (i) act as a constraint on the representation formed by an agent, (ii) distribute the computational load between neural and non-neural structures, (iii) act to coordinate cognitive processes in space and time (R. A. Wilson & Foglia, 2017). Each of these points will come up in Knott (2012)'s sensorimotor account of reaching to grasp a cup, discussed in detail in Sect. 3.

One further point to note with respect to embodied cognition now, is the use of the term "simulation". Crucially, embodied simulations (see Box 2.1) reflect a specific understanding of the term, which is not the same as what Knott (2012) refers to in his discussion of the "simulation mode". In Knott (2012)'s case, and that of this thesis, simulation refers to a reactivated working memory representation that will be argued to act as the basis for linguistic structure of transitive sentences about concrete actions. Indeed Knott (2012) uses simulation only in the limited sense that refers to the activation of a motor plan through working memory in the absence of any overt motor action (see Sect. 3.3). Despite this, the two understandings of "simulation" are related. What they share, is an emphasis on the potential of sensorimotor processes to subserve other cognitive processes, including those related to language

Box 2.1

Embodied Simulation

The term "simulation" is used to refer to a top-down reactivation of sensorimotor representations (see e.g. Jeannerod, 2001). Or rather, it refers to a reactivation, during some "higher" cognitive process, of neural circuitry that is associated with perception and action (Svensson, Lindblom, & Ziemke, 2007). This circuitry is the shared neural mechanism behind embodied simulations, and the idea is that all cognitive processes involve simulations of sensorimotor processes to some extent (Barsalou, 1999). This has been argued to be the result of a redeployment of existing neural processes over the course of the evolutionary process (M. L. Anderson, 2007). On this view, it is the underlying simulation of related sensorimotor processes that leads to the apparent influences of action on behaviour (Grezes & Decety, 2001; Barsalou, Simmons, Barbey, & Wilson, 2003; Feldman & Narayanan, 2004). Simulation allows a decoupling of embodied representations from activity, but maintains that these representations are always sensorimotor in kind. Moreover, this also questions the differentiation into networks of perception and action, which appear to depend on each other throughout their course (Gallese & Lakoff, 2005).

2.2.1.1 Embodied Representations

On an embodied view, human cognition does not consist of amodal representations that bear arbitrary relations to their referents in the world. Rather embodied representations are thought of as activation patterns that include information from various sensory modalities (Barsalou, 1999; Barsalou et al., 2003; Barsalou, 2008, 2009). This is the sense in which Knott (2012) uses representation, and also the understanding of representation at work in this effort. In many ways, the embodied approach is a response to the symbol grounding problem, formulated by Harnad (1990). This raises the challenge of accounting for how words and concepts get their meaning. In this context, grounding meaning refers to the process of accounting for it in terms of non-symbolic, sensorimotor capacities. These are the means by which an agent interacts, by which she engages with herself and with the world she is embedded in. It is through these capacities that the agent experiences the referents of words. As the argument goes, the capacity for grounding through sensorimotor processes is a necessary aspect of any account of meaning and cognition (Cangelosi & Harnad, 2002).

In contrast to amodal symbols (see e.g. Mahon & Caramazza, 2008), embodied cognition builds on (multi-) modal symbols. These symbols always remain tied to the agent's experience. In the reach-to-grasp episode, as we will see, transient sensory states play a central role in all aspects from understanding the situation, to deciding which object to grasp, and to guiding the action of grasping itself (cf. Knott, 2012). Moreover, these transient sensory states are part of the representation that will ultimately serve as the basis for linguistic structure. That argument will be developed in the next chapter (Sect. 3). For now, the point with respect to mental representation is that it is a specific kind of symbol, and that the concept continues to play a role in this thesis. That way, I try to take a middle path between on the one hand i) the shortcomings of so-called "traditional", also known as "symbolic" cognitive science, and on the other hand ii) its over-zealous rejection by non-representational approaches to cognition.

It is worth noting though, as Machery (2007) points out, that subjects' performance in a number of the experiments interpreted in favour of embodied cognition can be accommodated by both modal and amodal explanations. Indeed proponents of an amodal view often do not deny involvement of modal representations (see e.g. Mahon, 2015). However, the same transient sensory states mentioned above emphasise two points that make it worth insisting on an embodied perspective. First, such representations cannot be autonomous from the perceptual system and from bodily action. In fact, central aspects of cognition may be inextricably linked to moment-by-moment dispositions of sensorimotor processes, such as eye movements and hand movements (D. H. Ballard et al., 1997). Second, these representations cannot be separated from the actions an agent takes based on them. Based on results such as those of Zwaan and Taylor (2006), it seems impossible to separate the cognitive and the motor aspects of an action, which appear to permanently and reciprocally influence each other (for an example concerning eye-movements during reading, see Myachykov, Garrod, & Scheepers, 2009).

Moreover, Barsalou (2009) goes on to argue that cognitive and perceptual processes not only share representations, but that a simulation of perceptual states plays a central role in cognition (see also Knott & Takac, 2016). On behavioural terms, there is evidence of an interference of actions with linguistic processing (e.g. Glenberg & Kaschak, 2002). In neural terms, this has been related to an apparent re-activation of sensorimotor areas during comprehension of action-related words (Pulvermüller, 1999; Hauk, Johnsrude, & Pulvermüller, 2004; Kiefer & Pulvermüller, 2012). To avoid drifting into another discussion of the respective merits and demerits of embodied or non-embodied views, note that one key point of this discussion is make clear what understanding of embodiment is at work in this effort. At this point it can be said that, throughout this work, mental representations are taken to exist, and to do so in a manner that depends on sensorimotor processes.

2.2.1.2 Shared Mechanisms Hypothesis

The requirement that meaning, and words, must be grounded somehow in order to have meaning is the basis for the assertion that language has to somehow interface with sensorimotor processes. In Knott (2012)'s approach, this interface is created by a set of cognitive processes that occur both during sensorimotor and linguistic processing. Such an interface would consist of shared mechanisms. This is a hypothesis about sentence representation, which predicts that the syntactic representation of some sentences can be understood as a description of the sensorimotor processes involved in the experience of what is described in the sentence. By the example of the woman who grabbed a cup, the claim is that an observer (or indeed she herself) understands the sentence only (though not exclusively) through the sensorimotor processes involved in the experience of her reaching to grab a cup (where a memory of this experience is obviously sufficient). That being the case, there is enough in common at the interface between language and sensorimotor processing to make it worthwhile to study them together (Knott, 2012).

One key point is that, if the shared mechanisms hypothesis holds, then a discussion of the processes of sensorimotor cognition is relevant to our understanding of language. Following this hypothesis, the two domains share some of their mechanisms, which means these mechanisms must account for aspects of both language and sensorimotor cognition. Knott (2012) suggests that shared mechanisms should account for some of the regularities in both domains, in particular. This goes beyond the relationship between sensorimotor processes and individual words that was discussed above (Hauk et al., 2004; Tettamanti et al., 2005). Instead, the object of the present work is sensorimotor correlates of the linguistic structures that govern how words can be linked together and how sequences of words get their meaning.

That such a set of shared mechanisms exists would be appealing from the perspective of an evolutionary account of language. In giving such an account, Hornstein (2009) suggests that linguists should posit as little specifically linguistic cognitive machinery as possible. Indeed, the idea that evolution works by reusing existing structures has been put forward as another fundamental organising principle of the brain (M. L. Anderson, 2010). In fact, their potential to establish this link between motor cognition and language is one of the central factors that motivated the early mirror neuron hypothesis (see e.g. Rizzolatti & Arbib, 1998). One's perspective on mirror neurons aside (cf. the debate in Hickok, 2009; Gallese, Gernsbacher, Heyes, Hickok, & Iacoboni, 2011), the idea that aspects of language supervene on sensorimotor mechanisms suggests a parsimonious account that appears consistent with what we understand of the dynamics of cognitive evolution.

Finally, one intriguing similarity between linguistic and sensorimotor structures is that they are both compositional (Knott, 2012). This means that their meaning arises out of the individual components plus the way they are put together. For example, sensorimotor processes are compositional in the sense that it is the interactions between many different processing streams that produce the actions we are aware of. In the cup grabbing episode, processes relating to attention, reaching, grasping, and to understanding actions to name just a few, all merge into an assemblage that we recognise as grabbing a cup. The action of reaching-to-grasp breaks down into components that stand in a certain relation to each other, as a result of which the woman is able to reach to grab the cup. Granting that this is the case, that sensorimotor cognition and language both break down into components, we face the possibility that they might do so in the same way. So, the suggestion is that the same mechanisms that evolved to subserve sensorimotor cognition are reused and adapted as the basis for language. Specifically, we are interested in the structure of language (see Sect. 1.2).

2.2.2 Embodied Cognition and Related Concepts

This section attempts to clear a small patch of the murky terminological waters, but only in its immediate conceptual vicinity. That means, for example, that it forgoes a discussion of what has been come to be known as "4E Cognition" (referring to "embodied", "embedded", "enactive", and "extended" cognition; for an introduction, see Newen, De Bruin, & Gallagher, 2018). Instead, I briefly outline how the view of embodied cognition advocated for here stands in relation to i) traditional cognitive science, ii) grounded cognition, iii) situated cognition, and iv) enactivism. The comments in this section are kept very short. Their only purpose is to clearly position this thesis within the conceptual landscape of cognitive science.

Doing so, I take an explicit position regarding each of the four point listed above. Respectively, the perspective on embodied cognition adopted by this thesis can be summarised as follows, namely i) partially opposed to traditional cognitive science, ii) localised within grounded cognition, iii) seen as an approach within situated cognition, and iv) as partially opposed, again, to enactivism.

2.2.2.1 Embodied as opposed to "Traditional" Cognitive Science

Granted, there continue to be principled (and justified) doubts about the idea of an embodied cognitive science (Goldinger, Papesh, Barnhart, Hansen, & Hout, 2016). Now, these typically follow one of two routes, by either (i) rejecting the significance of empirical findings that support an embodied account in opposition to "traditional" cognitive science, or by (ii) rejecting the interpretation of results as evidence for an embodied account (R. A. Wilson & Foglia, 2017). Concerning words and the meanings of concepts, for example, Mahon (2015) insists that the embodied and non-embodied views are no longer coherently distinct theories. This is an important challenge to meet. The embodied approach has certainly contributed novel, and unexpected findings. But, a number of self-declared proponents of embodied cognition also insists that these results are not incompatible with many aspects of the "traditional" view (e.g. Shapiro, 2011). The present work insists there is strong evidence in favour of an embodied view, and takes up the need to offer further evidence in favour of an embodied account that can be differentiated from a non-embodied perspective.

2.2.2.2 Embodied as one approach within "Grounded" Cognition

Simply put, embodiment is one kind of grounding (Barsalou, 2010). On the other hand, this may be misleading, since the conceptual commitments of radical embodied theories extend well beyond those of grounded cognition, for example, on the subject of representations. Perhaps, whereas grounded cognition is the broader framework, embodied cognition can be thought to make more detailed claims. For example, the generic grounded cognition perspective does not necessarily emphasise embodiment, and may even adopt a focus on cognition 'independent of the body' (Barsalou, 2008). In this regard, the present work takes place in the more narrow focus of embodied cognition. Note that, in doing so, this effort leaves out other aspects that do play a role in grounded perspectives, such as the role of social context (see e.g. Agre, 1997a; Steels, 2003). In fact, this picks up on one major limitation of Knott (2012)'s approach, namely its individualistic perspective on language learning (more on this in Sect. 3.4).

2.2.2.3 Embodied as one approach within "Situated" Cognition

Embodied perspectives constitute one set of perspectives within the wider umbrella of situated cognition. The 'specific 'situatedness" of the embodied perspective emphasises that cognitive processes depend on, and occur in and through a body of which the brain is just one organ. According to R. A. Wilson and Foglia (2017)'s survey, this role of situatedness means embodied cognitive science can be distinguished from both the study of embedded cognition (e.g. L. A. Suchman, 1987) and the thesis of extended cognition (e.g. Rupert, 2009). Theoretically, embodied situatedness has been distinguished from an emphasis on how an agent may offload cognitive processes onto its environment (embedded cognition), an environment that we may have reason to think of as part of that agents cognitive system (extended cognition). As much as they may meld together in practice, distinguishing these perspectives in theory is a crucial means of articulating their commitments and assumptions.

So, for example, the cup-grabbing episode discussed in this work makes little explicit reference to offloading cognition onto the environment. Moreover, various points of the proposed account refer to a mental representation of spatial locations and other features. On the surface, this may seem at odds with the embedded perspective. Then again, there does seem to be some being embedded going on, since many of the representations discussed later include only task-relevant information (see e.g. D. H. Ballard et al., 1997), while task-irrelevant information is acquired through attentional processes from the environment when it is needed. Where the present views differs from embedded cognition is in its relatively strong emphasis on brain and computation as factors in the explanation of how the investigated behaviour is produced (though for a neuroscientific embedded cognition perspective, see Haselager, van Dijk, & van Rooij, 2008). This is in contrast with the equal emphasis on brain, body and world in the embedded view (e.g. Richardson, Shockley, Fajen, Riley, & Turvey, 2008).

2.2.2.4 Embodied as opposed to "Enactive" Cognition

One form of strong embodied approach is enactivism. This is a diverse field in itself, with different, largely consistent, but partially exclusive perspectives (R. A. Wilson & Foglia, 2017). Within this field, there is also at least one distinct understanding of embodiment (see Varela et al., 1991). The present work picks up only one of the aspects of this definition, namely that the cognitive capacities of an agent depend on having a body with various sensorimotor capacities. This view is also reflected in the sensorimotor version of enactivism (Degenaar & O'Regan, 2017), with which the present work is consistent to some extent. The important difference at this point is enactivists' rejection of computationalism and mental representations. Note, that this rejection of information processing has been called into question even from a sympathetic perspective (cf. Clark, 2006b). Meanwhile, sensorimotor enactive and active perception approaches drawn on in this effort (e.g. D. H. Ballard, 1991) share an emphasis on the active exploration of the environment. Where they differ, is regarding the role of motor activity. This might be said to be more instrumental (in service of sensation) in the present approach, while being on equal footing with sensation in enactive accounts

Another reason to stop short of a strong embodied approach with this thesis rests with the conceptual commitments of enactive approaches. For example, the concept of embodiment at work in Varela et al. (1991)'s enactive approach builds on the idea of *structural coupling* (Maturana & Varela, 1980). While in the ecological camp of the strong (or 'radical' Chemero, 2009) embodied cognition theories, roots are traced to efforts in theoretical biology (von Uexküll, 1928), as well as to Gibson (1966)'s concept of *affordances* (see Box. 2.2). Again, while both of these views and all three of these concepts offer important challenges and fascinating insights, I expect each to be less successful in dealing with the challenges of this project than the weak embodied view I have adopted. Simply put, remaining consistent with Knott (2012) requires a weak embodied perspective, while a stronger or enactive position would demand a thorough reformulation and reinterpretation of many (at least, if not all) of the studies upon which this thesis is built. Finally, as far as this thesis is concerned, the utility of a weak embodied view outweighs the concerns raised by enactivism (for a contrary perspective, see Van Elk, Slors, & Bekkering, 2010).

So, for example, the understanding of affordance at work in this thesis is more limited than in an enactive view. It is more similar to other applications of the term, in which practical use has led to some degree of divergence from the strict Gibsonian notion (for a discussion of this problem in the field of robotics, for example, see Horton, Chakraborty, & Amant, 2012). To clarify, then, the particular sense of affordance at work with Knott (2012) is the "motor affordance" of graspable objects (for a computational implementation, see Fagg & Arbib, 1998; Murata, Gallese, Luppino, Kaseda, & Sakata, 2000). The idea is that affordances parameterise the sequences of basic motor commands that make up an action like grasping. That means that default motor sequences are adjusted based on information about the current target object. An example is hand preshaping, which describes the automatic adjustment of hand shape early during a reach-to-grasp movement (seminally by Jeannerod, 1981; M. Arbib, 1981). The key point from this, in relation to affordances, is that hand preshaping reflects a fundamental relation between agent and environment that is present throughout the cognitive processing of reaching to grasp (for evidence that hand preshaping is hardly affcted by vision during the action, see e.g. Winges, Weber, & Santello, 2003). On this issue, the weak embodied view of this thesis shares the impetus of enactivism, but does not take it as far.

Box 2.2

Affordances

Gibson (1966) introduced the "affordance" as a concept that refers to what the environment offers an animal. In the original account, affordances are relations between the animal and environment that may be specified by anything from the animals anatomy, to its posture or goals (Soylu, 2016). Thus, affordances are necessarily relative, emergent out of the interaction between animal and environment (Gibson, 1986). And affordances are learned relations, specifying the particular meaning something has for the agent in question. The cup from the example, for instance, affords being grasped in a certain way, to a certain kind of animal that has an opposable thumb. Ecological approaches, just as the perspective of this thesis, are criticised from an enactive perspective for keeping motor activity as an instrument of sensation (Mossio & Taraborelli, 2008).

2.2.3 Computationalism

The present work stands in the computationalist tradition in cognitive science. In this view, cognition is best studied in terms of representational structures and computational procedures that operate on those structures (Thagard, 2005). This rests on a complex analogy between mind, brain and computers. More recently, it has come to include more of the body as well (for discussion, see e.g. Pezzulo et al., 2011; Allen & Friston, 2018). Indeed, there is a prominent research direction within embodied cognition that continues to emphasise both computation and mental representations (e.g. Barsalou, 1999). Concurrently, there are equally prominent approaches within embodied cognition that remain sceptical not only of mental representations, but also of computationalism, though perhaps less so (Boden, 2008, Chapter 16, among others).

In this section, I will outline the computationalist position taken in the present work. The role and character of computation varies across authors in cognitive science, so it is useful to be clear about the understanding at work here. To begin, I suggest that a computational theory of mind is not antithetical to the embodiment of cognition. Indeed, examples abound in which computational models take the interaction of body and environment into account (see e.g. Feldman & Narayanan, 2004; Knott, 2017b). In both, the underlying argument is that computational theories of mind can incorporate the insights offered by an embodied perspective. Such authors maintain that a computational perspective on cognition continues to offer the best overall explanatory framework.

We must be careful though, because not every kind of computationalism can be reconciled with the principles of embodied cognition. For example, embodied theories stand in opposition to a thorough detachment of cognition from the embedding environment. Here, it is not necessarily symbol manipulation that is the problem, though the amount of emphasis that should be placed on symbolic operations does remain contested (e.g. Schmolesky et al., 1998). Instead, the challenge concerns the role in which body and environment causally or constitutively shape mental activity. Concerning mental computation, the point is that this must be cast wider, to include agent-environment dynamics (e.g. Thelen & Smith, 1996).

In contrast to an eliminativst attitude towards mental computation, Knott (2012) assigns it a central role. In this view though, mental computation is always tied to sensorimotor processes. Those processes, in turn, are necessary antecedents of cognition. Recall that the basic computationalist claim is that information-processing is necessary for cognition (Miłkowski, 2018). That is not to say that this it is sufficient (Newell, 1980). The following section will address a few common objections to computationalism in order to show the position is not as vulnerable as is prominently argued (by many enactivists, for example, such as e.g. Chemero, 2009).

To begin, note that I concede that some of the simplifications that arise out of an emphasis on information processing and on computational modelling of cognition are misleading. However, this is not an argument against computationalism. All models are, by definition, simplifications, and simplifications will always be misleading about some aspects of the original phenomenon (see Sect. 2.1.1). Contrary to arguments expressing as much (e.g. Daugman, 1990), calling computation a metaphor also does not entail that computational models are of little use for studying and coming to understand cognition (for a related discussion of metaphors in relation (working) memory, see Box 2.3). This recalls the issue of levels of analysis. True, computational approaches at best constrain our search for the biological basis and neural correlates of cognition (cf. Friston, 2009). An example problem from the present work is the computational *unit*. In Knott (2012), *units* are not individual neurons but functional units, be they local cell assemblies or distributed activation patterns or any other form of functional grouping. Neuromodulation, synchronisation and neural noise all play a role in the same way that neural firing does. These aspects are not specified on the level of computational description or modelling. Nor can they be specified on then computational or algorithmic level. Then, though, there does not appear to be argument against computationalism here.

What computational approaches do offer is a description of what kind of processes cognition may be based on. That, after all, is the functionalist aspect of computational functionalism. Here then, it is important to note that, despite the abstract descriptions computationalism offers, the analogical computers themselves are also physical things. Computers do things in order to manipulate symbols, some of which are not computational (Miłkowski, 2018), like consuming energy, for example. Again, computationalism and embodiment do not appear to be mutually exclusive. Returning to levels of analysis, it is not an argument against computationalism to say that it does not give a complete account of all aspects (levels) of cognition. For, computationalism does not give a complete account of all aspects of computers, either – it's claim was never to explain everything.

The theme is that computation is necessary, not sufficient for cognition. Making that point alone removes the opposition between computational views and some approaches that emphasise that cognition is the result of unique biological structures or processes (cf. E. Thompson, 2010). Granted, there may be features, present in (some) organisms, that are necessarily absent in computers. However, allowing this does not mean cognition cannot be described in computational terms. It means that cognition is not reducible to computation. To the same effect, bodily structures beyond the neural system have been implicated in cognition and in a computational account of cognition (Miłkowski, 2013; Nowakowski, 2017). In fact, the computationalist view advocated for in the present work demands that cognitive processes are understood as being situated, embedded in agent and environment (Miłkowski, 2017).

Box 2.3

Storage Metaphor of (Working) Memory

Another prominent metaphor in this thesis is the image of storing a representation in (working) memory. Storage, however, was a doubtful metaphor from the beginning (Roediger, 1980), and is now an outdated one at that (Harré, 2002). Embodied accounts of memory offer alternative perspectives (Engelkamp, 1998; Sutton & Williamson, 2014) This view is based on data indicating that imagined actions facilitate remembering, that the external location is an aid to memory (Cole, Hood, & McDermott, 1997) and that the imagined spatial layout of a scene constrains the time taken to remember (e.g. Bryant & Geoffrey Wright, 1999). From other research, it appears that the body posture is a cue for memory (Dijkstra, Kaschak, & Zwaan, 2007). On this basis, memory is not process of abstract information storage an retrieval. Instead, it can be understood as (partially) an embodied skill (Sutton & Williamson, 2014). So, memories are not stored to be recalled, as much as they are constructed online, in the context of a specific action and situation. In doing so, it makes reference to the agent's body and situation. Here, again, there is no contradiction with computationalism, as indicated by computational accounts of embodied memory (e.g. Barsalou et al., 2003).

There are many of challenges to computationalism. Harnad (1990)'s symbol grounding problem, for example, is another. This is related to the problem of semantics in computers and Searle (1980)'s Chinese Room (for a discussion, see Miłkowski, 2018). One more point to note in this context is the claim, out of ecological psychology (e.g. Gibson, 1966), that humans do not process information. Responding to this, Miłkowski (2018) points out that even accounts of 'direct' awareness of ones environment make use of a minimal notion of information. Sensitive to this rebuttal, Chemero (2009) argues that while information does play a role, cognitive system do not process that information. While it is unclear what this means, that may be the case under some definitions of information processing. By this move, more radical accounts seek to reject computationalism outright. This thesis does not go as far. Indeed, I have argued in this section for an understanding of information processing that is compatible with embodiment. Since, one the one hand, many prominent objections are against straw-man versions computationalism, as has been shown. And on the other hand, justified opposition is often only partial. So, while many strong embodied accounts appear to be irreconcilable with computationalism, I conclude that weak embodied perspectives, such as the one adopted here, are not.

2.2.4 Minimalism

The Minimalist theory of syntax is a later formulation of Chomskyan universal grammar (Chomsky, 1965). Its basic feature, from the perspective of this thesis, is that it distinguishes two syntactic levels: (i) a phonetic form (PF), and (ii) a logical form (LF). PF describes the surface characteristics of a sentence as a sequence of words. This varies across languages, for example across linguistic typologies (e.g. Subject-Verb-Object, or others; see Tomlin, 1986) LF, meanwhile, describes how that sentence is constructed and why it means what it does (Knott, 2012). LF is taken to be universal, both in the original interpretation and by Knott (2012).

Now, in Knott (2012) already, we find a rejection of Chomskyan modular LF (and Fodorian modularity too, cf. Fodor, 1983), in favour of a sensorimotor interpretation. That means this account does not rely on any language-specific machinery in order to ground the underlying structure of language. Instead, I argue that there are shared mechanisms, originally sensorimotor in kind, that are taken to be mostly invariant across speakers of different languages. Those different languages exists, because any number of surface structures (PF) can have developed based on the same underlying structure. Note the assumption that individual and culturally-mediated differences in the sensorimotor processes are small enough to be negligible.

What Knott (2012) retains from Minimalist theory is precisely the idea of a universal underlying structure. The point of the sensorimotor reformulation of that proposal is that the syntactic structure (LF) of a transitive sentence, like *the woman grabbed a cup*, can be thought to describe the sequence of sensorimotor operations involved in experiencing the event. This approach is appealing because, if the shared mechanisms hypothesis (see Sect. 2.2.1.2) is true, we should expect a hidden layer of language processing that is rooted in sensorimotor processing.

In this regard, interesting data were reported by Gershkoff-Stowe and Goldin-Medow (2002) and Goldin-Meadow, So, Özyürek, and Mylander (2008). According to their interpretation, the linguistic typology Subject-Verb-Object (SVO, see Tomlin, 1986) is a “natural” sequence in which to represent transitive actions nonlinguistically, even by speakers with different word order. This is interesting because that is the same thematic sequence, namely agent-action-patient (see Box 2.4), that Knott (2012) imputes in the series of sensorimotor processes that are involved in reaching to grasp a cup. As such, this might be taken as a first indication there appears to be some default order to how humans think about transitive actions.

Box 2.4

Subject, Verb and Object vs. Agent, Action, and Patient

The sentence “the woman grabbed a cup” is separated into three components at times throughout this thesis: the woman (subject / agent), the cup (object / patient) and “grabbed” (verb / action). To be clear though, the terms originate with different lines of research in linguistics. “Subject-Verb-Object” is a linguistic typology, the study of which investigates similarities and differences of word ordering, and the consequences this has (Whaley, 1996). “Agent-Patient-Action”, on the other hand, is a characterisation in terms of thematic relations (see e.g. Fillmore, 1969). These refer to the semantic roles of a noun phrase, usually in relation to a sentence’s main verb (Jackendoff, 1992) – for example, one can think of the woman as doing the action, while the cup is thought of as being acted upon (by the woman).

One final point of (Knott, 2012)'s is that his is not an attempt to reduce syntax to semantics. The sensorimotor interpretation of Minimalist theory is not a semantic interpretation. So, LF is not taken to encode semantics, but the processes involved in generating a semantic representation. On the sensorimotor side, for example, the discussion will focus on the principles and demands that produce a canonical sequence of operations for actions like reaching to grab a cup. This concerns the (temporal) structure of sensorimotor processes, and how the sequence in which operations occur constrains the representations involved. Something similar is true on the linguistic side, where we are also concerned with an underlying structure of language. The idea is that some features of syntax reflect contingencies in sensorimotor processing. That means that syntax carries information about the mechanisms by which we construct our world. This is different from information about the world itself, which is semantics.

Starting from this sensorimotor interpretation of Minimalist theory, Knott (2012) addresses two related questions: (i) what is the interface between language and the sensorimotor questions and (ii) how does syntax work? But, while Knott (2012) is always also concerned with how syntax works, the main focus of this current thesis is on the interface between language and sensorimotor processes. As a result of this, the term syntax is not adequate for what is investigated here. For that reason, linguistic structure is used instead.

Indeed, after this section, the present work will abandon the term "syntax" entirely, in order to avoid being dragged into neither the persistent debate in linguistics between the pro and contra-Chomsky factions, nor the evaluation of competing theories of grammar (for two contenders with Minimalist theory, see Jackendoff, 2002; Pollard & Sag, 1994). Again, the reason why the Minimalist perspective on syntax is important nonetheless, is because Minimalist theory makes the strongest claim about underlying syntactic universals. That is appealing, given that the sensorimotor apparatus of humans is taken to be invariant. Based on Minimalist theory, Knott (2012) develops a mapping between the universals of both sensorimotor cognition and natural language syntax. Based on Knott (2012), this thesis replaces syntax with the idea of underlying regularities in "linguistic structure" that are not tied to the formalisms of Minimalist theory.

2.2.5 Relation to the History of Cognitive Science

When I explain that the modern English "context" comes from Latin "contextus" – formed by joining "con" and "texere" – meaning "weave together", I have conjured a nice image but contributed little substance. In the same way, pointing out that events occur(ed) in a context is true by definition, but not very meaningful by itself. To make that point something other than trivial, an argument has to be made how a given context produced an event we are interested in, and what we gain by apprehending this. Only then do we actually examine the phenomenon in context, as opposed to simply pointing out that it occurred "in context". In fact, it may be more interesting to forgo the container-like context altogether and consider, instead, how certain events are related to one another (L. A. Suchman, 1987).

In this sense, one way thinking about history is useful, is when it helps us understand the present. Concerning the role of linguistics in cognitive science, for example, it is illuminating to see its decline in relation to a distinction between competence and performance, (cf. Jackendoff, 2003, pp. 652) introduced by Noam Chomksy during his mid-1960's retreat from empirical psychological investigations of generative grammar (Boden, 2008). For, until the publication of Chomsky (1965)'s "Aspects of the Theory of Syntax" , psychologists had interpreted Chomsky's claims as hypotheses about actual mental processes. This changed with

Chomsky's withdrawal into mathematical linguistics, which he justified with the competence/performance distinction. One effect of this distinction was to isolate a certain brand of linguistics from psychology. As a result, this line of linguistic research has faded from the centre of cognitive science, despite the fact that it was early work of this kind that set psychologists on a path to consider language in computational terms in the first place (Boden, 2008). I point this out, because cognitive linguistics (e.g. Lakoff & Johnson, 1980) arose, in part, out of the effort to bridge the gap that was produced through that competence/performance distinction.And with it came the groundwork for work like (Knott, 2012)'s and this thesis.

When that distinction was revised, the essential differentiation remained intact (see Chomsky, 1986). In contrast, the original formulation has always been seen critically by many (e.g. Lakoff, 1973; Milroy, 1985). The gist of that rejection is that knowledge of language is derived from patterns of language use (Evans & Green, 2006). Knott (2012) argues in the same vein. Here, the underlying linguistic "competence" of all humans is reflected in the structuring role of sensorimotor processes. These include limitations of memory (Knott & Takac, 2016), attention (Walles, Robins, & Knott, 2013), and articulation (Takac et al., 2012), as a result of considering which all aspects of grammar become a question of performance (Labov, 1971). In this sense, Knott (2012) is an example of recent work that recalls 1960's efforts to interpret the concepts of grammatical theory in relation to actual mental processes (e.g. G. A. Miller & Chomsky, 1963). However, the discrete language module has been replaced with a mechanism for language that is based on sensorimotor processes and shared with them. In doing so, it establishes an embodied account of linguistic structure (Knott, 2017b).

On a wider stage, it is worth pointing out one more historical relationship namely the role of machine translation, as a precursor to computational linguistics and factor contributing to the development of contemporary cognitive science. At the outset of cognitive science, work on machine translation set in motion computational efforts concerned with language (Boden, 2008) for examples of its influence, see Hutchins (1999). This was before the 1956 Dartmouth Summer Research Project on Artificial Intelligence (McCarthy, Minsky, Rochester, & Shannon, 1956). In fact, it was also long before Longuet-Higgins (1973) coined the term "cognitive science". But, incidentally, that term was coined in comments on a thoroughly negative report of the state of AI in the United Kingdom (Lighthill, 1973) that came as a result of the disappointing progress with that self-same machine translation, as well as with computer language systems in general. Now, in many ways, the present work continues the research tradition that emerged to answer the challenges to our understanding and our study of language that emerged at this time. For example, we draw on the framework proposed by Knott (2012) as part of an ongoing shift towards understanding the role of multimodal grounding for language acquisition (e.g. Suanda, Smith, & Yu, 2016), and towards implementing models that can do so (see e.g. Pezzulo et al. (2013) – see also the discussion in Sect. 6.3.1).

It may appear strange to refer back to machine translation at this point. However, the statistical or rule-based modelling of natural language and its study in computational terms was pioneered in machine translation. Through efforts in this field, it became apparent that human language was more difficult to model than had been assumed. Recognising this, efforts to automatically translate languages led to work that considered grammar, morphology, syntax, then lexica, pragmatics and semantics, too (Boden, 2008). To do so, expertise had to been drawn together from disciplines including linguistics, computer science, and psychology. So, machine translation was not only the subject concerning which cognitive science was named (by Longuet-Higgins (1973) in his defence of machine translation), but also an impulse for both the broad scope and the multidisciplinary character of cognitive science, which is taken up in this thesis.

Chapter 3

A Sensorimotor Model for Sequences of Actions

> "The woman grabbed a cup."

Though it is an integrative effort, the present work was mainly inspired by the idea that linguistic structure is derived from the structure of sensorimotor cognition. The temporal organisation of sensorimotor processes specifically, is put forward as the basis of the underlying structure of language. That structure of language refers to fundamental patterns observed across human languages. These patterns are taken to exist below the variety of linguistic typologies that characterise different languages (e.g. Subject-Verb-Object (SVO), see Tomlin, 1986). In contrast, we are concerned with patterns across languages that explain, not how to construct the correct sequence of words, but how and why such a sequence means what it does.

Knott (2012) proposes one model of how this happens. In that account, concrete action episodes (see Box. 3.1), like reaching to grab a cup, are made up of a default sequence of actions (see also D. H. Ballard et al., 1997). Unsuprinsingly then, Knott (2012)'s account begins with a summary of the sensorimotor processes involved in reaching to grasp a cup (see Chapter 2, in Knott, 2012). That summary concludes by suggesting there is a canonical temporal structure to the actions involved in an episode, which means that the sensorimotor process involved occur in a specific order. Of the ordering constraints upon this sequence, some must hold, while others only usually hold. That this sequence is "canonical" or a "default" also suggests, though, that a mature agent can depart from it when performing the action. [1] Some constraints, Knott (2012) suggests, must hold at some point during development, but may be more flexible thereafter.

Box 3.1

Actions and Episodes

Action: is a unit of analysis that refers to some one thing an agent does. In the present work, the concept operates on various levels of granularity. For example, it includes the automatic and unconscious recall of a mental representation as much as it includes a voluntary and conscious reach to grab a cup. Thus, "action" is a general functional concept, usually in terms of the outcome of some set of processes. An action at a higher level of abstraction can be composed of a sequence multiple actions at a lower level. The decomposition of the action of reaching to grab a cup into its component actions is a prime example.

Episode: refers to a sequence of actions, such as those grouped together by the description that the woman "grabbed a cup". The reach-to-grab "episode" includes all actions that are immediately involved in the agent's doing so. The term is carried over from Knott (2012), where it has the same meaning.

[1] Concerning action observation, rather than action execution, it means that the agent can infer that she is observing a reach to grasp action based on sensorimotor processes that do exactly correspond to the default sequence.

For developing the framework further, the default sequence is key. That is because the reliable sequential order of sensorimotor processes is what is used to structure episode representations. These are expected to be established early enough during development that the actions involved will still be represented using the canonical sequence. Having been generated out of this sequence, Knott (2012) proposes that episode representations are prepared sensorimotor routines (Takac & Knott, 2015). That is to say, the episode representation is a prepared sequences of attentional and motor operations (see Box 3.5). As such, the episode representation reflects the entire sequence of actions that make up an episode, including their order (for a further discussion, see 3.3).

Next, the central claim is that the episode representation can be replayed through working memory (for a discussion, see A. Baddeley (2000) and Sect. 3.3). Since it draws on the episode representation, this processing of replaying mirrors the sequence of attentional and motor actions that makes up the sensorimotor sequence. In fact, both the processes of performing the action and the replay process follow the same general schema (see Table 3.1). [2] Moreover, they proceed through the same sequence of actions, which is referred to as "sensorimotor sequence" (or "routine") regardless of whether it is generated by replay through working memory or by any other mechanism.

Finally, the reason Knott (2012) builds up this account of episode representations is that it is this representation that he puts forward as the basis of linguistic structure. So, linguistic structure should be thought of as a description of the replay processes discussed above. Since the replay process mirrors the performing of the action, this means that the linguistic structure of a concrete sentence, like "the woman grabbed the cup", can be thought of as a kind of summary (a "description" Knott, 2012) of the sensorimotor processes involved in performing that action (Knott, 2014). This is not a claim about the sequence of words in an utterance. Nor is it a claim about the processing involved in producing (or interpreting) the utterance "the woman grabbed the cup". Instead, this is a proposal about a shared mechanism, based on which the linguistic structure of a sentence can be grounded in sensorimotor mechanisms. That shared mechanism is captured by a sensorimotor representation of the entire action episode, replayed in working memory. This is the episode representation.

Table 3.1: Each cycle begins with an initial context (the specific notion of "context" is discussed in Sect. 3.3). Given the goal of performing a particular action, and the intention of doing so in a particular way, this triggers a sensorimotor action. This, in turn, triggers a reafferent sensory signal. The action and the reafferent signal together establish the next context. From this point, the cycle is repeated as long as agent is pursuing the same goal by the same means.

Context signal	Action	Consequence
initial context		
	sensorimotor action	
		reafferent sensory signal
next context		

[2] Both the episode and the replay process can be thought of as a process of cycling through this general schema. One mechanism that plays a central role in the cycle is reafference. When an agent performs an (sensori)motor action, reafferent signals reflect the perceived changes following that action (for a discussion of the idea, see e.g. Grüsser, 1995).

Outline of This Section

Above, this chapter gave a general summary of the sensorimotor model of linguistic structure. Necessarily, there is much to be clarified. In the remainder of the chapter, I consider individual aspects one after another. To do so, I will go back to embodiment (discussed in Chapter 2.2) and gradually build up to (Knott, 2012)'s model. The span of the explanation is broad, so the impatient reader is encouraged to skip to Sect. 3.2 to keep reading about the model immediately.

For, before delving into Knott (2012)'s model properly, a number of points require clarification. Specifically, Sect. 3.1 provides an overview of the concept of attention and of the role it plays in the present work. Building on this, I outline D. H. Ballard et al. (1997)'s framework of "deictic routines". This framework is a central aspect of this thesis, because it suggests a mechanism for how embodied attention shapes cognition. Only based on this introduction, does Sect. 3.2 summarise the framework proposed by Knott (2012).

To do so, Sect. 3.2 presents the big picture first. That means I briefly outline the entire sensorimotor sequence of reaching to grasp. Afterwards I discuss some key points in more detail. Nonetheless, they only provide an more in-depth summary. For a detailed version of the argument, the reader is referred to Knott (2012, 2017c). In order to focus on the most important aspects, many aspects of reaching to grasp will also not be discussed, either in detail or at all, despite the fact they are relevant. These include, in no particular order, object classification, conversion of visual signal into a stable (non-retinotopic) coordinate system, visual search and target selection, action categorisation, motor planning and many more – (for summary, see Ch.2 in Knott, 2012). [3] Instead of addressing all of these aspects, I focus only on a small number of essential processes, as well as those that could be addressed empirically within the experimental set-up we designed. To do so, the summary of sensorimotor processes that follows below is interleaved with short discussions of the key processes that are relevant to a specific aspect of the episode.

Next, following the overview of the sensorimotor processes involved, I briefly introduce the memory mechanisms specified by the sensorimotor model (see Sect. 3.3). In addition to this, I offer a few more comments on the interface between sensorimotor mechanisms and linguistic structure (see Sect. 3.3.1). Then, I conclude the overview of the sensorimotor model by summarising it into a nutshell, for ease of reference during the rest of the work (see Sect. 3.3.2). Finally, the last section of the chapter offers a short discussion of advantages of and objections to Knott (2012)'s proposal (see Sect. 3.4). Before, at last, I round out that section by summarising the concrete predictions this model allows us to make concerning the data collected here.

3.1 Embodiment and Attention

To begin where Sect. 2.2.1 left off, note that (Knott, 2012)'s framework builds on a particular understanding of embodiment, drawn from D. H. Ballard et al. (1997)'s framework of deictic routines. There, embodiment refers to constraints imposed upon cognitive processes by the physical and physiological limitations of the human body. This reflects a level of analysis, the "embodiment level", on which cognitive processes reflect the temporal constraints of motor actions, like saccades or fixational eye movements (D. H. Ballard et al. (1997); see Box 3.2). At this level of abstraction, the appropriate model of computation should be expected

[3]Indeed, this thesis (like Knott, 2012) does not claim to offer a complete account of the cub-grabbing episode. L. A. Suchman (1987)'s point holds here too, that a complete account of any behaviour is a conceptual and practical impossibility. Simply put, this is because the list of all component actions is not enumerable.

to be different from those at shorter and longer timescales (Newell, 1994). The basic idea about the level is that body movements at the embodied timescale (circa 1/3 second) provide an essential link between elemental perceptual and "higher-order" cognitive processes (for a review of temporal constraints on neural processing, see e.g. Pöppel, 1994).

On this embodiment level, the current state of the body is essential for neural information processing. Schematically, the current orientation of the body – where the agent is looking, for example – can point toward some aspect of that location of the scene. This picks up the notion of a "pointer" (Pylyshyn, 1989), introduced as an abstract representation of spatial locations independent of their features. Usefully, one aspect of pointers is that, if sufficiently few are required to solve a task, then that task can be accomplished by explicit pointing (e.g. a fixational eye movement), or "neural" pointing (D. H. Ballard et al., 1997), in the form of certain forms of attention (see Box 3.4).

Box 3.2

Saccades and Fixations

Briefly, a voluntary saccade is rapid eye movement, initiated voluntarily, that suddenly and quickly changes the area of fixation (J. E. Hoffman & Subramaniam, 1995). The term "visual fixation" refers to fixational eye movements, periods during which the eyes appear to be immobile (though they actually move constantly), while visual information is being collected and processed (Rucci, McGraw, & Krauzlis, 2016). Such periods can indicate visual attention (Weichselgartner & Sperling, 1987). In light of this, the voluntary saccade between fixations has become associated with the shifting of visual attention (D. H. Ballard, Hayhoe, Li, & Whitehead, 1992; Webb et al., 2010).

3.1.1 Attention

Attention plays a central role in Knott (2012). One way is by making whatever aspect is being pointed to available for other cognitive processes (D. H. Ballard et al., 1997). This occurs via a transient representation of that feature. So, for example, the generation of such transient representations via attention appears to be an essential regularity of action preparation (Averbeck & Lee, 2007). Performing an action, then, has been argued to draw on a constant process of updating, based on the current relation between agent and environment, that also depends precisely on transitory, and sensorimotor representations (e.g. M. Arbib, 1981; Thelen & Smith, 1996). This suggests that computation is limited to what is needed for the current task. From that vantage point, attention need not be thought of as a limitation on processing capacity, but as a necessary aspect of action. This gives the selectivity of attention a computational rationale (D. H. Ballard et al., 1997).

This also means attention is key in the present framework, because it is attentional processes that generate the transitory representations that subsequent processes operate upon. It is important to be clear about what this means; so, there are three ideas at work behind this claim, namely (i) that orienting in physical space, by moving one's eyes, for example, simplifies cognitive processing (conceded even by opponents of the pre-motor theory of covert attention, which stresses this point; see e.g. Smith & Schenk, 2012), (ii) that important information is drawn from sensory processing when it is needed, rather than being recalled (D. H. Ballard et al., 1997), and (iii) that sensorimotor processing is functional, meaning that it is built primarily on task-relevant representations (D. Ballard & Rao, 1995).

Moreover, attention is not unitary. Indeed, there are different modalities of attention (see Box 3.3). These are each involved at different points during the reach to grab episode and contribute different information to the process of reaching to grab a cup (Knott, 2012). This is the result of the timing of their contribution, as much as it is the result of selective inputs and processing. For example, it can be key that the agent looks at something at one time and not at another. So, for example, participants performing the task reported in D. H. Ballard et al. (1997) reliably acquired different information (e.g. location-specific, or colour-specific information) at different points during the task. Indeed, their actions during the task appear structured by the sequence of visual operations. This is not a claim about a default sequence, yet, but one about the central role of attention. Namely, shifts in attention i) simplified the task, ii) occurred at specific moments in time, and iii) were shaped by task demands. [4]

Box 3.3

Modalities of Attention

Visual: selective visual attention, for the present work, should be understood exactly as the ability to quickly direct one's gaze to important objects in our visual surroundings (Itti & Koch, 2001). The model of attention drawn on in the present work (e.g. Koch & Ullman, 1987), compromises between early (Treisman, 1991) and late selection (Duncan & Humphreys, 1989), by including two levels, one for each type of process. In this study, we focus on the processes on the second, selective level, where attention is considered focused. For the study, we used eye-tracking to access this form of visual attention.

Motor: operates according to the same principles as visual attention, but in the proprioceptive modality. Evidence for motor attention comes from studies of action preparation, for example (M. F. Rushworth, Nixon, Renowden, Wade, & Passingham, 1997). Like visual attention, this process generates a stable neural representation, formed through parietal sensorimotor circuits (M. Rushworth, Johansen-Berg, Göbel, & Devlin, 2003). Using motion capture, especially of the hands and arms, we are able to cover aspects of this modality, too.

Haptic: is based on mechanoreceptors, sensory neurons that respond to mechanical pressure (Lederman & Klatzky, 2009), embedded in ligaments, muscles and tendons (for a discussion, see e.g. Hajnal et al., 2007). In this work, haptic attention is exclusively covert. Haptic attention is put forward as an important feedback indicating that the agent is grasping the object at the end of a reach-to-grasp episode. In the absence of neuroimaging data, this modality could not be represented here.

Besides different modalities, different types of attention are also relevant (e.g. "overt" vs. "covert", see Box 3.4). [5] Like the different modalities of attention, overt and covert attention may also generate functionally equivalent representations in whichever sensory domain (Chun, 2011; Chun, Golomb, & Turk-Browne, 2011). Indeed, there is evidence for a broad overlap in the neural circuitry associated with covert and overt attention (Nobre, Gitelman, Dias, & Mesulam, 2000). This suggests that two processes that appear distinct

[4] Note that the function of visual attention at one point in time may be as different from that of visual attention at another point during the task, as it is different from the role of haptic or motor attention. On the other hand, haptic and motor attention could serve some of the same functions that visual attention appears to, given the appropriate inputs at an appropriate time during the task time (for fascinating work, see that concerning sensory substitution devices Bach-y Rita, 2004; Nagel, Carl, Kringe, Märtin, & König, 2005).

[5] A further distinction is into "top-down" and "bottom-up" attention. Here, top-down attention refers to any processes whereby attention is allocated based on some "higher-order" cognitive function, as opposed to being attracted by a salient stimulus, as is characteristic of bottom-up attention (for different views of the (neural) distinctness of the two types of attention, see e.g. Connor, Egeth, & Yantis, 2004; Pinto, van der Leij, Sligte, Lamme, & Scholte, 2013). Note simply, that this thesis is concerned with a common mechanism of attention and with how the pointing function of attention that can take place across its modalities and types.

in our descriptions are not as neatly differentiated in physical terms. Nor can the two kinds of attention be readily dissociated based on observed data (Viviani, 1990). That means we have to be attentive to the types of attention in which we describe observed behaviour. So, while haptic attention is taken to be exclusively covert, similar internal attention processes also exist in the visual and motor domain. In each case, these are difficult to address empirically. Covert visual attention is only an exception to some degree, despite the fact that a number of eye behaviours have been associated with internal attention (Benedek, Stoiser, Walcher, & Körner, 2017; Walcher, Körner, & Benedek, 2017). Significantly, the empirical study reported here will be concerned almost exclusively with overt, and then mostly with visual attention.

Box 3.4

Types of Attention

"Covert" vs. "Overt": whereas overt visual attention refers to the act of attending to an object or event by moving the eyes to point in that direction (Posner, 1980), "covert" visual attention describes a mental shift in one's focus without moving one's eyes (Eriksen & Hoffman, 1972). Accounts differ on whether covert attention substitutes or supplements overt attention (Findlay & Gilchrist, 2001). Here, covert attention is also referred to as internal attention (Chun, 2011). That means that the object of attention is not present through the senses, but by being maintained in the agent's working memory (or recalled from long term memory, Chun et al., 2011). "Internal" attentions is not to be confused with "endogenous" (vs. "exogenous") attention, which is not discussed in the current work.

Based on this general overview of types and modalities of attention, the next section can outline a key theoretical framework concerning attention. D. H. Ballard et al. (1997) "deictic routines" suggests that attention not only generates transient representations, but also offers a computationally efficient means of linking successive actions to form a canonical sequence of sensorimotor actions for any well-practised task (for some examples, see M. Land, Mennie, & Rusted, 1999).

3.1.2 Deictic Routines: A Framework for Sequences of Actions

Indeed, Knott (2012) builds on a particular notion of attention. This comes from D. H. Ballard et al. (1997), who make the particular point that attention acts a kind of pointing device. That pointer is a cognitive operation that specifies a momentary reference (Pylyshyn, 1989). In doing so, it accomplishes the linking operation required of attention. It does so, by making momentary sensory information available to other cognitive processes in the form of a transient representation that is maintained as long as attention remains on the same feature of the scene. Indeed, to deal with the challenge of temporarily binding the operation of vast numbers of neurons and rapidly changing those bindings, it seems some kind of pointer system must exist (D. H. Ballard, 1986). To be clear, this is not some abstract mental operation, since D. H. Ballard et al. (1997) fundamentally tie the pointer to bodily states, and specify explicitly how pointers operate on an embodiment level of cognition.

Based on the concept of a pointer, operations involved in an episode (e.g. reaching to grab a cup) are considered to be *deictic* when they create representations tied to specific cognitive programs (D. H. Ballard et al., 1997). [6] One such specific cognitive program might be identifying the location of a target object. Here, the location could be selected by a fixational eye movement, which is used to guide the agent's reaching

[6] i.e. the cognitive (inseparable from perceptual and motor) means of achieving some goal

for the object. Such pointing is called "deictic" because it refers to a target in terms of its essential features. The term is taken over from (Agre & Chapman, 1987), building on (Ullman, 1984). One advantage that arises immediately is that the fixation can easily be changed. Such a change in fixation, by voluntary saccade, for example, captures the dynamic aspect of deictic reference. The benefit, here, is that deictic reference captures the agent's momentary intentions (D. H. Ballard et al., 1997). This kind of action we call a "deictic operation" (see Box 3.5).

Box 3.5

Processes, Operations, Routines

Processes: are those cognitive phenomena that are described in terms of a flow of information or in terms of their suggested function. While a special focus is placed on neural units and connections, processes should by no means be understood to refer only to events in the brain. Nor is the term restricted to biological or internal events. Given our emphasis on the situatedness of action, some processes are best understood by also considering what happens between the agent and her environment (for example, see the us of the term "process" in D. H. Ballard et al., 1997).

Operations: a different term for processes, that places the emphasis on the manipulation/transformation of information (Newell, 1994). A given operation, if successful, brings about a certain state, such as the agent attending the cup. In this sense, the "operation" is more abstract than the "process", referring more to a mental level of analysis than to its physical implementation. In this work, the use of "operation" is limited to so-called deictic operations (see D. H. Ballard et al., 1997).For clarity at the outset, these can be considered special kind of mental process.

Routine: another term used exclusively in reference to the work of (D. H. Ballard et al., 1997). "Routines" are functional primitives, which can be thought of as canonical sequences of operations that achieve some specific function, such as being able to point to aspects of the physical environment. Importantly, within a routine, the end state of each previous operation is the initial state that enables the subsequent operation. As such, the "routine" is roughly analogous to Knott (2012) "episode".

Deictic operations are attentional actions first and foremost (even though a deictic operation like a saccade is also a motor action). One example of an attentional action, meaning an action that reflects an assignment of attention, is the voluntary saccade. The additional term "action of attention" is introduced despite this overlap in order to emphasise that by directing her attention, the agent is doing something. Specifically, she is creating the sensory state that is necessary for the next operation to be performed. That means that individuals operations are linked by actions of attention. By implementing these links, actions of attention create a chain of actions that comes to be a characteristic sequence. In such a chain, attentional actions create transient sensory states that enable the execution of the next action. So, for example, the agent may fixate (very briefly) to create stable image on retina, in order to then classify that object based on the transitory representation generated

In the context of deictic routines, the transient representations generated by attention are called "deictic representations". Part of what defines them is they are best understood, not as self-standing representations, but as the product of the operation that produced them (Knott, 2012). So, the transient representation of the location of a target object may be best understood as the result of a voluntary saccade to that location plus a process of attention that maintains that location in working memory.

At the same time, within a deictic routine, operations themselves are based up the previous deictic representation, since this establishes the context of the next operation. As a result, the meaning of the action rests in the sequence of operations and representations (D. H. Ballard et al., 1997). Indeed, a sequence like this is the basis for much of this thesis (outlined in Chapter 3). The core message, here, is that some action episodes (e.g. reaching to grasp a cup) can be thought of as canonical routines of sensorimotor operations, in which successive actions of attention to one or another aspect of the scene determine how the task is performed. For any structured or well-practices task, this leads to characteristic sequences of attentional actions that define the episode as a whole.

Given editorial constraints, the reader is referred to D. H. Ballard et al. (1997) and commentaries for a review of the relevant empirical literature and justification of its interpretation. One benefit acknowledged throughout is that an external frame of reference that can be shifted rapidly, such as is established by deictic routines, promises a simplification in algorithmic complexity (see also D. H. Ballard, 1991). Put in less computational terms, it makes things easier to do. So, for example, fixating a target object simplifies reaching to grab it by directing the action (Milner & Goodale, 2006). It makes sense, then, that reaching to grasp appears to usually proceed by first looking at the target object and then moving one hand towards the fixation (Jeannerod, Paulignan, & Weiss, 1998).

So, what deictic reference seems to provide is a mechanisms that links sensory information about an important aspect of the scene with motor actions, as well as internal cognitive programs generally (D. H. Ballard et al., 1997). Moreover, given the sequential character of deictic routines, events on this level can be thought of as revealing the temporal structure of sensorimotor processes that make up the action episode. For example, by modifying aspects of the scene in their task, D. H. Ballard et al. (1997) were able to identify the target of individual saccades, i.e. whether they were location or colour-specific. This, in turn, provides insight into the order in which information is (usually) obtained in that specific task.

As we will see in the next section, Knott (2012) will decompose the action of reaching to grab a cup in the same way. Here, thinking of attention as a pointer is not only computationally efficient, but also gives its selective character a computational rationale. So, the deictic understanding of attention offers an account of the utility of selectivity, rather than simply re-describing the phenomenon of attention as a limited resource and failing to explain that selectivity (for a detailed version of this criticism, see Allport, 1989). Moreover, this account of selectivity reiterates the importance of sequences of actions, and offers mechanism for how these actions are linked. In the next section, at last, this understanding will be applied to the cup grabbing example and, by extension, to the original task reported on in this thesis.

3.2 A Sensorimotor Sequence for Reaching-to-Grasp

The sequence of actions in a reach-to-grasp episode can be thought of as a series of operations and corresponding states that the agent is in as a result of the operation (for an overview of the whole sequence, see Table 3.2). This follows the general schema outline in Table 3.1. In this section, a series of operations and their sensory consequence will be discussed in turn. In every case, there will be a short summary, followed by a discussion of a number of key concepts that come up in relation to that specific operation/sensory consequence.

The following points will be clarified about a deictic interpretation of reaching to grasp: i) in relation to Operation 1 (*attend to the agent*), Knott (2012)'s use of the term cognitive "mode"; ii) relation to State

1 (*attending to the agent*) the understanding of agency that is at work in the proposed framework; iii) concerning Operation 2 (*attend to the cup*), the role of attention in terms of a specific example (I will also suggest an ordering constraint upon the execution of reach-to-grab actions; iv) for State 2 (*attending to the cup*), the concept of distinct reach and grasp motor pathways and their involvement in two aspects of the episode, v) regarding Operation 3 (*activate 'grasp' action*), the properties of the reach and grasp pathways and the notion of "action schemas" in relation to action selection; vi) in relation to State 3 (*reattention to the agent*, the dynamic unfolding of action execution, based on a continuous sensorimotor interactions; finally, vii) concerning State 4 *reattention to the cup*), the issue of successful completion of the action and the processes of attention involved in establishing that the actions has been completed.

Table 3.2: The default sequence consists of four cycles, comprising three operations and four sets of sensory consequences. The final cycle has no dedicated operation, since it reflects the end of the grasp action, activated during cycle 3.

	Operation	Sensory Consequence
Cycle 1	Attend to the agent	
		Attending to the agent
Cycle 2	Attend to the cup	
		Attending to the cup
Cycle 3	Activate 'grasp' action	
		Reattention to the agent
Cycle 4		
		Reattention to the cup

Sequence of Actions for Reaching to Grab

Instead of charting reaching to grasp chronologically through different neural systems. [7] , this section follows the sequence of outline in Table 3.2. I will comment on just one functional differentiations as it arises. Namely, the distinctions between a reach and a grasp-pathway (based mainly on studies with monkeys; see Jeannerod, 1996). On the one hand, the differentiations face conflicting results (e.g. Kilner, 2011), and cannot be drawn as neatly in the brain as it is theory. On the other hand, it allows us to highlight different (even if not separate) sets of processes. Moreover, introducing them emphasises, again, the interrelated processes of sensorimotor cognition.

Operation 1: Action of Attention to the Agent

Summary – Prior to any object classification, Knott (2012) postulates a mode-setting operation that identifies the agent as either executor or observer of an action (see Box 3.6). If this is required, it follows that establishing "action execution mode" should precede other component processes of any action. On this account, a configuring for action execution takes place before

[7] Schematically, the series of operations that Knott (2012) charts through the brain involves the following systems. First, there is the early visual system, including object classification and object categorisation (for a discussion, see e.g. Carandini et al., 2005) This is joined by the primary processing of somatosensory input (see e.g. Mountcastle, 2005). In addition, there are a number of bottom-up and top-down attentional processes operating through a somewhat distinct pathway (for seminal research on monkeys, see respectively K. G. Thompson & Bichot, 2005; Kusunoki, Gottlieb, & Goldberg, 2000) Besides these are also the multiple routes of action recognition (Kilner, 2011), including the much-contested mirror neurons (Gallese et al., 2011). Knott (2012) also discusses a "who"-pathway (see Box 3.7), that is associated with making the distinction between executed and observed actions (Georgieff & Jeannerod, 1998).

any voluntary movement can follow (see also Fagg & Arbib, 1998; Haggard, 2008). Further, through a feedback process, establishing this mode constitutes an action of the agent attending to herself and prefigures her as the author of the action to come (see Box 3.7). [8]

This operation is included for completeness. Since this thesis is concerned with action execution, this part of the sequence is of secondary importance.

Box 3.6

Cognitive Modes – Action Execution vs. Action Observation

The concept of a "mode" refers to an idea of distinct whole-brain states, implemented in more or less tightly coupled, large-scale neural circuits (e.g. the default mode network; for a review, see Buckner, Andrews-Hanna, & Schacter, 2008). For Knott (2012), modes are important because his account of action observation draws on the mirror neuron theory (for a critical discussion, see Gallese et al., 2011). Since neural activation is suggested to be the same whether the agent is executing or observing an action, such accounts require an explicit operation to distinguish those modes. That operation would occur prior to action monitoring, and constitute as a kind on nonspecific decision to act, i.e. to perform some motor behaviour, rather than to observe another agent doing so.

State 1: Attending to the Agent

Summary – Granting operation 1, the result of entering action-execution mode and the reafferent consequences this triggers is that the agent is aware of herself as the actor. Haggard (2008) discusses this role of preparatory processes, i.e. that a good part of becoming aware of ourselves, and of our feeling of agency, results from the preparation of volitional actions, during which bodily experience is brought to attention to constitute the subject "I" (see Box 3.7).

Agency, like mode-setting, is mentioned mainly for completeness. In contrast, this first appearance of the agent as the target of an attentional agent will later be shown to be a crucial aspect of the episode representation (see Sect. 3.3.1). This has to do with how a universal LF – after all a recapitulation of sensorimotor routine currently being described – can act as the basis for the varied surface forms (PF) of different natural languages.

What is important to note concerning agency nonetheless, is that Knott (2012) presents an account that is supposed to hold equally for performed and observed actions. There are differences though, of course, between action execution and action perception. In one case, it is the agent herself who is acting, and in the other it is someone else. That means that one difference is whether or not the agents is aware that it is her that is acting. This is what brings us to the topic of agency. In the limited sense discussed here, the term originates in a process of entering an "action-execution" or "action-observation mode", that is based in a "decision-to-act" cognitive operation. Trevena and Miller (2002) suggest that the general readiness potential (GRP) of voluntary actions is candidate for such an operation (for a translation of the seminal work on GRP, see Kornhuber & Deecke, 2016). Indeed, brain stimulation studies suggest that this preparatory operation takes a crucial role in attribution of agency (e.g. Desmurget et al., 2009). This is

[8] Establishing agency through such an action of attention does not explain how the agent attributes involuntary actions to herself. However, this could be accomplished via a process of observing her own action and retrospectively attributing agency to herself for it. Indeed, this does not seem too far from the phenomenology of involuntary actions.

said to happen because the reafferent consequences of this cognitive operation are interpreted as entering action execution mode, which also appears to contribute to self-representation. One key point for Knott (2012) is that the distinction between execution and action perception must occur at the very beginning of any concrete action episode. This leads to the following constraint:

> **Ordering constraint 0:** biologically-inspired computational models of action recognition assume that the agent's intended target is identified before the process of action monitoring/classification begins

Note, this is different from a premotor signal, which would reflect the preparation of a specific movement, and which has been localised to different areas and takes place later (Shibasaki & Hallett, 2006; Smulders & Miller, 2012).

Box 3.7

Agency

Knott (2012) does not imply that there is no representation of the animate self prior to the action-execution mode-setting operation (for en explicit argument against this claim, see Damasio, 1999). Nor is the operation discussed here in any way exhaustive of self-concept or agency. It refers only to the question: *am I performing this action?*. Further, all of this reflects a special sense of agency, summarised as an agent's sense that she is the one who is producing an action (Gallagher, 2000). As such, it is different from the broader understanding of agency in situated action perspective, where it denotes the capacity for action (L. A. Suchman, 1987). As a caveat, it may be worth thinking of agency as a relationship and not something that an individual "has" (Barad, 1999).

Operation 2: Action of Attention to the Cup

> **Summary** – Once in action execution mode (see State 1), the first operations is target selection. This selection process results in an action of attention to the object via a distributed attentional network that includes several premotor regions.

Recall that the structure of sensorimotor processes, such as saccades and fixations, reflects the demands of the task (for further evidence, see Flanagan & Johansson, 2003; Webb, 2015). That is key, because it suggests that actions are performed precisely when they are required in order to achieve a goal. So, for reaching to grasp, (D. H. Ballard et al., 1997) present results in favour of a characteristic sequence of saccades and fixations. Moreover, these data come in addition to similar results for action observation (see e.g. Webb et al., 2010). Indeed, this is relevant especially since it has been suggested that action observation uses the same sensorimotor processes that is involved during action execution (Flanagan & Johansson, 2003).

If we accept the default sequence discussed by these authors, then the object tends to be attended to early during a reach-to-grasp episode. In conjunction with the neuroscientific evidence of signal timing in macaques (Schmolesky et al., 1998), this leads Knott (2012) to formulate the following ordering constraint:

> **Ordering constraint 1:** an object must be attended to before it can be categorised

This constraint exemplifies the central role of attention, which appears to be necessary in order for a certain task to be completed. The underlying attentional network is suggested to includes premotor regions, since several of these appear to be associated with controlling eye-movements (Colby & Goldberg, 1999; K. G. Thompson & Bichot, 2005). For our purposes the constraint implies that participants will tend to attend to the object early in a reach-to-grasp episode, i.e. before they begin to move their hand to grasp it. Testing this claims, as this thesis does, is an attempt to replicate the characteristic sequence.

State 2: Attending the Cup

> **Summary** – Taking the example of vision, the action of attention in operation 2 modulates visual processing to permit classification of the target. This generates an enriched representation. Persistent visual attention to the object enhances and maintains the motor representation of the reach pathway. Simultaneously, it also results in the generation of affordance-based representation of the cup through the grasp pathway.

Here, I will forgo further discussion of visual classification and comment only on aspects relating to the reach and grasp pathways. Concerning reaching to grasp an object, we can define two components with distinct, though interacting pathways (Jeannerod, 1996). Namely, it appears that separate motor circuits are primarily responsible for reaching and grasping. The former is related to movements of the shoulder, elbow, and wrist joints to transport hand to appropriate positions in relation to object, the latter to movements of fingers and thumb, the object-specific hand preshape (see e.g. Winges et al., 2003), as well as for controlling contact with object.

Concerning the role of visual attention in the reach pathway, the pattern of saccades during the reaching movement appears closely related to motor components of action (M. Land et al., 1999). As a result of this integration of the pattern of saccades and the motor components of an action, previous research has pointed a number of regularities. For example, the agent typically fixates the target object before the action is properly underway (Johansson, Westling, Bäckström, & Flanagan, 2001). This leads Knott (2012) to the following ordering constraint:

> **Ordering constraint 2:** during execution of reach-to-grasp action, agent typically attends to target in early stages of the action

Indeed, data on sequences of saccades suggest that information is acquired incrementally during the action (for a summary, see Hayhoe & Ballard, 2005).

Meanwhile, affordance-based representations are associated with the grasp pathway. This, in contrast to the reach pathway, is associated with hand, rather than arm movements. The key to these is not object location but object shape (Jeannerod, 1996). Throughout this pathway, the co-ordinates of representations are not person-centred but object-centred instead (Jeannerod et al., 1998). That means that the articulation of hand proceeds in relation to the shape of the object. Since action-guidance in this pathway is always in relation to the object, this entails that its representation can be thought of as being affordance-based.

For example, in addition to grasp preshapes, the grasp pathway also appears to be associated with aspects of hand trajectory that have to do with how the hand is shaped in relation to how far it is from the target object (Knott, 2012). Importantly, this preshaping appears tied to some representation of the target object, since continuous visual feedback of the hand and/or object is not necessary to allow the hand to gradually conform to object contours (Winges et al., 2003). Also, hand preshaping is one example of a process based on repeated contact between reach and grasp pathways, and their respective embedding in the broader neural architecture (e.g. Fagg & Arbib, 1998; Oztop & Arbib, 2002).

Note that, during a reach-to-grab episode, this is the point at which overt movement towards the target object begins.

Operation 3: Activation of the Grasp Action

Summary – On the basis of the representations computed in the motor circuits and by visual classification (see State 2), a grasp action is selected to be performed. Here, bottom-up processes are tightly coupled with top-down influences. In this case, the reaching and grasping components are likely selected as components of a high-level action category.

Knott (2012)'s model builds on the idea that actions are represented functionally, meaning they are defined in relation to their intended effect (see e.g. Hommel, Müsseler, Aschersleben, & Prinz, 2001). One account of action categories is that of a "schema". This concept comes from a high-level (read "abstract") functional vocabulary, and denotes persistent patterns in actions that can be selected for execution (M. A. Arbib, 1992). Specifically, the term refers to commonalities between various repetitions of the same action (this reflects the original formulation by Piaget, 1971) and involves an expectation that the action will follow a characteristic course. Given such a characteristic sequence, such schemas may develop for all highly practiced actions (M. Arbib, 1981; Oztop & Arbib, 2002). These sequences, so M. F. Land and Furneaux (1997) show, can include patterns of eye movements, as has been discussed here. Examples might be steering a car around a corner or playing a stroke in table tennis (for a summary, see M. F. Land, 2009). Throughout, this emphasises a continuous and pervasive interaction of sensory and motor processes.

This view of actions builds on the observation that actions become more stereotyped and efficient as they are repeated (for a discussion of the neural basis, see e.g. Censor, Sagi, & Cohen, 2012). That streamlining is what is modelled in the action schema, which is proposed as a coordinating influence on the different movements needed to complete an action. For grasping, for example, which movements are needed depends on the target object. This suggests that it is unlikely that action selection would ever precede the classification of the target, which occurs only during the preceding state of attention (Knott, 2012). This reiterates ordering constraint 2 (above). In part, this sequence appears to be the result of an interaction of top-down (goal-related) and bottom-up (stimulus related) attentional processes through sensorimotor processes related to motor control and object categorisation reciprocally bias each other toward the same target (for a review, see Schneider & Deubel, 2002).

State 3: Reattention to the Agent

Summary – Once the grasp action can begin (see Operation 3), this triggers the dynamic process of action execution. During this, efferent motor signals constantly update the selected movement vector, which continues to guide the behaviour being performed. A similarly dynamic representation of the body state offers the second representation of the agent, attended to via a complex of somatosensory signals.

One aspect of action execution is the coupling of sensorimotor processes to guide action. For example, during reaching actions, saccades are mostly to control points along the movement trajectory (Johansson et al., 2001). Control points include, among others, where the hand will contact an object, borders of obstacles, and the location an object is being moved to. Meanwhile, the moving hand is almost never tracked. This is unsurprising, since there may be a specialised selective attention to the body within the haptic modality (for a review, see e.g. Holmes & Spence, 2004). Indeed, there is evidence to support a distinct motor subsystem for action guidance, which is based in a relatively distinct modality for attending to parts of own body, referred to as "body schema" (Schwoebel & Coslett, 2005; Berlucchi & Aglioti, 2010). Therefore, in the case of action guidance, haptic attention is expected privileged over visual attention.

Indeed, the somatosensory and proprioceptive aspects mentioned above are part of larger, dynamic representation of the body that is active during action execution. This also reflects the continuous interplay of processes in the reach and grasp pathways throughout action execution. Taken together, this leads Knott (2012) to interpret action guidance as a form of attention to self, the experience of which reflects a gestalt of the various modalities involved.

State 4: Reattention to the Cup

> **Summary** – Achieving a stable grasp of the object terminates a successful reach-to-grasp action. The agent is not only holding the cup, but can also be argued to shift from a state of attention to herself (State 3) into a state of attention to the cup, possibly (also) in the haptic modality. This reflects the epistemic change resulting from the new configuration of agent and target object and the novel information now available about this situation, i.e. sensory feedback that the agent is holding the cup.

Concerning this moment of grasping, we can think of it as being successful when the agent is holding the object in a sure grasp. According to Knott (2012), this is a motor operation, but also an attentional one. On this perspective, that is because of a change in what the agent is able to know, which occurs when a new source of information about object becomes available (Goodwin & Wheat, 2004). In this case, that information is the haptic information from touching the cup. To Knott (2012), this raises the issues of haptic attention. Indeed, the information could be conceptualised as a kind of feedback, indicating/confirming the successful completion of the grasping action. This suggests a state of haptic attention to the object (see Box 3.3). That, in turn, leads to the following ordering constraint:

> **Ordering constraint 4:** after completion of reach-to-grasp action, the agent reattends to the target object [9]

The reattention is supposed to be in the haptic modality. That necessarily makes it covert. Either way, it would not be possible to address it in this thesis. However, there may also be an action of visual attending to the object. Particularly, since there is evidence for a common reference frame for both modalities (seminally, see Driver & Spence, 1998). Moreover, and in relation to multimodal perception, there could be an overlap of visual and haptic processing at this point (Maravita, Spence, & Driver, 2003). No doubt, haptic attention is distinct from visual attention in other ways but they may contribute similarly or jointly to the generation of a signal that marks the successful completion of the action.

Finally, this reattention, especially inasmuch as it is in the haptic modality, could also serve as the basis for learning cross-modal representations, such as grasp affordances for example, and as feedback towards the generation and maintenance of action schemas (Knott, 2012). So, it appears that the grasp pathway is not

[9] For ease of reference, the original numbering of ordering constraints by Knott (2012) is maintained here.

only instrumental to the act aspects of hand shaping. Instead, it may also have a broader significance for the reach-to-grab episode. Again, a crucial point is that this is achieved through a sensorimotor interplay. Here too, there appears to be reciprocal influences between sensorimotor processes, for instance to generate motor affordances.

3.3 Models of Memory for Sensorimotor Sequences

Above, I have only considered transient signals that reflect the sequential operations of the sensorimotor routine of reaching-to-grasp. To recap, these are (i) an action of attention to the agent, with the consequence being a representation of the agent, (ii) an action of attention to the cup, which generates a sensorimotor representation of the cup, (iii) the evoking of a motor action and a resulting reattention to the agent, and (iv) a process of reattention to the cup, at least partly in the haptic modality. Together, these are the transient signals in Table 3.3 (see Sect. 3.3.2).

One question that remains, is whether, and if so, how the individual operations are bound as a single unit. For Knott (2012), this question arises in response to the problem that the agent must prepare the whole sequence of actions in advance, to make sure they occur in the right order. Preparing the whole sequence of actions in this way is intended to contribute to the mechanism that keeps the agent doing one thing, even in the face of unexpected events. In (Knott, 2012)'s view, that requires some kind of representation of the whole sequence to be active throughout the execution of the action. Such a tonically active representation would consist of a prepared sequence of actions, or "plan" (quotation marks in Knott (2012), too; see Box 3.8). Note that performing the action is a situated practice that is neither fully-specified nor determined by this plan, but guided by it (read one way, both Vera & Simon, 1993; L. A. Suchman, 2007, agree on this point).

Box 3.8

Plans

Plans take a central role in Knott (2012)'s account of the sensorimotor aspects of a reach-to-grasp action. Note that, the "plan", too, is a metaphor that illustrates a certain cognitive operation. Here, a plan is a prepared sequence of actions. However, these actions are of a different granularity to those discussed by L. A. Suchman (1987); Vera and Simon (1993) (recall Box 3.1). Concretely, "plan" here refers to a distributed neural representation of a particular sequence of sensorimotor operations (Takac & Knott, 2015). Despite referring to a different domain, L. A. Suchman (1987)'s point is well taken that the plan does not determine the sequence of actions that is actually carried out. Instead, the actions that are performed reflected an unfolding resonance between agent and environment, perhaps especially in neural terms (see also Winograd & Flores, 1986). Rather than determine the action, the "plan" orients the agent, based on the expectation that a certain sequence of actions is beneficial in the current situation (L. A. Suchman, 2007; Knott, 2012). Whether canoeing down white-water rapids or reaching to grab a cup, the action that is performed emerges out of the relation between the agent and her circumstances. These circumstances are not static, or prefigured by the "plan", but are always in the process of being generated as the agent acts and time passes (L. A. Suchman, 1987).

In this thesis, a plan is a representation of a sequence of actions that specifies the order in which the individual actions are to be performed. As a brief clarification, this action plan is the same as the "episode representation", which is the sustained signal that complements the transient signals discussed above. Both together make up the sensorimotor model (for a formal implementation, see e.g. Takac et al., 2012). Of this model, the plan is the last remaining aspect that I will consider ahead of a final summary in Sect. 3.2.

However, to explain what a plan is and how one is formed, I have to begin by briefly outlining the understanding of memory at work in Knott (2012). In particular, Sect. 3.3 addresses the corresponding model of working memory (based on A. D. Baddeley & Hitch, 1974; A. Baddeley, 2000, 2003). For, this thesis is specifically concerned with the way sequences of actions, i.e. plans, could be represented in working memory.

Working Memory Plans: Representing Sequences of Actions

The term "working memory" is attributed to G. A. Miller, Galanter, and Pribram (1960), though seminal experiments relating to a transient memory representation and its neural basis were conducted earlier (for a overview of some of this early neuroscientific work, see Benton, 1991). Given editorial constraints, a discussion dedicated only to memory models for sequences of actions is not possible here. Instead, I will just mention the key concepts and models that Knott (2012) draws on, before summarising what these imply for sequential episode representations.

To begin, Knott (2012) distinguishes two uses of the term "working memory". The first account of working memory comes from behavioural psychology, and is related to action preparation specifically (see e.g. Monsell, 2003). This understanding of working memory builds on an idea of "task sets", which are top-down, learned biases on an agent's actions (Knott, 2012). That means that task sets refer to a mechanism that prepares the agent to solve a certain problem in a certain way, for example by cuing certain cognitive skills. This view is related but separable from the specific model of working memory that is important in this thesis (for a discussion of task sets in context of that model, see A. Baddeley, Chincotta, & Adlam, 2001).

The second understanding of working memory (see Box 3.9), and the more important one for this thesis, builds on seminal work by A. D. Baddeley and Hitch (1974).[10] In it, emphasis is placed on short, transient representations (a later formulation that expands on this point is A. Baddeley, 2003). There is a difference between these representations and task sets, which may persist for a longer amount of time (on the order of seconds, see A. Baddeley, 2000). As far as this thesis is concerned, the "episodic buffer" A. Baddeley (2000) is the essential notion of this second account of working memory. In terms of mechanism, this is a modular construct, added to A. D. Baddeley and Hitch (1974)'s seminal model. It was introduced to accommodate empirical results that appeared to demand a specific module to process memory of sequences of events, the internal structure of which is retained because it is meaningful to the agent. Sentences are one example, since longer sequences of words can be remembered when they form a sentence, rather than when the words are in a shuffled order (Hulme et al., 1997).

The role of the episodic buffer is to integrate information across modalities and maintain a special, "episodic" representation via rehearsal (A. D. Baddeley & Andrade, 2000). [11]. In the model, this rehearsal occurs based on a mechanism that sequentially directs attention to the different components of the representation (A. Baddeley, 2000). Crucially, how and when to direct attention is learned, and likely based on embodied experience (see discussion of embodied memory in Sect. 2.2). [12]

[10]Do note the staunch cognitivism of this model, though not necessarily as a criticism. Instead, bear in mind that it is a paradigmatic example of a set of ideas about cognition (and embodiment too, though it may come as a surprise) that constitute and construct a particular, historically-embedded computational theory of mind (E. A. Wilson, 1996). I take up this issue, along with a few others, in Sect. 7.

[11]"rehearsal" is also closely tied to the notion of embodied simulations, discussed in Sect. 2.2.

[12]This also reiterates to the role of attention discussed in D. H. Ballard et al. (1997)'s account of deictic routines. In fact, the episodic buffer introduces a specific model of memory for such sequences of events.

For the sensorimotor model, the episodic buffer offers a functional perspective on memory for sequences of actions (recall Marr (1982)'s computational level). Expanding on it, the complete model of memory for sensorimotor sequences in Knott (2012) specifies two mechanisms for generating the representations of actions sequences through the episodic buffer. These mechanisms are competitive queuing (e.g. Grossberg, 1978; Rhodes, Bullock, Verwey, Averbeck, & Page, 2004) and associative chaining (e.g. E. K. Miller & Cohen, 2001). For an account of action plans the two mechanisms are not mutually exclusive. Instead, the two can thought together, as different mechanisms (Marr (1982)'s representations & algorithm) behind the same function (computation).

Box 3.9

Working Memory

A. D. Baddeley and Hitch (1974)'s model and its' successors are modular. That means they consist of a central executive plus a changing number of memory modules. A considerable volume of empirical data have been interpreted to support this model (for an introduction, see A. Baddeley, 2012). While there are alternative models, the difference may be more of emphasis and terminology. Some persistent concerns with the model include criticisms that individual aspects are under-defined and that the modulating effects of emotion and somatic aspects are not sufficiently accounted for (Andrade, 2001). Sensorimotor aspects, at least, have subsequently begun to be considered (for computational work, see e.g. Schoner & Spencer, 2006). The success of A. Baddeley (2000)'s model is how productive it has been (for the significance of this, recall Lakatos, 1976).

Competitive queuing is the idea that all actions in a sequence are prepared on two levels. On one level they are active in parallel, but on the second, only one action can be activated. What happens during actions execution is that, one after another, whichever of the parallel actions is most strongly activated triggers its analogue on the competitive level and then becomes inhibited to avoid repeated actions (Averbeck, Chafee, Crowe, & Georgopoulos, 2002). [13] In addition to this, a tonic representation of each action in the prepared sequence is active throughout the sequence (Averbeck & Lee, 2007). Knott (2012)'s adopts this model for short, well-learned sequences with predictable movements.

For less predictable or longer actions, associative chaining is considered more suitable. The basic idea here is one of recurrence (seminally discussed by Elman, 1990). Here, the consequence of one action within the sequence is that it automatically triggers the next action. This model builds on the image of a sequence of actions as an ensemble of stimulus-response pathways that are encoded across distributed neural cell assemblies (Fagg & Arbib, 1998), where "stimulus-response pathway" refers to a mapping between sensory states and motor actions. Because the model is recurrent, the current event (e.g. sensory input) is represented alongside feedback about the previous event (e.g. a motor action (Dominey, Arbib, & Joseph, 1995) or previous sensory state (Beiser & Houk, 1998)). Again, Knott (2012) predicts that there is a tonic representation of the entire action sequence. In this case, this is reflected in the biases that exist throughout the underlying cell assembly.

Both mechanisms are supplemented with a "context" signal. This refers to a representation of aspects including how far into the prepared sequence the agent is, as well as information on the agent's circumstances, which may require adjustments to the course of action. For competitive queuing, a representation of context is useful for dealing with repeated actions during a sequence (see e.g. Houghton & Hartley, 1995). For

[13] To account for repeated actions, a third "context layer" is added to the model (Houghton & Hartley, 1995).

associative chaining, one benefit of recurrence is that it provides an encoding of recent actions. This deals with the challenge of repeated actions, because, unlike the plan, this context signal is continuously updated. That means this signal can represent the agent's progress through the plan, and thereby disambiguate which of the possible actions that follow the same signal should be executed next (Knott, 2012).

In both models, sequences are learned. [14] Moreover, both proposals are premised on the idea of a tonically active representation of action sequences. Indeed, one key point for this thesis is that the two mechanisms are not mutually exclusive, though each is supported by distinct sets of empirical data. In Knott (2012)'s combined account, the agent's plan of reaching to grasp a cup consists of a particular sequence of sensorimotor operations, encoded across a neural cell assembly and activated because this sequence of events has been beneficial in a similar situation to the one the agent is currently in – recall that the situation includes her goal. When the assembly is activated, if it is not interrupted, its internal dynamics carry the agent through the entire sequence of actions. The central message of this is that the plan is a tonically active representation of the entire sequence of sensorimotor operations that comprises an episode, like our reaching-to-grasp.

To round out this account, (Knott, 2012) also considers competition between different action plans, and the neural instantiation of the entire model in prefrontal cortex and with essential involvement of the hippocampus. That is all beyond the present scope. For the purpose of this thesis, the final point to consider is the mechanism behind the replaying of "plans", since this is the crucial interface to linguistic structure.

3.3.1 Interface of Sensorimotor Cognition and Linguistic Structure

How a sensorimotor sequence provides structure to language still remains an open question, both within the scope of the present work and in the literature in general. That has to do with the question of how sensorimotor sequences are represented, and how these representations can subserve language. For, to act as the basis for language, sensorimotor sequences need to have a representation that can be separated from the time during which the action is being performed (Glenberg & Kaschak, 2002; Knott, 2010; M. A. Arbib, 2019).

The previous section has outlined some of the memory mechanisms that Knott (2012) suggests are involved in generating representations of sequences of actions in working memory. Next, I summarise a mechanisms for how to bridge the gap between the time during which the agent reaches to grab a cup and the time during which a sensorimotor representation of the corresponding sequence of actions might serve as the basis for linguistic structure. This builds on Hommel et al. (2001)'s idea that motor actions may be cognitively represented in terms of their expected reafferent consequences. A similar process, but in relation to sensory information, plays a role in mirror neuron theory. The theory also assumes that an action representation can be simulated ("rehearsed"), including the expected sensory consequences of the simulated action (for related accounts, see e.g. Jeannerod, 2001; Barsalou et al., 2003; Grezes & Decety, 2001; Feldman & Narayanan, 2004). In each case, this is accomplished through the biasing of attentional and motor actions. The novelty is that the attentional actions may be mostly covert and that no overt movement is carried out. [15]

[14] In both models, this occurs in a manner that is modelled as reinforcement learning (respectively for competitive queuing and associative chaining, see e.g. Averbeck et al., 2002; Dominey et al., 1995).

[15] It is this emphasis on simulation that recalls (A. Baddeley, 2000)'s "episodic buffer". For, as specified there, the working memory sequence discussed here i) draws on multimodal sensorimotor operations, ii) is inherently sequential, and iii) can be internally rehearsed (Knott, 2012) – each of these aspect is a feature of the episodic buffer, too.

Crucially, this account proposes that memory is not merely a store for information, but that it represents (actions in terms of) perceptual and attentional operations. The same idea is found in other models of perception (D. H. Ballard et al., 1997) and memory (prominently in e.g. Kolers & Roediger III, 1984; Kent & Lamberts, 2008). Knott (2012) introduces the term "episode representation" in relation to the plans discussed above, because of its sequential structure. For, only sequential structure allows the representation of a sequence of actions as a sensorimotor routine. Since the episode representation is brought forth through working memory, it is also referred to as a "working memory episode". In sum, this is a sequence-based working memory construct of the experienced episode. Now, its role concerning language is to facilitate the rehearsal, or internal replaying of the individual operations involved in a sensorimotor routine, together with their reafferent sensory consequences. To Knott (2012), the reason that the agent can utter (and understand) a sentence describing a reach-to-grasp episode, is that she has, in working memory, a planned sequence of the working memory operations that make it up, which she can replay. This is the sustained signal in Table 3.3.

Then, it is the replay process, as a whole, that constitutes the interface between linguistic structure and the sensorimotor operations discussed above (see Sect.3.2). So, working memory is the interface. In a sentence, what happens through working memory is that linguistic structure develops as a description of the process of replaying the entire sensorimotor routine.

That means that elements of the linguistic structure (loosely, e.g. subject, verb, object) are mapped onto different components of the sensorimotor routine. This mapping is learned based on the language that infants are exposed to (for some interesting results relating to language developing based on a computational implementation, see Takac et al., 2012). For readers wishing to pursue the developmental angle, Knott (2012, Ch. 7) discusses a model of language development in relation to the sensorimotor model outlined here.

Now, the interface between sensorimotor processes and linguistic structure is not itself the target of this thesis. One reason the topic is not pursued further at this point, is that the line of research proposed here and Knott (2012)'s diverge at this point, since this thesis does not share the earlier emphasis on syntax. For Knott (2012), the replay of each cycle of the sensorimotor routine corresponds to a certain grammatical construct. [16] This thesis does not make a similarly specific prediction. As a result, the investigation of relations between sensorimotor processes and linguistic structure in this work focuses on patterns in the measured data (eye-tracking and motion capture – see Sect. 6.2.4).

Finally, I defer further discussion of this model of memory to Sect. 3.4, in which I assess the sensorimotor model as a whole. Readers interested in the mechanistic details of memory episode representations are referred to Section 3.6 – 3.9 in Knott (2012, pages 121–152). Readers interested in computational implementations are referred to Takac et al. (2012); Takac and Knott (2015); Knott and Takac (2016); Takac and Knott (2018). Since computational modelling is one aspect of theory testing, the success of these models can be interpreted as support for the model of working memory outlined above. One success among others, in this case, was the production of different word-ordering conventions (e.g. Subject-Verb-Object, SOV, OSV, OVS, VOS) at will. This may, in fact, be a shortcoming, since the different word orders are not equally prevalent within human languages (Tomlin, 1986). One way to assess this proposal, is not primarily in terms of its consistency with one set of beliefs, but in terms of productiveness and the ability to produce novel results (see Sect. 2.1.1). Certainly, that the computational implementations satisfy certain specific predictions is positive. On the other hand, productiveness can be a dangerous criterion for neural networks,

[16] This is Minimalism's "X-bar schema" (Chomsky, 2014; Knott, 2012, chapters 4 and 5)

since we cannot assume that the model is necessarily doing what we think it is doing, just because it produces the expected output (Rodríguez-Algarra, Sturm, & Dixon, 2019). This concern should be kept in mind alongside concerns about the relationship between human memory and artificial models of memory.

3.3.2 Summary of Knot (2012)'s Sensorimotor Model of Reaching to Grasp

Having described the sensorimotor sequence of reaching to grasp, and the corresponding model of working memory for sequences, the entire sensorimotor model is summarised in Table 3.3.

Table 3.3: Summary of the Episode Representation according to Knott (2012)

Sustained Signals	Transient Signals		
	Context Signals	Actions Signals	Reafferent sensory signals
$\text{plan}_{attend-agent/attend-cup/grasp}$	C_1		
		attend-agent	
			attending-to-agent
$\text{plan}_{attend-agent/attend-cup/grasp}$	C_2		
		attend-cup	
			attending-to-cup
$\text{plan}_{attend-agent/attend-cup/grasp}$	C_3		
		grasp	
			attending-to-agent
$\text{plan}_{attend-agent/attend-cup/grasp}$	C_4		*attending-to-cup*

Crucially, the sequential structure of this representation is preserved by working memory. Moreover, that sequence is also maintained when the sequence is reproduced by a process of replaying it through working memory. Table 3.3 summarises this replay process. Broadly, it consists of two components: sustained and transient signals. The sustained signal is a plan, held in working memory, that represented the entire sequence of actions to be performed. This representation is complemented by transient signals, which are associated with the current action and its reafferent sensory feedback. The final transient signal is a representation of "context", which facilitates the execution of the actions in the proper order (Knott, 2012). A new phase of the replay begins with each context signal, as previously outlined in Table 3.1. If successful, the sequence of reaching to grasp concludes, following C_4, with the agent in a state of haptic attention to the cup. With that, the woman (has) grabbed a cup.

3.4 Evaluating the Sensorimotor Model of Linguistic Structure

Knott (2012) proposes that the model outlined above should be judged in terms of the success it has in accounting for the grounding of linguistic structure in sensorimotor representations. Certainly, there are empirical results that do not match individual aspects of the model. Likewise, there are competing accounts for individual mechanisms put forward above (for broad reviews of the neural correlates of prehension (reaching to grasp), see e.g. Castiello, 2005; Grafton, 2010; Turella & Lingnau, 2014). Clearly, the model summarised above is one possible synthesis. Unsurprisingly, then, there are other models with which this one is not always, or even not at all compatible (see e.g. Van Elk et al., 2010; Harvey, 2015, for some pro's and con's of an enactvist grounding of language respectively).

That being the case, neither Knott (2012) nor I make any claim that this is *the* model. To underwrite such a claim, the model would have to be tested as whole. Especially given the breadth of the model this promises to be exceedingly difficult. That is, if it is possible at all (for some suggestions why it is not, see e.g. Lakatos & Feyerabend, 1999). Instead of attempting any such argument, the remainder of this thesis will focus on testing a number of specific predictions (these were emphasised as ”ordering constraints” above). Before getting to that, I make some general comments on the quality of the model, as promised in Sect. 2.1.1.

A first point concerns the parsimony of the synthesis that Knott (2012) offers. Concerning the episodic buffer, for example, the sensorimotor model integrates conceptions of working memory (e.g. A. Baddeley, 2000), ideas about episodic memory (e.g. E. K. Miller & Cohen, 2001), as well as two mechanisms to implement these ideas. Individual aspects of the different accounts are mapped onto one another and integrated into a single model. While drawing these results together is not a unique strength of the sensorimotor model, its potential for doing so suggests that i) the model is broadly consistent with existing work, and that ii) the model provides a framework within which disparate results can be reconciled. Both points suggests a model that is greater than the sum of its parts.

Another point to consider is the model's potential to account for phenomena and processes are considered unexplained. The near universality of some aspects of language, for example, continues to be a source of controversy. Meanwhile, the sensorimotor model offers an evolutionarily plausible account of these observations (Knott, 2010, 2014). That, in itself, is considered a strength of any model able to do so (Fum et al., 2007). Besides that, this account does not propose some additional or specific mechanism. Instead, grounding linguistic structure in the replay of working memory episodes takes advantage of a mechanism that was independently proposed, and which is supported by a distinct set of data. Granted, this is a reflection of the parsimony of the shared mechanisms hypothesis, rather one specific to (Knott, 2012)'s framework.

An independent strength, though, is the generality of that framework. Despite focusing only on reaching to grasp here, and the sentence “the woman grabbed a cup”, it is not limited to utterances concerning transitive concrete actions like this one. So, Knott (2012) also considers a mechanism to account for the grounding of abstract sentences, while (Knott, 2017a) considers a number of additional cases (see also Knott, 2017c). Likewise, the sensorimotor operations described above are not specific to reaching to grasp. Recall for example, evidence of the canonical temporal structure of a variety of actions (M. Land et al., 1999), that deictic routines are a domain-general framework for the role of attention (D. H. Ballard et al., 1997), or that the models of action referred to in this section were not, themselves, developed with a single example in mind (e.g. Fagg & Arbib, 1998; Oztop & Arbib, 2002).

As it stands, the effectiveness of the model is being tested. Computationally, it is promising (see e.g. Takac et al., 2012). Empirically, the model has proved difficult to investigate. So, for example, (Webb, 2015) reports a failure to find evidence for some of the mechanisms proposed in the sensorimotor model. However, whether this has to be interpreted as evidence against the model remains an open question. Part of this specific challenge, is that definitive statements about the role of fixational eye movements are impossible when we cannot be certain what exactly the agent is currently doing (Viviani, 1990). Indeed, this outlines one persistent concern with attempts to test the sensorimotor model, the attempt reported in this thesis included.

So, while the sensorimotor model offers a number of strengths, it also faces concerns. To wrap up this brief evaluation of the model, the next section will consider some concerns about the model. Following this, the final assessment of the model is left to the reader in any case. Instead, I conclude this section by noting three concrete predictions that were tested as part of this thesis, and which were derived from the sensorimotor model.

3.4.1 Concerns About the Sensorimotor Model of Linguistic Structure

As mentioned above, Knott (2012) does attempt an account of abstract sentences on the basis of the sensorimotor model. On the one hand, there is persistent debate about the grounding of abstract concepts (see e.g. Mahon, 2015). On the other hand, the sensorimotor model is not primarily about conceptual knowledge, let alone is it an account of the meaning of words. Instead, it is an account of the underlying structure of language, and in this respect, there is no distinction between abstract and concrete sentences. So, abstract sentences turn out not to offer a particular challenge.

A more important challenge is posed by the need to make falsifiable predictions. [17] To address this, Knott (2012) outlines a number of predictions about other kinds of grammatical constructions than the concrete, transitive sentence. Actually, this thesis contributes to addressing this concern, because the utterances of our participants were frequently in the progressive aspect ("I am grasping the cup", rather than "I grasp the cup"). In fact, such progressive utterance raise concerns of their own about the sensorimotor model (Knott (personal communication) – for further discussion, see Sect. 6.2.4).

A very different challenge is posed by cases of congenital paralysis. Specifically, the sensorimotor model has to offer an account of how individuals who are born paralysed are able to acquire language and ground (the meaning of) linguistic structure. For each assigned cause and for every progression of the different cases of paralysis, there must be some mechanism that substitutes the overt performance of motor actions. While each specific case will be a different, the same general mechanism may be relied upon. Embodied simulations are likely to play a central role here, and as is action observation. This issue recalls the challenge that Piagetian developmental theories (Piaget, 1977; Gelman & Baillargeon, 1983) face more generally. In fact, paralysis is also raised as a major challenge for enactive cognition and its emphasis on the exercising on sensorimotor skills (Aizawa, 2007). In a nutshell, if sensorimotor processes are essential for a specific cognitive function, then how does it happen that some individuals develop that cognitive function despite the case that such processes are different or absent in their case (for some comments on this concerning Piagetian developmental theories, see Segalowitz, 1980)?

However, this need not be a question of contrary evidence, but one that asks about the minimal and redundant mechanisms of the model in question. This is an instance in which the weak embodied position of the sensorimotor models comes good (recall Sect. 2.2.1). To Knott (2010), sensorimotor processes in part cause, but do in no way constitute linguistic structure. Instead, the crucial episode representation is grounded in working memory replay, which relies on embodied simulations rather than overt action itself. This takes it out of the line of criticism directed at enactivism by Aizawa (2007). It also leaves open the possibility of an explanation for cases of congenitally paralysed individuals with no apparent linguistic difficulty. For now, it is important that while daunting, this is a challenge that can be answered, rather than a principled case against the sensorimotor model.

The concerns mentioned above arise within the sensorimotor model. Other concerns question the model from the outside. An illuminating point to consider is language development. Knott (2010) offers a relatively

[17] Note, that this remains a legitimate method of evaluating knowledge claims (i) even if the privilege that it is granted by being the "scientific" method is withdrawn, and (ii) even where the utility of such a method is generally called into question.

individual-centred approach, in which meaning is primarily related to sensorimotor processes. Another position is that language learning must be considered in social terms (see e.g. Steels, 2003). Both approaches affirm two ideas; namely, (i) that language development is tied to acting and interacting – i.e. it is not a passive activity that mainly takes place in the brain; and (ii) that the foundations of language are acquired through this interaction – i.e. that they are not pre-determined. However, it seems (Knott, 2010)'s approach may not depart far enough from (ii) in the perspective of a social-interactive perspective on language development. The legacy of a kind of parameter setting persist in the sensorimotor model, in which observed regularities in linguistic input influence the agent to produce the components of a sentence in a specific sequence. Part of the difficulty, here, is that social interaction accounts of the grounding of meaning deal with concepts and words, rather than linguistic typologies (see also Steels, 2008). In effect, they look for meaning in different places.

Put another way, the sensorimotor and the socially-interactive approaches focus on different aspects of the meaning of an utterance. The former deals with underlying structures, the latter on words and concepts. Here, the open question becomes whether linguistic structure and its meaning, like words and concepts and their meaning, is necessarily based in social interaction. One the one hand, for any human child, the active learning of the sensorimotor model is always also social learning. On the other hand, the acquisition of linguistic structure might be adequately modelled without reference to this social dimension of the learning situation. In Knott (2010)'s model, for example, linguistic structure is strictly identified with sensorimotor processing. So, possible shortcomings of the model might be usefully related back to this omission of social learning.

At this point, it is an open question to what extent the sensorimotor model can and/or should be extended to also include social grounding, in addition to grounding via sensorimotor processes. Perhaps, behind the scenes of the model, social learning is simply assumed to be going on anyway. For example, there are accounts in terms of social learning of object categorisation and action recognition, not to sense of agency (see e.g. Bruner, 1990; Tomasello, 2009). The sensorimotor aspects of these processes are described in the (Knott, 2010), but these are not necessarily an exhaustive account of these abilities. From this perspective, social interaction does contribute to the sensorimotor model, but in the background, rather than as one of its main aspects

In this view, learning is both grounded both socially and in sensorimotor processes. While, in the first place, the sensorimotor model was not at odds with a social grounding of concepts. Moreover, the influence of conceptualisation and the development of concepts over time on the grammar of a specific language (as argued in e.g. Heine, 1997) does not have to pose a problem, because these effects can be argued to concern only the surface structure of that language. That returns us to a basic question about the sensorimotor model, namely the plausibility of a distinct, abstract representation of the underlying structure of all languages, such as the (sensorimotor) logical form in Knott (2010). Some arguments in favour of the idea were made in Sect. 2.2.4, but as many can be made against it. But this is a fundamental difference, and one which cannot be arbitrated here. The upside, at this point, is that the sensorimotor model can accommodate many of the conclusions of a social interaction perspective, rather than needing to contradict them.

The astute reader will doubtlessly have her own concerns in addition to those discussed above. Indeed, this has been only a brief discussion of some of the concerns that could be expressed. Some matters are open questions that can be addressed empirically, others are questions of emphasis and perspective, while yet others again are principled criticisms that cannot be explained away, because they reflect a difference in epistemology or ontology. Note, again, that the sensorimotor model is one among several proposals to account for the grounding of linguistic structure. Also, the data it is built on have been interpreted in other ways. The present work concedes all of this. However, there are good reasons to test the model at least, which is what this thesis attempts. In doing so, it contributes to an attempt to respond to some of the concerns raised above and to gather more data on the basis of which to assess the model.

3.4.2 Testable Predictions Based on the Sensorimotor Model

Like any model of similar detail, the sensorimotor model is difficult to test as a whole. Instead of trying, this thesis derives a number of specific, testable predictions from the model. The approach reported here is empirical. Using audio and video recordings, an eye-tracker, and a motion capture system it is possible to try to identify aspects of overt attention – and to some extent covert attention, too (see e.g. Benedek et al., 2017; Walcher et al., 2017). Inferences about attention are mainly related to eye-tracking data, since eye movements appear to heavily implicated in the direction of overt attention (D. H. Ballard et al., 1997). Meanwhile, motion capture provides a timestamp for movement onset, as well as a representation of some dynamics of the movement (e.g. velocity). This leads us to a particular testable prediction that we worked towards in this thesis:

1. during a reach-to-grasp episode, participants will typically attend to the target object in the early stages of the action, i.e. before the directed motor action is underway;

Based on this hypotheses, and the sensorimotor model as such, the thesis also proposes to investigate whether there is a systematic relationship between certain sensorimotor processes and the surface structure of language. That being the case, we formulated the following additional hypothesis:

2. for the proposed task, there is a systematic relationship between sensorimotor processes like saccades, fixations and overt movements, and parts-of-speech (e.g. agent, patient);

These hypotheses will be addressed again in Ch. 6. Though, this thesis did not undertake to test these hypothesis *per se*. Recall that, instead, the objective was to produce a novel data set, prepare data for future analysis, offer means to investigate it and to develop a further lines of inquiry based on the novel data. Before getting to this, though, the following two chapters will describe the methods (Ch. 4) and data analysis (Ch. 5). Both reflect the novel combination of Knott (2012)'s proposal and earlier work by Gross, Hirschmanner, Krenn, Neubarth, and Zillich (2018). To encourage further development of this approach, both of the following chapters include an evaluation about the design reported here, which could be modified to make it more suitable to test the various proposals outlined in this chapter. Again, (Knott, 2012)'s is an expansive model, as the scope of this chapter has shown. Now the focus shifts away from the theoretical background and on to how to test the model based on empirical data.

Chapter 4

Methods

The data collection reported here is a novel design, inspired by Schreitter and Krenn (2016) and Gross et al. (2018). This chapter will explain the task (see Sect. 4.2) and the set-up (see Sect. 4.3). Also, it includes a reflection on challenges encountere during this study and suggests how they can be addressed in future (see Sect. 4.5).

4.1 Participants

A 13 participant convenience sample (mean age = 24.08 years, range 21-28; 8 female, 5 male) was recruited from a population of university students. One participant was left-handed, the others (N=12) were right-handed. This study was conducted in English to facilitate the dissemination of the corpus, given limited interest in German-language corpora. Moreover, English was chosen for the sake of consistency with related empirical work (e.g. Webb, 2015). In order to participate, participants had to report an English-language proficiency of B2 level or above. English was the native language of one participant, the other native language being German (N=12). Participants were naive to the purpose of the study, and debriefed upon completion of the task. Prior to their participation, all participants gave informed consent.

4.2 Action Description Task

The overall goal of this data collection was to gather multimodal data on basic TAKE, PUT, and PUSH actions. The general set-up is similar to that reported by Gross et al. (2018). In this case too, participants received instructions to perform actions with three objects, placed on a table in front of them, and to verbalize these actions as they carried them out. In the instructions, participants were told: " *Your task in this experiment is to change the positions of objects and to describe the actions as you perform them. You are asked to do this as naturally as possible and to pay attention only to the task. Please look through the glasses normally and try not to get distracted. During the task, you will move objects from place to place within the marked area and describe what you are doing.*"

Participants were also given two example utterances ("I am reaching for Object X" and "I am pushing Object X to the edge of the table")In the present study, participants did not receive specific instructions about which action to perform when, or which aspects of an action to focus on. Participants did receive suggestions about what kinds of actions they could perform, with suggested verbalizations including: *I take the bottle and move it to the top left of the table.* Participants were instructed to return to a neutral position between actions and to continue performing actions until they were told to stop. They were informed that they would perform around 10 actions during each trial before being stopped by the experimenter. We chose this approach instead of a fixed number of actions to avoid participants performing what they would otherwise know to be the final action(s) differently.

Participants included both left- and right-handers, and both were instructed to perform actions only with their dominant hand. This reflects a concern that hand dominance may result from neural asymmetries in motor planning (Sabaté, González, & Rodríguez, 2004). Since motor planning takes a central role in Knott (2012), all actions performed with the non-dominant hand (N = 2) were excluded from the analysis.

Figure 4.1: The task is performed seated at a table, with the three objects placed centrally on the table.

4.3 Data Collection Set-Up

The data collected for this study include audio data (recordings of participants describing their actions), video data (recordings of participants field of view), as well as the eye-tracking data and motion capture data. Data collection was conducted at the Performance Science Lab of the Department of Music Acoustics – Wiener Klangstil, University of Music and Performing Arts Vienna (mdw). [1]

The user sits in front of a table (see Fig. 4.2 a.), wearing Sensorimotor Instruments (SMI) ETG 2 wireless glasses, which tracked eye gaze at 120 Hz. The glasses featured a custom arrangement of motion capture sensors for detection by the motion capture system. One marker was placed at each corner of the top of the frame. A third marker extended down from the left arm of the glasses, attached to a stick (see Fig. 4.2 b.).

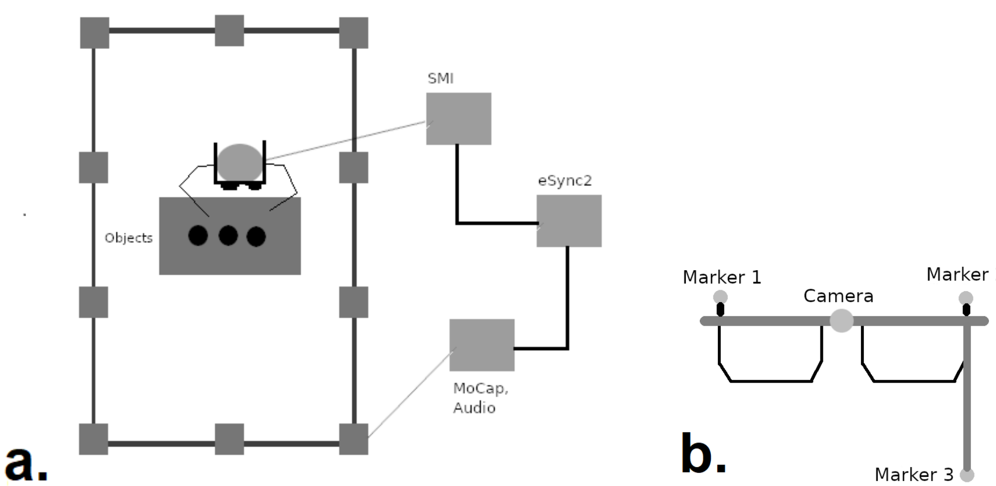

Figure 4.2: (a.) the set-up of the motion capture array and assorted recording equipment. (b.) eye-tracking glasses with centre-mounted camera and three markers attached

[1] https://iwk.mdw.ac.at/?newlang=en&PageId=43

Magnetic snap-on corrective lenses were available for participants requiring a distance correction (within the range -4.0 to +4.0 dioptres). Eye-gaze data was collected with a laptop connected to an OptiTrak eSync 2 device that transmitted TTL triggers from the motion capture software to the eye-tracking software to mark the start and end of a motion capture recording. The video recording from a camera mounted in the glasses constitutes the primary video recording in the Multimodal Action Verb Corpus.

For the motion capture, a 10-camera (Prime 13) OptiTrak motion capture system was used, tracking participants upper body and head at 240 Hz. Each performer was fitted with a custom arrangement of 28 reflective markers. Additional markers (4) were attached to each of three objects the actions were performed with, as well as to the table. Audio data were recorded using two microphones mounted on the ceiling, via a Focusrite Scarlett 18i8 sound card and recorded in Ableton Live.

4.4 Correspondence with Knott (2012) – Data Modalities

These methods allows us to investigate aspects of the sensorimotor model proposed by Knott (2012) (see Table 4.1). Specifically, this thesis aims to aspects of the model related to early visual attention to the cup, and concerning the events around the time of movement onset (see Sect. 3.4.2). Eye-tracking will be used to investigate visual attention (similar to the approaches reported in Webb et al. (2010); Bishop, Cancino-Chacón, and Goebl (2019). While motion capture will provide information on movement, in a manner similar to that reported by Bishop and Goebl (2018). One important difference to Webb et al. (2010) is that participants in the study reported here did not observe someone else perform an action (like they did in Webb et al. (2010)), but were recorded while they performed (and described) an action themselves. This allowed us to study utterance, eye-tracking and motion capture data from the same episode, without separating the data collection from different modalities across time. The first person perspective also means this paper can remain agnostic about the correspondence between action performance and action observation.

Table 4.1: Summary of data modalities of the LKG-Corpus and their correspondence to aspects of the sensorimotor model (Knott, 2012).

Deictic Operation	**Sensory Consequence**	**Measure in LKG**
Attend to the agent (interoceptive)	Attending to the agent (interoceptive)	*Not applicable*
Attend to the cup (visual)	Attending to the cup (visual)	Eye-tracking
Activate motor action	Reattention to the agent (motor)	Motion Capture
–	Reattention to the cup (haptic)	*Not applicable*

4.5 Data Collection Set-Up: Lessons Learned

Based on experience with this experimental set-up, some of the outcomes of this thesis are suggested improvements. Here, I propose both technical improvements, for example to increase data quality, and principled improvements, which aim to improve general aspect of the design, such as increasing reliability, validity, or generalisability of the results.

4.5.1 Technical Modifications

The placement of markers on the target objects ("Bottle", "Can", "ObjectX") led to difficulty during data processing. To deal with this, future versions of this study will use a different marker placement. Options include using more markers, specifically eight per object, rather than 4. However, this bears this risk of poor tracking quality given the proximity of these markers, as well as the interference of the markers of participants' hands. That loss in data quality was the reason such an arrangement was not used for this study.

One way to respond to this challenge is to use different hardware and software. For example, a developing cooperation that builds on this study would use Qualisys Miqus M3 or Oqus 7+ motion capture technology[2] to improve tracking accuracy. This would also allow automatic synchronisation of motion capture and eye-tracking data. This would further reduce the time required for data processing.

Indeed, synchronisation turned out to be a challenge in this study. Whereas, the alignment of different data modalities is essential for all subsequent tasks. Given that, it would be worth considering extending the role of the trigger box (see Fig 4.2) to include a synchronisation across all data modalities. However, this could pose a discouraging electrical and software engineering challenge. In that case, the issues of synchronisation alone might be reason enough to consider the use of different hardware in future.

4.5.2 Principled Modifications

In addition to technical modifications, future work would benefit from a number of other adjustments, as well. For example, issues with data quality led to an unexpected decrease in the sample size of this study. From the original 45 recordings, 22 were of sufficient quality to be included in the study – for the other 23 there with quality issues with either motion capture and/or eye-tracking data. Similarly, three of the original 15 participants had to be excluded completely. Especially the exclusion of participants raises the issues of missing data (cf. Rubin, 1976). It is easy, but not necessarily justified to assume that data quality in a given study is independent from the variables being tested. In the case of this study, participants were excluded both due to issues with eye tracking, and due to issues with the motion capture system. It was assumed that both kinds of exclusion were due to technical limitations of the data collection set-up. The final number of instances of objects being moved (N=209) was also only around half that reported for three different studies by Webb (2015). This raises concerns about the original sample size as well, particularly for any future work intent upon calculation various statistics relevant to the sensorimotor model. With this and data quality issues in mind, a larger sample size is recommended.

[2]see https://www.qualisys.com/hardware/ [accessed 25.05.2020]

A point to bear in mind going forward, is whether the task reported here is suitable to more specific research questions. When the time comes to test specific aspects of the sensorimotor model based on similar data, there are both advantages and disadvantages of the less constrained type of task we chose here. For example, this thesis cannot report an exact action onset, nor the time when a participant decided to manipulate a given object. Whereas controlling these aspects by presenting cues for participants to act on (e.g. visually presenting the next object to be grabbed on a screen) would have allowed us to define the onset of an action, in relation to the presented stimulus. On the other hand, such a design might distort participants habitual actions, which would raise concerns about the ecological validity of the study. The question here is about which design is appropriate for what kind of research question. Future work investigating specific aspects of the sensorimotor model may need to constrain participants actions more than was done here (see e.g. Webb et al., 2010). At the same time, less constrained behaviour as was recorded here, offers an wider perspective on the question of which predictions of the sensorimotor model continue to hold for adults, for example.

On another note, two issues concerning language are i) the low number of native language participants, and ii) the absence of a comparison across languages. In the end, only one of the total number of participants in the final study (N=13) was a native English speaker. This makes several avenues of linguistic research impossible. For example, being non-native speakers of English might influence the word choice as well as the choice of word order. It is not clear what significance this might have for studying underlying linguistic structure. At least, though, the present result may be more difficult to generalise. However, an English language corpus, even containing non-native speakers, may still also be more accessible internationally that a non-English resource, and is more easily compared to previous work that was also conducted in English.

Finally, maintaining the focus on English in research on the sensorimotor model would assume that the sensorimotor model is, in fact, universal. To test this though, similar studies would have to be conducted with speakers of different languages, especially languages with different grammatical structures (SVO, OVS etc., see e.g. Tomlin, 1986). From this perspective, the non-native English sample is cast in a different light, although the majority of participants spoke German as a native language, which follows the same word order as English – namely SVO, compare: "the woman grabs the cup" and "die Frau greift den Becher"). To respond to these issues, future studies might preferably recruit native speakers, or specifically study languages with a different word order.

Similar issues will be raised again in the final section of the following chapter (see Sect. 5.4), which outlines data processing. While some of the principled modifications developed based on methods and data processing were the same, both chapters also raise specific modification. Before getting to the difficulties involved, though, Chapter 5 will explain how the data acquired as explained in this chapter is represented (see Sect. 5.1) and processed (see Sect. 5.3).

Chapter 5

Dataset and Annotation

The data collected have been made available as the Linguistic, Kinematic and Gaze Information in Task Descriptions Corpus (LKG-Corpus: Reinboth et al., 2020). The Corpus contains all the data that was analysed for this project and is publicly available. [1] Table 5.1 summarises the data that the LKG-Corpus comprises.

Table 5.1: Summary of the current state of the LKG-Corpus. Each instance of an object being moved contains at least two actions (first TAKE, PUT/PUSH), though not necessarily two utterances referring to actions.

Participants	12
Recordings	22
Utterances referring to actions	324
Instances of objects being moved	209
Gaze Behaviours	4239

5.1 Representation of Information

Information recorded for each trial include audio recordings, eye-tracking, motion capture, and video recordings from the user's perspective. In addition to this, each trial is represented by:

1. manual orthographic transcriptions of speech (conducted in Praat [2])
2. a video from the participants perspective, automatically overlaid with a circle indicating the target of the participants' gaze by SMI software
3. schematics of the gaze vector every 5 frames (see Fig. 5.3), also compiled into an animation of gaze behaviour
4. information on gaze behaviours, for example whether a participant is looking at/close to an object (manually annotated)
5. information on whether the user's dominant hand touches an object (manually annotated)
6. information on when an object is being moved (manually annotated)
7. information on velocity and acceleration of participants' gaze [3]

[1] Open Science: This work is also an effort in open science, in the democratic, pragmatic, public senses (Fecher & Friesike, 2014). Democratic, in the sense of providing access to knowledge, pragmatic in that we hope to facilitate collaboration, and public, in that we aim to make the process of knowledge production transparent and open to available to scrutiny. Alongside the corpus, for example, we provide key scripts of code used to process and analyse the data. That way, not only the data, but also the tools for transforming the data into results is made available. This open code practice is a significant step towards reproducible results, among other things (Stodden, Guo, & Ma, 2013). For data sharing, there are similar reasons related to reproducibility, as well as utilitarian arguments related to the same research being carried out multiple times or the enabling potential of publicly available data (e.g. Poline et al., 2012).

[2] http://www.fon.hum.uva.nl/praat/

[3] This provides a continuous rather than discretised measure of gaze. This has the advantage, for example, that it leaves open the parameters of what counts as a visual fixation (see e.g. Rayner (2009) for parameters approximating a typical fixational eye-movement)

8. information on velocity and acceleration of a subset of motion capture markers [4]
9. segmentation of each trial into situations referring to action episodes – in our specific dataset, an episode begins after the last utterance referring to the previous action, and ends at the start of the first utterance referring to the subsequent action.

All raw data is provided as CSV files. In addition, the data from 1. and 4. to 8. (henceforth annotation tiers) is represented as an Elan[5] file (.eaf), exportable to CSV from Elan (for a look at ELAN, see Fig. 5.1). All animations are synchronised with the annotation tiers. In the following, we describe in more detail (a) the transcribed utterances (annotation tier 1, in Fig. 5.1), (b) the calculated gaze vectors (tier 2, 3, 8), (c) the annotated gaze behaviours (tier 4), and (d) the motion capture data (tier 5 to 7).

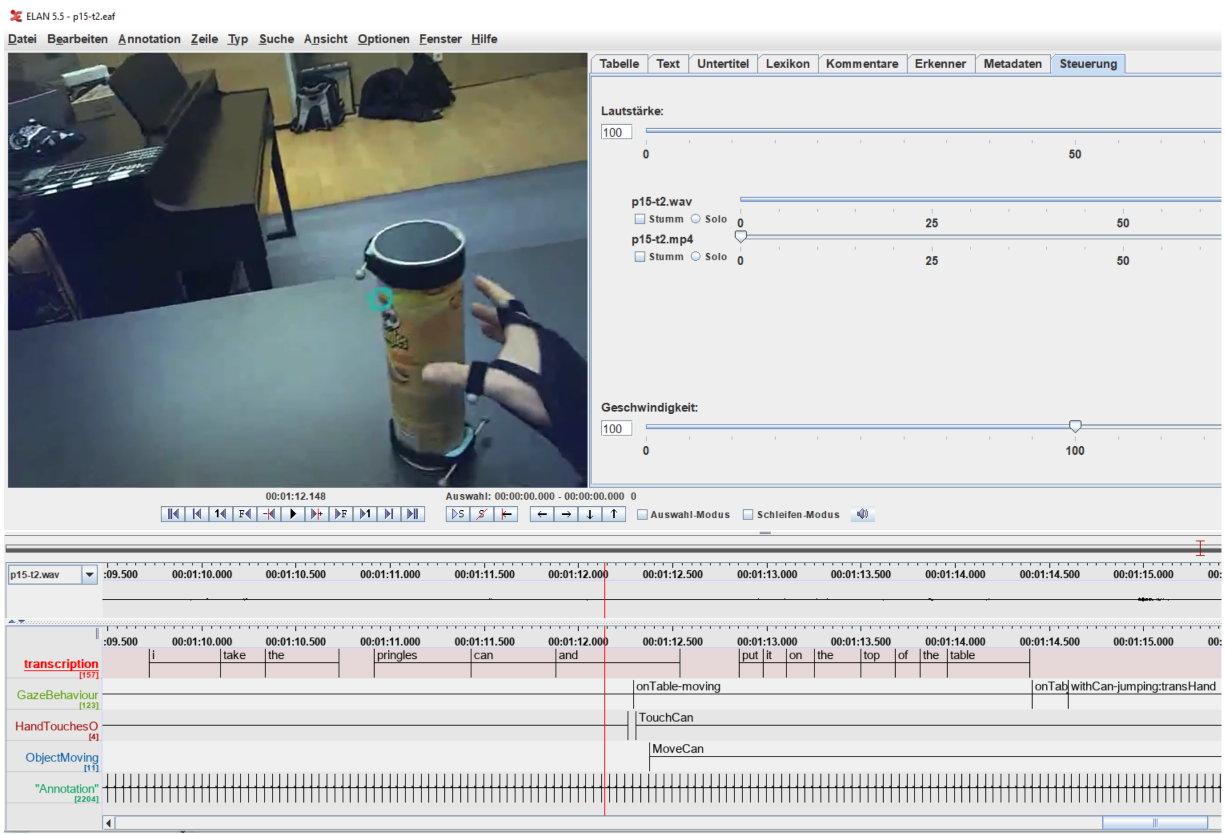

Figure 5.1: a screenshot of an ELAN-window. The top left displays the video, annotation tiers are represented at the bottom.

5.2 Action Segmentation

To analyse individual actions, we also performed a segmentation of each trial by a procedure similar to (Gross et al., 2018). In our case, we focused on situations, rather than individual actions. In this study, segmentation of the actions depends on the utterances. Specifically, the key aspect is whether participants used multiple verbs. For example, if a participant said "I take the can and put it next to the bottle", this is output as two actions; the first concerns the action "I take the can" and the second concerns the action

[4] This subset includes the markers of the head, and dominant hand and arm.
[5] https://tla.mpi.nl/tools/tla-tools/elan

"and put it next to the bottle". Meanwhile, the statement "I put the can next to the bottle" is output as a single action, since the utterance tier does not contain a reference to a TAKE action.

This produced an overlap between segments. The time window between the last utterance referring to the current action, and the first utterance referring to the subsequent action is included as the end of one situation, and as the beginning of the next situation. Given our design, it was impossible to determine the point between two sets of actions, where one ended and the next began. Recall, this was done intentionally as part of an effort to construct a design in which participants would be constrained as little as possible in terms of aspects like the timing of their actions. In fact, the possible overlap between post-movement events and pre-movement events is a salient factor in experimental design. Since we could not determine the end and start points of actions, the overlap allows us to consider the recorded data as either indicative of pre-movement (e.g. motor preparation, see D'Esposito, Ballard, Zarahn, & Aguirre, 2000) or post-movement actions (e.g. attending to target, see Knott, 2012). Segmentation was conducted in Python, using a procedure similar to the one reported by (Gross et al., 2018).

5.3 Summary of Data Modalities

For the LKG-Corpus, data was recorded from three modalities: linguistic, eye-tracking, and motion-capture data. Each of these is discussed in turn below.

5.3.1 Linguistic Data

The present study focused on utterances, rather than finer parts-of-speech. The LKG-Corpus comprises 369 (minus) utterances out of which 338 (minus) contain an action verb, such as 'I am grasping the can' and 'I am moving it to the middle of the table'. In the utterances with verbs, 15 different verbs are used by the 13 participants: 'move', 'grasp', 'lift', 'put', 'pull', 'push', 'take', 'hold', 'try', 'grab', 'stand', 'shift', 'pick', 'place', 'reach'. In 25 (minus) cases, the verb is missing, as in 'then the can next to the bottle'. In addition to utterances describing actions, there are four scene and sensor descriptions. Two utterances were excluded from the analysis, as they are only partially audible.

One point to note about the utterances of the LKG-Corpus is that they are largely in progressive aspect ("the woman is grasping a cup", as opposed to "the woman grasps a cup"). This is different from the results reported by Webb et al. (2010), in which participants did not describe actions while they were performing them. In fact, the case of progressives poses a particular challenge to the sensorimotor model (Knott, personal communication). This is because the model assumes that, stored as a working memory representation, the sensorimotor sequence that corresponds to the utterance being produced is replayed as the basis for that utterance. During the task reported here though, the episode is still underway, so there can be no complete working memory representation of this particular episode, yet. Now, parts of the utterance may be associated with sensorimotor operations that have not yet occurred. For example, the agent might describe that "I am grasping the cup", but do so prior to attending the cup, and therefore prior to the sensorimotor operation that is the basis of the patient object. This suggests there is some additional memory mechanism that allows the agent to refer to a sensorimotor operation that has no yet taken place (Knott, personal communication). On the other hand,

5.3.2 Eye-Tracking Data

Prior to analysis, the in-built functions of SMI software were used to compute the target location of participant's eye gaze. A small circle to indicate this point, including a margin of error, was inserted into the video recorded from the front-facing camera mounted in the eye-tracking glasses (see Fig. 5.2). The area of the circle was defined by the SMI-software independently of the study design.

Figure 5.2: Participants' gaze was reduced to the recording area of the camera for analysis. The orange circle indicates the area that is calculated as being the current target of the right eye.

5.3.2.1 Gaze Vectors

A vector was calculated that indicated the direction of the gaze for the right eye of each participant. From this, a "gaze target" was identified, either one of the three target objects ("Bottle", "Can", "Object X"), the table ("Table") or elsewhere in the room ("Other"). This gaze vector is based on a precise calibration of the glasses when the participant first puts them on. Given the length of a recording session (15-45 minutes), a degree of imprecision in the gaze vector coordinates resulting from calibration errors (i.e. if participants bumped or shifted the glasses) cannot be ruled out.

An automated procedure for identifying gaze targets was adapted from Bishop et al. (2019). This involved remapping gaze coordinates into motion capture space. First, all data were temporally aligned using the recorded trigger values (see Subsect. 4.3) and downsampled to the lowest used recording rate (audio recording – 100 Hz) by linear interpolation.

In the following step, all points in the motion capture space were rotated around the origin of the gaze vector (i.e. the right eye) using a 4 x 4 rotation matrix. The three dimensions of the matrix were defined by the pair of markers on top of the glasses (x-component), the axis defined by the two markers on the top and bottom left of the glasses (y-dimension), and the axis defined by the normal vector to the plane containing the three glasses markers (z-component). This allowed the generation of a schematic representation of each frame (see Fig. 5.3)

In every case, the direction of the gaze vector had to be corrected to compensate for imprecision in the placement of markers on the glasses, and that of the glasses on the participant (i.e. imprecision in gaze vector origin). For this, a separate 3 x 3 rotation matrix was constructed for a set of angle values, based on an automated optimization algorithm. These angle values reflect the extent to which coordinates were shifted in the corresponding dimension, e.g. one unit to the left, or two units downwards.

Starting from 0, and alternately sampling from the ranges [1:13] and [-1:-13] the optimisation algorithm computed the total number of frames in which a target object was identified as "gaze target" by the original procedure. For each range, the procedure ended after the total number of frames was lower than the current maximum number of frames on two consecutive samples, i.e. when the categorisation was worse than it had been before twice in a row. This approach relies on the assumption that the most accurate configuration is the one with the highest number of frames annotated as the gaze being on an object. To check this was the case, the adjusted eye gaze vectors were tested by visual comparison with the video and gaze marker display. This revealed a number of difficulties, which led us to annotated gaze data manually instead.

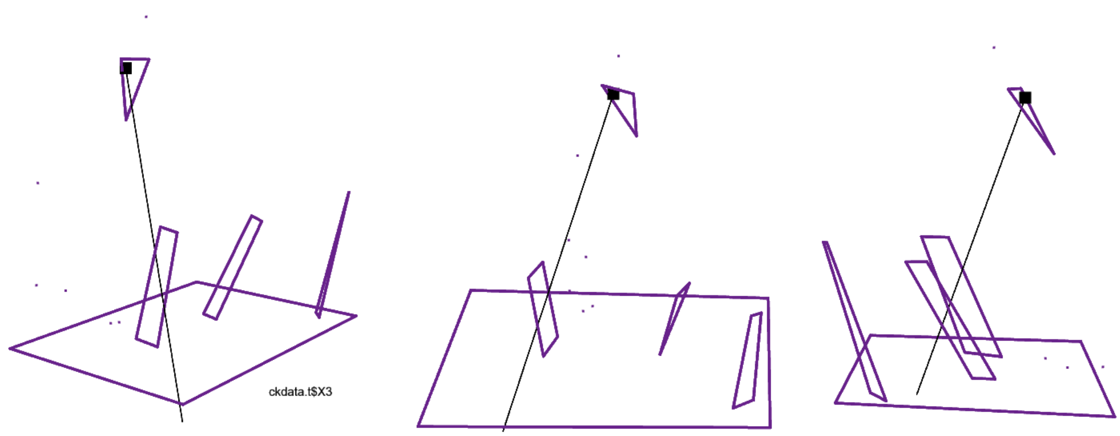

Figure 5.3: Gaze vectors were computed originating from the right eye of the participant. The opposite end point of the gaze vector is not informative, since Z-coordinates, representing the depth of the area of focus, were not reliable enough to include in the analysis. The small black dots correspond to the motion-capture markers on the participants right arm.

The automatically generated schematics and animations are accurate to the recorded data, however. This means they remain a valuable additional representation. The animations of gaze behaviour are based on these schematics. To generate the animations, a schematic was saved every five frames, leaving the animation with a frame rate of 20fps. The individual schematics were combined into a single animation using the free VideoMach software.[6]

Missing Data: Throughout recordings, some participants eyes could not be tracked for brief periods of time. A number of participants (N=2) had to be excluded due to unreliable eye-gaze recordings.

5.3.2.2 Gaze Behaviours

Manual annotation of eye-gaze behaviours (annotation tier 4) became necessary because of persistent irregularities of the representation of the objects in motion capture space due to marker placement. Specifically, target objects were represented as asymmetrical shapes that were insufficient for analysis (see Fig. 5.3).

[6]http://gromada.com/videomach/, [accessed: 20.05.2020]

This led to a number of difficulties. One the one hand, objects could be turned so that the surface area of the plane that represented them was orthogonal, rather than parallel to the participant. In this case, a participant's gaze would not be annotated as intersecting the object, even when the video clearly showed that her gaze was on/through the object. On the other hand, the asymmetry of the representations made it difficult to construct virtual 3D-representations out of the recorded data, i.e. the planes the objects were represented as. This made the transformation of the plane-like into a volume-like representation uneconomical. As a result, we were prompted to consider manual annotation.

For this, we identified distinct patterns of eye movements from the video of the front-facing camera mounted on the glasses. To do so, we developed a label set (see Table 5.2), built on consideration of existing data on eye movements during tasks similar to our own, e.g. D. H. Ballard et al. (1997); Webb (2015). The development of the label set was an iterative process. Three authors of Reinboth et al. (2020) (TR, SG, BK) developed categories as a team, annotated the data independently and finally annotated data together in a forced choice process to ensure data quality. This process went through several iterations, during all but the last of which the label set was revised and retested. This is a characteristic process for the development of a coding scheme (Weber, 1990).

Table 5.2: Coding scheme for gaze behaviours – Primary labels distinguish three different types of eye movements: gaze on object, from/to a certain location, and with an object/hand being moved. There can be up to two reference objects per annotation. Further labels describe features of the gaze behaviour including whether it is (a) smooth, (b) jumping, (c) stable, (d) crossing an object without change in speed (trans), or (e) alternating between two locations/objects. Aspects of the scene include whether another object is close to the current gaze location.

Primary Label	Reference Object	Secondary Label	Addition	Supplement
on	$Object^1$	stable	$transObject^{1\text{-}n}$	
	$Object^1Object^2$	moving	$closeObject^{1\text{-}n}$	
		jumping	$altObject^{1\text{-}n}$	
to	$Object^1Object^2$	smooth	$transObject^{1\text{-}n}$	
	$Object^1*$	jumping		
	$*Object^2$			
with	$Object^1$	on	smooth	$transObject^{1\text{-}n}$
	$Object^1Object^2$	close	jumping	$closeObject^{1\text{-}n}$
		trans		$altObject^{1\text{-}n}$

Since categories were updated according to patterns in the data, the method is partly inductive. When research informs the development of new theories, as is the case here, such a method is beneficial (Rivas, 2018). One challenge was that of identifying the appropriate level of granularity at which to annotate. An iterative approach allowed us to continually refine our label set until it fit the data. This is a common and effective pattern of developing novel coding schemes (Weston et al., 2001). In contrast to automatic annotation, manual annotation allowed us to distinguish between eye-movements in a more qualitative manner.

This way, we were better able to represent the rich dynamics of visual behaviour. For example, we distinguished various forms by which a participant's gaze could traverse an object; for instance, her gaze might simply pass over the object during a voluntary saccade, or she might move the object through her field of view while fixating a single point on the table. These instances of the gaze being on the object during a given frame would be not be distinguished by a simple automatic annotation. By annotating manually, we succeeded in extracting features we would not have distinguished otherwise.

5.3.2.3 Conventions for Gaze Behaviours

To deal with ambiguous situations in the gaze videos, a number of conventions had to be introduced in order to ensure consistency of annotations. In this section the term "object set" refers to the following objects: BOTTLE, CAN and OBJECTX. The conventions relied upon during annotation were as follows:

- when the circle indicating the gaze target was directly next to an object from the object set, this was annotated as "on" as long as the circle partially overlapped the object.
- no annotations lasting less then 50ms.
  - exception being some short "to" annotations that were unmistakable. The duration is related to lower limit of the duration a fixational eye movement during which some further processing can be expected to take place.
- no "to" was annotated from AND to the same object of the object set
- "to" was annotated from AND to either ENVIRONMENT or TABLE. This was done to capture instances when the gaze moved to a (future) target location (i.e. where an object would be placed next)
- while an object from the object set is being moved, a "with" rather than "on" was annotated if in doubt
- while a participant's hand touched the object from the object set, but this object did not appear to move, then "with" annotations ended as soon as the direction of the eye gaze went against the direction the object had been moved in
- no "close" annotations for either TABLE or ENVIRONMENT
- anything was annotated as "close", if the distance to the the "'close" object was less than twice diameter of the gaze circle
  - exception being "with[object]-close" annotations for objects in the object set, as long as the gaze is moving in the same direction as the object. In this case, gaze often often jumps further ahead of the object than twice the diameter.

5.3.3 Motion Capture Data

The present motion capture data were recorded from an adapted skeleton with 29 upper body markers, each represented by a set of x-, y-, and z-axis coordinates (see Fig. 5.4). The standard upper body + hand skeleton offered by the *OptiTrak Motive* software was modified by removing two markers from each hand, one from the index and one from the little finger. This was done after piloting indicated that it improved tracking accuracy of finger-movements. It should be noted that removing markers can negatively affect overall tracking, since the skeleton is best tracked when all markers are visible.

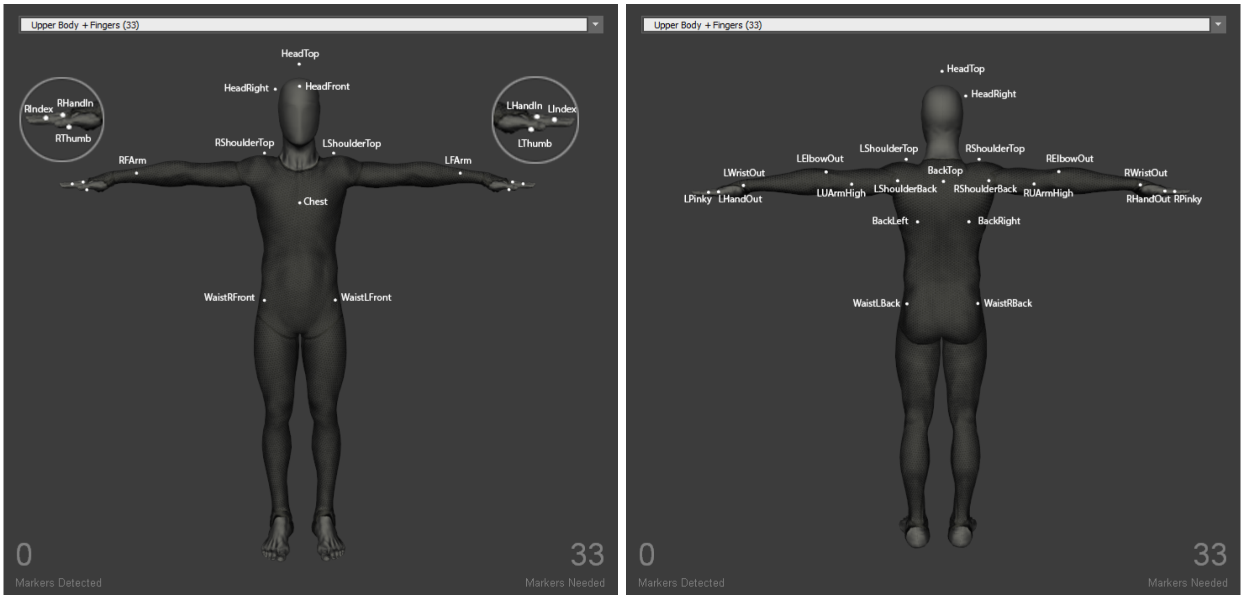

Figure 5.4: the above configuration of markers was used to record motion capture data. This configuration was preferred over one with more markers on the hand to increase data quality.

5.3.3.1 Manual Preprocessing

Each instance of motion capture data was inspected visually, to search for labelling errors. Where errors occurred, markers where relabelled whenever possible. This involved changing the marker assigned to a given data point. For example, while the participant's right hand was close to an object, the software might assign the label referring to one of the object markers to a point on the participant's hand, and vice versa. To correct this, the labelling was adjusted manually. At other times, such relabelling was not possible. In this case, short segments of the data were deleted and filled in automatically (max. 150ms). The resulting trajectories of individual markers were visually inspected for plausibility. In addition to this, individual markers were occasionally untracked. Where the duration of untracked periods was less than 150ms, these gaps were also filled automatically. Longer gaps could not be filled in a manner than ensured fidelity to marker movement. An exception to the last point are untracked markers of objects that were not moving. In this case, untracked period of any length were filled automatically.

Missing Data:some of the trials included in the final analysis (N=6) contain periods during which one or more markers were untracked. These segments were excluded from further analysis. Since such untracked periods occurred during a single action, they did not warrant the exclusion, from the analysis, of the remainder of the trial.

5.3.3.2 Motion Annotations

For each trial, the velocity and acceleration of a subset of the makers were calculated in RStudio and exported to ELAN. The subset of markers were on the participants head, and the dominant arm and hand (specifically, the markers: “HeadFront”, “RThumb”, “RHandIn”, “RWristOut”, “RElbowOut”, “RFArm”, “RUArmHigh” – see Fig. 5.4).

Velocity and acceleration were each calculated over time windows of 50ms. To do so, we calculated the distance between the point where the same marker was at successive measurement times. The distance is calculated between two sets of three dimensional coordinates, as shown in Fig. 5.5. The distance between coordinates was converted to velocity by dividing it by the change in time, which corresponded to the frame rate of the (downsampled) motion capture recording (100Hz, i.e. 0.01s). For acceleration, we calculated the difference between successive velocities and divided that by the change in time over the course of the two measurements of velocity. To smooth the values for velocity and acceleration, we calculated the mean value of both properties for each successive time window of 50ms.

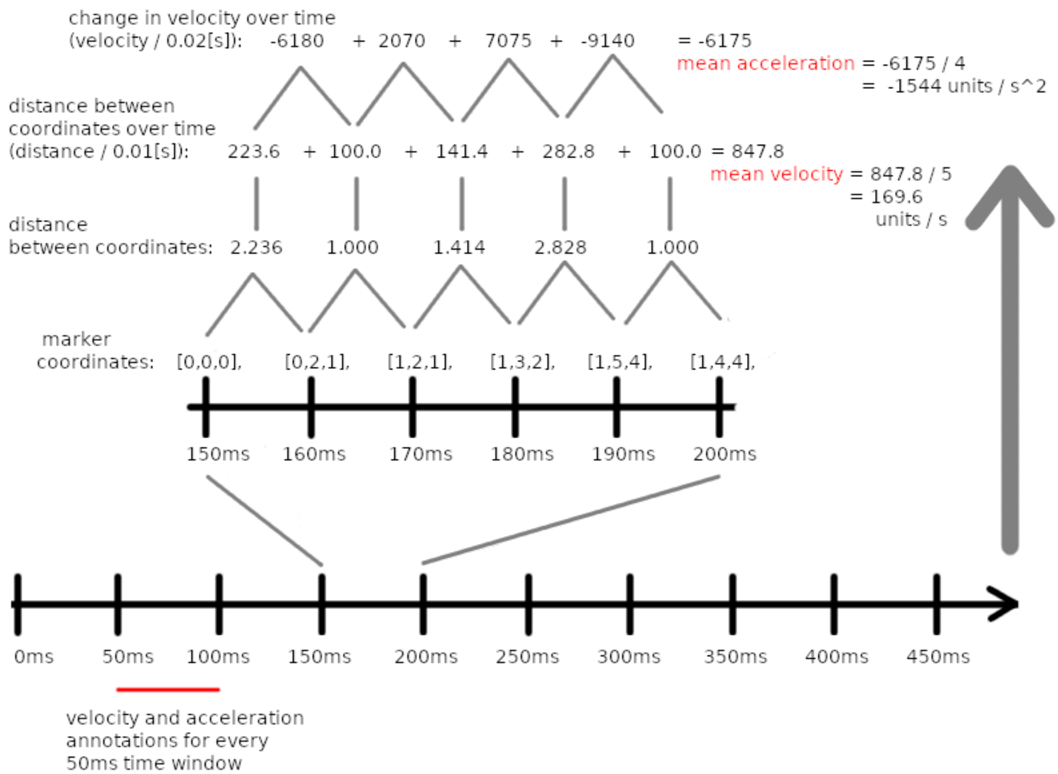

Figure 5.5: to compute velocity and acceleration of the limbs and joints, we used the change in position of motion-capture markers. Velocity and acceleration were calculated over an interval corresponding to the frame rate of the eye-tracking data, for ease of synchronisation. Velocity and acceleration were calculated based on the difference in 3D-position of each marker between the start and end-point an interval corresponding to that frame rate. It is assumed that both features remained constant during a single frame of the eye-tracking.

5.3.4 Extraction of Data Modalities For Analysis

To bring together the different data modalities, an R-script was written that takes the different modalities as input. Using this script, different features of the LKG-corpus can be extracted for analysis. This is done by specifying the data of interest using a 8 digit vector, according to the scheme outlined in Table 5.3. For example an 8-digit vector could be specified to extract from all trials the time periods during which the velocity of the "HandIn" motion capture marker (see Fig 5.4) was above a certain threshold, as follows: [2, 0, 0, 1, 3, 1, 1, 0]. This corresponds to selecting “Data”, “*none*”, “*none*”, “motion capture”, “HandIn”, “velocity”, “above threshold”, “*none*”, according to the feature outline in Table 5.3.

A variable containing the extracted data can be saved for later analysis. The script allows the data corresponding to any number of these 8-digit vectors to be exported simultaneously and stored together. The extracted data contains the time onset and offset of each data point and is labelled according to trial and segment. The data can be loaded later, for example to investigate the relationship between two different data modalities.

Table 5.3: Features of the data extraction tool developed to facilitate analysis of LKG-corpus data. I. corresponds to which data should be used for a variable, whether (1) from ELAN or (2) 'raw' data; II. corresponds to which ELAN tier to focus on; III. corresponds to which gaze behaviour annotation to focus on; IV. corresponds to which type of data to use; V. corresponds to which marker to focus on; VI. corresponds for motion data, to which value to analyse; VII. corresponds to whether to use all data points above or below threshold, i.e. only fixations (eye movement ¡ threshold); VIII. corresponds to whether or not to limit analysis to the 'main object' of the segment, i.e. object being moved; the value 0 can be entered as a placeholder when *none* of the possible values should be selected.

Position in vector	Possible **0**	Values **1**	**2**	**3**	**4**	**5**
1.	–	Annotation	Data			
2. (only annotation)	*none*	HandTouches Object	Object Moving	transcription	Gaze Behaviour	*none*
3. (only Gaze Behaviour)	*none*	on	to -on	withObject -close	withObject	stable
	6	**7**	**8**	**9**	**10**	**11**
	moving	jumping	smooth	close	alt	trans
4. (data only)	*none*	motion capture	eye-tracking			
5. (motion capture only)	*none*	Head	Thumb	HandIn	WristOut	ElbowOut
	6	**7**				
	Arm	Arm High				
6. (mo cap only)	*none*	velocity	acceleration			
7. (data only)	*none*	above threshold	below threshold			
8. (only annotation)	*none*	only main object	all objects			

Early attempts to develop a similar tool with which to facilitate the visualisation of data in a similar way were unsuccessful. The variety of data contained in the LKG-Corpus and the corresponding breadth of relationships between these data that might be investigated would have required creating a software tool beyond the scope of the present project.

5.4 Data Processing: Lessons Learned

With respect to data processing, too, this thesis offers a number of lessons for future projects. Especially given the early stage of this research, these are important outcomes that can facilitate future work. Just like the suggestions concerning methods (see Sect. 4.5), this section outlines both technical and principled modifications.

5.4.1 Technical Modifications

One technical improvement in data processing would be automatic annotation, for example of the target of the participants gaze (as was attempted here). To do so, future studies could build on a synchronised hardware set-up, that combines motion capture and eye-tracking during recording. For example, this could be achieved by integrating a Qualisys motion capture set-up with Tobii Pro eye-tracking glasses.[7] This would simplify the processes of translating the data from one modality (motion-capture or eye-tracking) into the other. So, one of the challenges faced by this study was translating coordinates from the coordinate space of one modality into the space of the other modality.

Another improvement would be to optimise the search algorithm that was used to find the best values with which to correct the eye-gaze data – specifically the subsequent correction of the focal point that is originally computed by the SMI-software (see Sect. 5.3.2). The current procedure is importantly based on two assumptions. First, that the correction that maximises the number of frames during which an object is tracked, is actually the best fit to the participants behaviour. Second, that if two successive adjustments of the correction (i.e. from two to three, then to four) produce worse results, then further testing in that direction (e.g. beyond a value of four on the x-axis) is unnecessary. It is also assumed that this procedure avoids local minima (i.e. a set of values at which a high number of frames have the participant's gaze intersect an object, but which nonetheless isn't the best possible set of values). In part, this is based on the additional assumption that the best correction values will tend to be close to zero, given based experience with the hardware involved. It is likely that this approach was sufficient for the current study, but future work might benefit from a more robust search algorithm.

5.4.2 Principled Modifications

A number of modifications were already put forward in Sect. 4.5.2. One different point that arises about data processing concerns the more qualitative approach to participants' gaze behaviours. Here, taking this perspective from the beginning, rather than in response to technical issues might have produced a deeper analysis. Especially in the context of cross-disciplinary research, one future line of research could take a very different approach and explore participants experience during tasks like the one reported here. In fact, there are even recent corpus-based approaches at the intersection of linguistics and phenomenology (e.g. Jugnet & Lhôte, 2019). Though, future work on the subject of the sensorimotor basis of language might be more closely related to Croft (2007)'s account of grammar in terms of the verbalisation of experience, or even to Zlatev, Blomberg, and David (2010)'s investigation into the intersections of linguistics and phenomenology of motion. If nothing else, a lesson learned during the manual gaze behaviour annotations is that the eye-gaze data that was recorded is an ambiguous record of participant's visual experience. Taking this thought further, and specifically when considering the role of participant's self-reported experiences, questions arise about the relative importance of conscious and unconscious processes, and even the (un)availability of certain cognitive processes to introspection (Zlatev, 2010).

[7] https://www.tobiipro.com/product-listing/

That is why, to deepen this analysis, bringing in phenomenological methods and epistemology could be an excellent contribution. Indeed, this could be to the benefit of a general linguistics, one that spans both neuro- and psycholinguistics, as well as traditional linguistic perspectives (Zlatev, 2016). Moreover, drawing on phenomenology in this way might facilitate a turn away from the mentalist emphasis the approach proposed here might be seen to have inherited from Chomskyan linguistics. While it is not clear what exactly what such an approach might contribute to testing the sensorimotor model, that is not the point. Instead, a phenomenological approach might present a successful countering force towards the "philosophical naiveté" of cognitive linguistics (Zlatev, 2010). For example, the presence of ongoing phenomenological inquiries into the phenomenon of sensorimotor involvement in language could encourage a broader perspective in natural scientific projects. From this perspective, there would be research with a different epistemology to reach out to and engage with, with an enriching effect on both avenues of research. In this way, such research, while value in its own right, makes the additional contribution of encouraging connections across the lines of disciplines and departments.

With this thesis, I have not made any attempts in this direction. Rather, I want to point to potential that qualitative and experience-based approaches offer concerning the topic of this thesis, especially because that has been discussed much more in computational terms. The next section continues this trend, since research on the sensorimotor model remains at the beginning of a large project. To contribute to testing the model, Chapter 6 introduces ways to visualise the LKG-Corpus data (Reinboth et al., 2020) and outlines how specific predictions of the sensorimotor model could be tested on that basis.

Chapter 6

Investigating the Sensorimotor Basis of Linguistic Structure

This chapter will outline different analyses that can be conducted based on LKG-Corpus data (Reinboth et al., 2020). As mentioned at the outset (see Sect. 1.3), this thesis contains no statistical analysis. Instead, I first introduce a way to visualise sensorimotor events. One of the aims with this tool was to capture the dynamic of an unfolding action, both to help test the sensorimotor model, and to generate predictions about how attentional, linguistics and sensorimotor processes may be related.

Building on these visualisations, the following sections suggests different empirical studies. Among these, I first outline future directions concerning fundamental (basic) research into i) the sensorimotor basis of language, and ii) linguistic structure specifically (see Sect. 6.2). These are two different lines of inquiry. The former is more general, while the latter specifically explores Knott (2012)'s sensorimotor model.

Then, the Sect. 6.3 discusses applied research, focusing on two specific applications. These are language structure learning (see Sect. 6.3.1) and robotic action 'understanding' (see Sect. 6.3.2). In each case, future work, benefits from a compact visualisation of the LKG-corpus data, as will be illustrated with examples generated using the software tool introduced below.

6.1 Visualising Sensorimotor Events

Fundamentally, each recording was decomposed into a number of action segments, each of which contains at most one utterance referring to an action. Absent a stimulus to mark the onset of an action, the action segmentation performed on the basis of the utterances plays a central role in this chapter (for details on the method, see Sect. 5.2). Below, these segments are represented graphically (for an example, see Fig. 6.2). In this case, the relationship being visualised is that between participants' right hand movement velocity and times when participants looked at the object they were grabbing / about to grab. This reflects Knott (2012)'s *Ordering Constraint 2*, that during execution of a reach-to-grasp action, the agent typically attends to the target object in the early stages of the action (i.e. before movement is properly underway).

To represent this, the focus is on continuous rather than discrete data. However only comparing the quantitative data is of little use in certain circumstances. [1] Consider Fig. 6.1; for a single segment, it shows the relationship between the velocity of one participant's eye movement and that of the movement of their right hand. One difficulty is that, while low eye movement velocity may indicate a fixation on an object, this is difficult to interpret without more information. Indeed, gaze behaviour annotations indicate that participants' gaze often lingered on an object without remaining stationary. In terms of movement velocity, such instances would be recognised as an action of attention to that object, since the gaze is not stationary. Though, the video may show the participant looking at a single object. So, one challenge is how to represent the data recorded here to make it more easy to interpret.

[1] this is also due to issues with the automatic annotation of eye-gaze data,

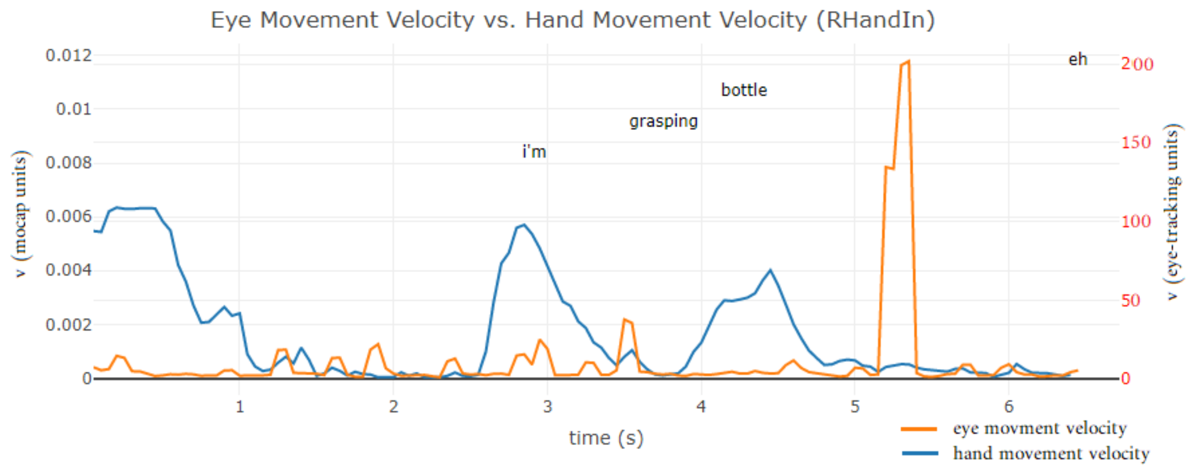

Figure 6.1: The first segment of trial p1-t1, showing hand movement and eye movement velocities. Low eye-movement velocity indicates a fixational-eye movement, but these are difficult to interpret absent information what the participants is looking at.

A different example is that action onset can be difficult to define categorically. A simple operationalisation might refer to the onset of hand-movement, i.e. the increase in the velocity of the marker after a period of rest. However, this only reflects movement onset. The action itself though also has attentional components related to action preparation (recall Sect. 3.1.2) [2] , meaning it should be considered to begin well before overt movement is visible. In fact, it is not clear that a discrete moment in time can be identified at which an action begins. This, after all, is what was discussed as the cascading unfolding of a deictic routine.

Now, the challenge of accounting for action onset, too, is not solved by representing the continuous motion capture data. However, this representation does avoid a misleading representation of participants behaviour, in terms of concrete action onsets and offsets, which might suggests that i) such points in time exist, and ii) that they can be identified based on the data – though, as it stands, neither appears to be the case. As a result, where gaze behaviour annotations replace continuous data on eye movements in subsequent visualisation, participants actions will continue to represented via the motion capture data on their movements, since any other representation risks a misleading reduction of "action" to "movement".

So, Fig. 6.2 shows the relationships between the gaze behaviour "onObject" and hand movement velocity. It represents time periods during which the participant looked at the target object, and puts these in relation to action. These two data modalities were chosen as a way to visualise variables relevant to Knott (2012)'s *Ordering Constraint 2*. In this regard, the initial increase of hand movement velocity during each segment could be interpreted as the action onset. While, the "onObject" gaze behaviour annotations might be interpreted as visual attention to that object. [3] It appears the participant looked at the object ahead of, or in the early stages of some movements. This is one point from which to start future work.

Also, Fig. 6.2 demonstrates the overlap between subsequent action segments (for a description of the segmentation procedure, see Sect. 5.2). Recall, a segment is the time span during which one utterance referring to an action occurs. That means, a segment begins some time after the last utterance of the previous

[2]This action preparation is not thought of as a deliberative process. It is entirely consistent with the meaning intended here if action preparation occurs without the participants explicitly reflecting on which action action to performing.

[3]The only "onObject"-annotations included in the figure are those in which the participant was looking at the object that is the target of the current action (as indicated by the utterance).

segment, includes one (and only one) utterance referring to an action, and ends immediately some time prior to the first utterance of the next segment. As can be seen in Fig. 6.2, the beginning and end of each segment are a few seconds in which the participant was not speaking. Subsequent segments usually overlap during these time windows, unless a participant was silent for a long time during actions. [4]

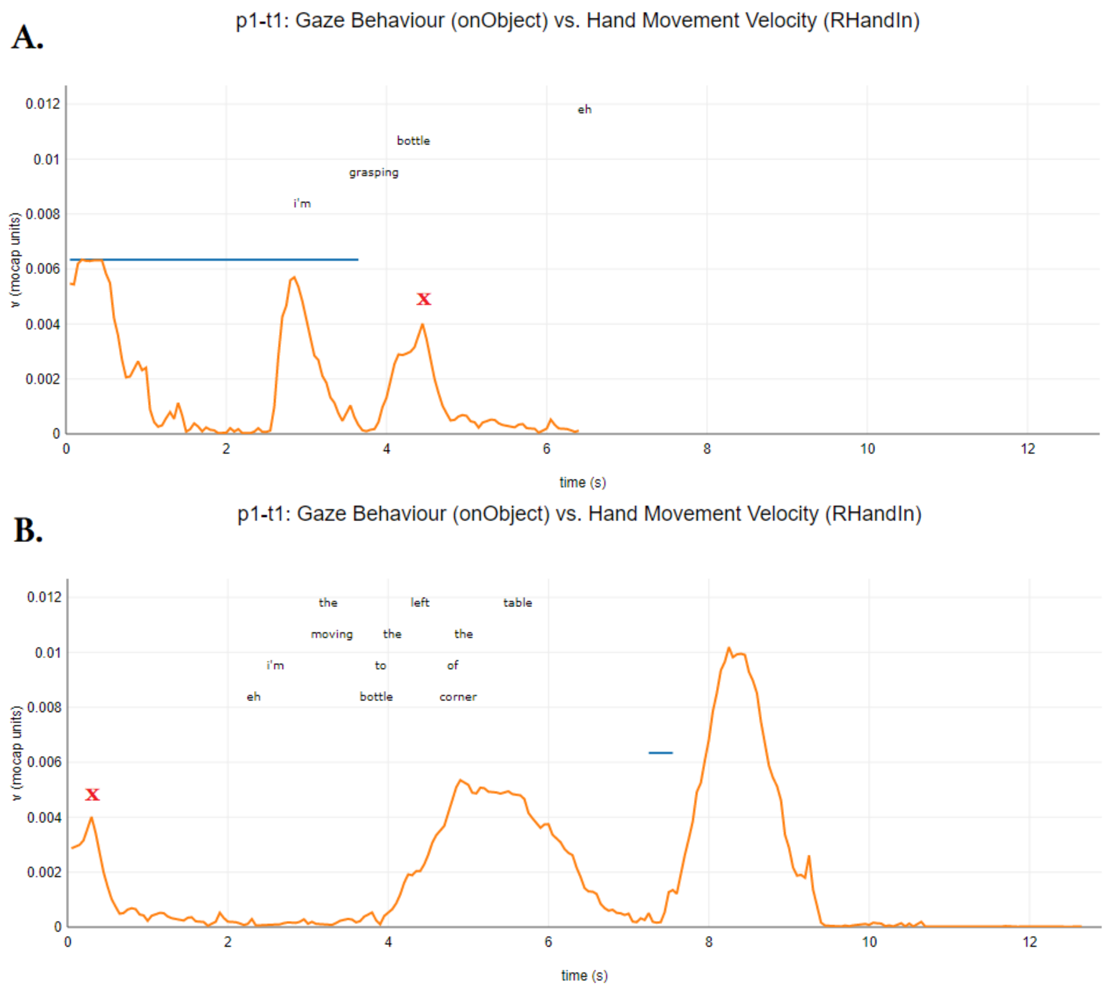

Figure 6.2: To exemplify the visualisation proposed here, **A.** shows the first and **B.** the second segment of the same trial. The utterances are placed on the figure so that the first letter coincides with the onset of the utterance. The vertical arrangement of the utterances was chosen for better legibility. The figure also demonstrates the overlap between subsequent action segments. The red “x” on **A.** and **B.** is the same point in time during the trial, present in both since the segment overlap.

Collecting the overlapping segments represents a complete trials (see Fig. 6.3; representative visualisations are also included in Sect. 9). Like comparing segments, comparing trials reveals various differences: patterns in utterances, idiosyncrasies in the dynamics of movements, or particular sequences of eye-movements that appear specific to certain individuals. Each of these points offers an avenue for future work. Indeed, the software tool to generate these visualisations was developed to facilitate hypothesis generation. Such hypotheses can then be tested by specific research designs, inspired by visualisations like Fig. 6.3.

[4]Since the segmentation was adjusted slightly for each participant to improve performance, the length of the periods of silence on either end of the utterances in each segment varies somewhat across the different trials.

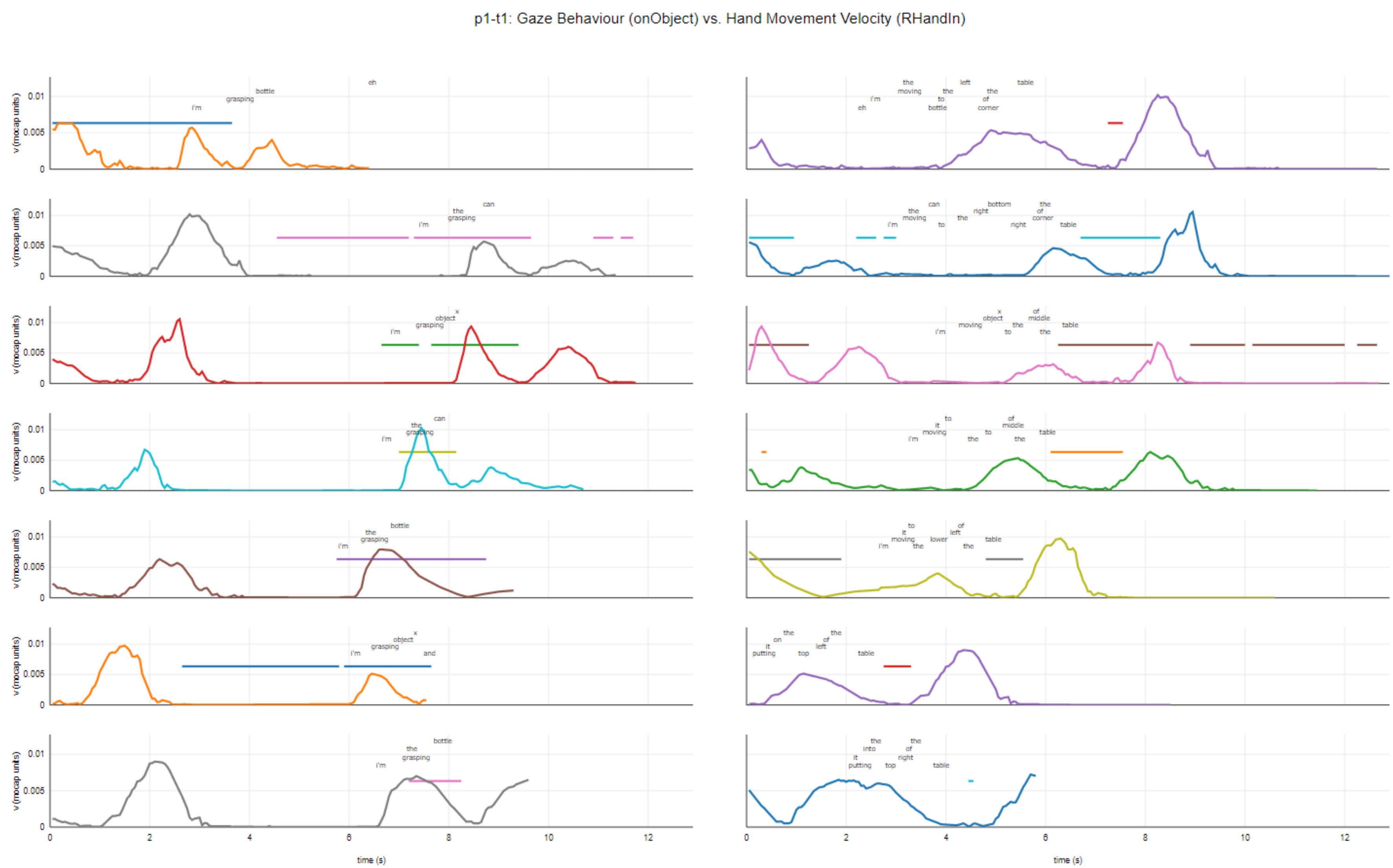

Figure 6.3: Overview of an entire trial of one participant. It appears this participant looked at the target object ahead of or early during every reach action (all the left-hand segments, in this case). The put actions (right-hand segments) also show some similarity (especially the top four), but appear more heterogeneous in general.

6.2 Suggestions for Empirical Studies

Based on the visualisation, one set of hypotheses and predictions concerns basic research, either concerning the sensorimotor basis of language generally, or specific to Knott (2012)'s sensorimotor model. In the following, Sect. 6.2.1 outlines how the LKG-corpus will be developed further in order to pursue broad questions about the structure of actions. While, Sect. 6.2.2 and Sect. 6.2.3 outline future studies investigating (Knott, 2012)'s sensorimotor basis of linguistic structure. Finally, Sect. 6.2.4 departs from this line of research, and deals with word order (comparable to Chomskyan PF rather than LF, see Sect. 2.2.4).

6.2.1 Future Directions for Investigating the Sensorimotor Basis of Language

Questions Raised by Existing Data

One question that might be addressed based on this data, is whether there are systematic difference across trials during which participants either did, or did not look at the target object before reaching for it. One possible factor is whether a participant was looking at another object immediately prior to action onset. To this end, more object than one could be represented in the visualisations (see Fig. 6.4). A different option would be to investigate decision processes during the LKG-corpus task. Visual inspection of the videos suggests participants may have paused after a grasping an object in order to decide where to place it. Given the information on where an object was later placed, future work might try to better understand the process with which the target location is selected, or to identify possible predictors for which location will be chosen.

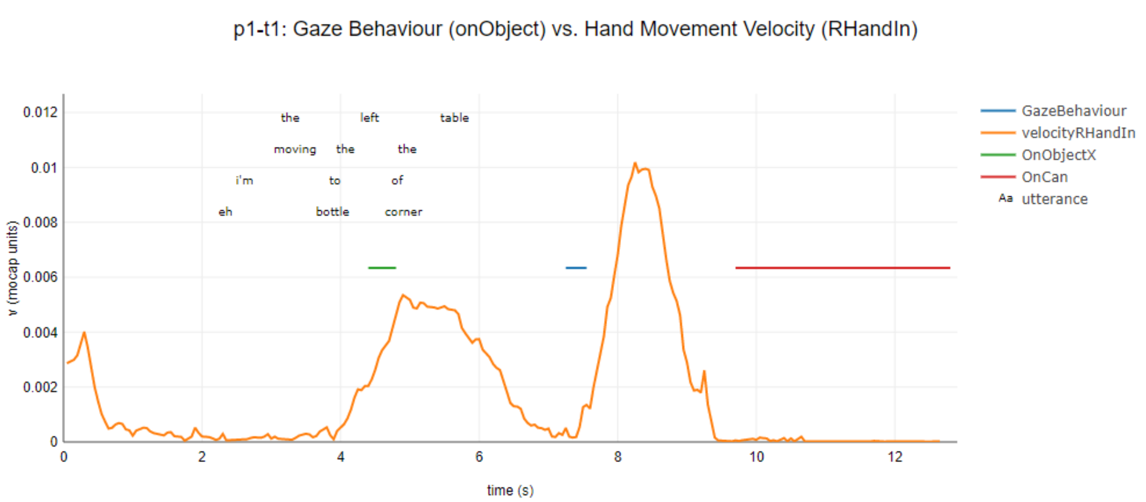

Figure 6.4: Adding additional objects to the visualisation might inform the observer about what happens around an actions, e.g. obstacle avoidance (possible explanation for "onObjectX" annotation) or preparation of future actions (possible explanation for "onCan" annotation).

Indeed, the visualisations turn out to be helpful tools for data-driven hypothesis generation. Another example, Fig. 6.5 includes the "HandTouchesObject" annotation, as a way of representing the successful completion of the reaching action. It may be possible to address haptic reattention to the target object (suggested by *Ordering constraint 4*) by a combination of hand movement and eye gaze velocity. Since the object has just been completed a period of low hand movement velocity would be expected before the participant begins the next action (see Fig. 6.5-A). Meanwhile, low eye movement velocity during the same time could indicate the absence of a visual attentional action during the same time period, as a result of an ongoing haptic attentional action. This raises questions about the relationship between different modalities of attention, and the way these interaction during reach-to-grasp actions.

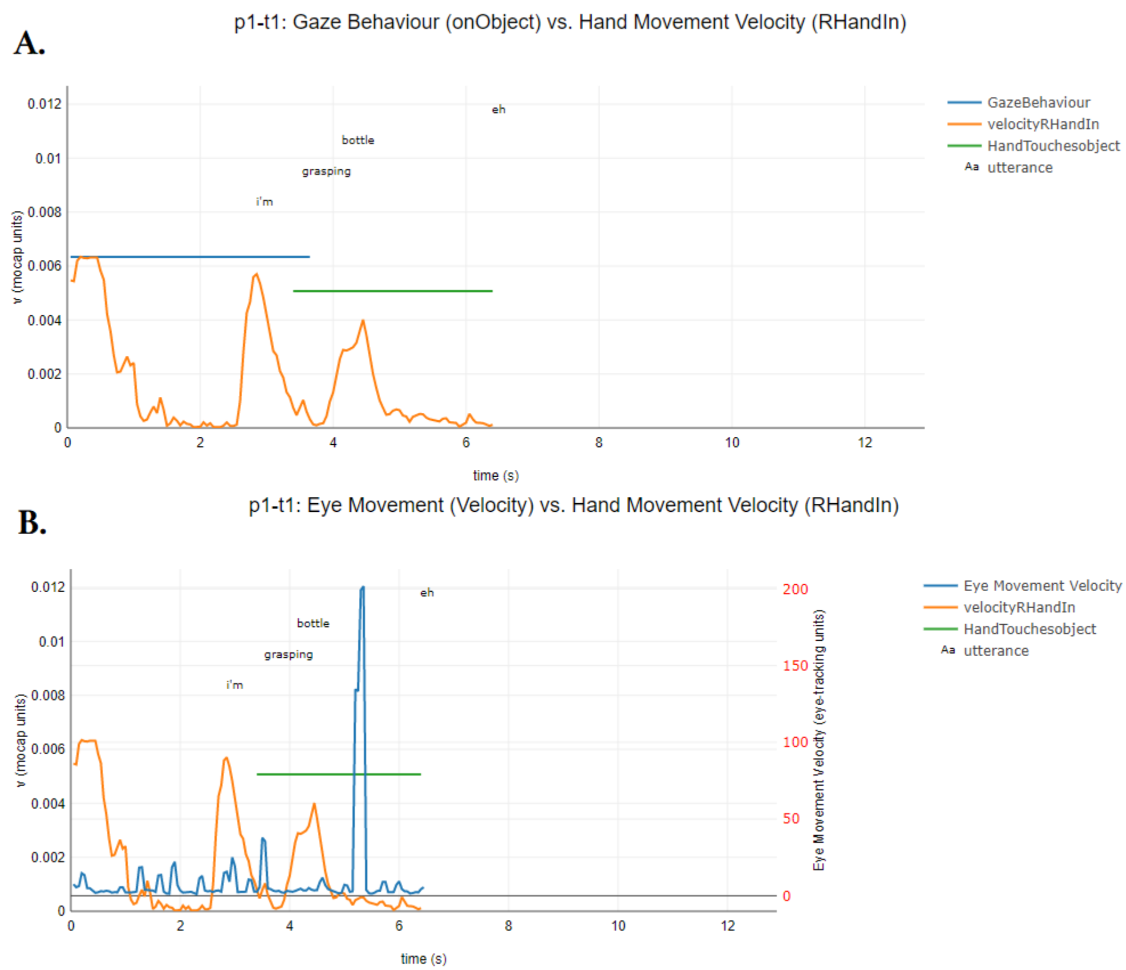

Figure 6.5: Adding information on when the participant was touching an object could raise a variety of questions. **A.** puts this in relation to gaze behaviour annotations, while **B.** does so in relation to eye movement velocity. While possibly informative, increasing the number of data modalities also makes the figures more difficult to interpret.

As a question of good practice however, this data driven approach should be approached with care. What the visualisations are useful for is representing the data in a more accessible form. On this basis, interesting or unexpected correspondences between data modalities might be discovered. However, such exploration of the data is no substitute for theory. [5] Indeed, at this point, the hypothetico-deductive model and traditional approaches to the scientific method would demand a theory and the formulation of specific, testable predictions. These would require additional testing, and the development of new research designs. Remaining agnostic with respect to the reader's position on the theory/data thing (Bowker, 2014), Marr

[5]For a contrary perspective, see C. Anderson (2008); Bowker (2014) meanwhile, offers an excellent discussion of the issue.

(1982) can be read to as emphasising the importance of specifying a computational, algorithmic, and implementational aspect of any cognitive process that is interpreted into the data based on this data.

Nonetheless, while it is worth, for example, warning about spurious correlations between data modalities on the one hand, on the other hand, there is also a place for exploring what participants were doing by inspecting the data visualisations, and then developing further research based on that. Rather than categorically rejecting one approach or another, the key points may choosing a suitable approach for the question one is interested in (recall (Agre, 1997b), awareness of the approach one is taking, and explicit acknowledgement of and reflection upon it strengths and weaknesses. [6]

Additional Data Modalities

On the data-driven side, preparations are currently underway to add a number of scenarios to the corpus. For example, future aims aims to investigate differences (and similarities) between action observation and action performance, which is not possible based on the LKG-corpus. In future, eye-tracking data will be collected from the active participant and an observer simultaneously. This assures that the recorded data, though from different perspectives, refer to the same situation, thereby removing difficulties that might have arisen when trying to compare data generated under different experimental conditions. Based on lessons learned with the LKG corpus, these studies will also include some number of trials in which the target object will not be chosen by the participant, but will be cued via a visual display visible only to the active participant. This may help identify processes related to target selection.

Future work will also add electroencephalography (EEG) as a data modality. This technique has been used to decode actions (Schwarz, Ofner, Pereira, Sburlea, & Mueller-Putz, 2017) and as the source for common neural features from which to estimate behaviour (Touryan, Lance, Kerick, Ries, & McDowell, 2016) – this idea is duly problematised in Sect. 7. Together with the revised design, bringing these various modalities together could offer a way investigate more aspects of a model as broad as Knott (2012)'s at once.

Taking neuroimaging data into account, it might be possible, for example, to disambiguate questions about the onset of action monitoring, or classification (*Ordering constraint 10*). However, this will require a larger number of trials per participant, and a more rigid data collection set-up. While this is limiting, it may also open ordering constraints to investigation that could not be addressed before. Consider, for example:

> **Ordering constraint 4** – after completion of a reach to grasp action, the agent typically reattends the target object (in a haptic modality)

Absent any measure corresponding to haptic attention, this aspect of the sensorimotor model could not be investigated based on the LKG-corpus. Haptic attention (see e.g. Hajnal et al., 2007; Lederman & Klatzky, 2009), too, could be investigated based on neurophysiological data. The same counts for aspects of motor attention (see e.g. M. F. Rushworth et al., 1997; M. Rushworth et al., 2003), for which the only indicators in the LKG-corpus are the properties of participants movements (e.g. direction, velocity). In particular, a neuroimaging study designed along the lines of the set-up reported here would contribute to a more detailed representation of the actions being investigated. If a suitable experimental protocol can be developed, the

[6]This may seem obvious enough, but given the hyperbolic tone on both sides (compare (C. Anderson, 2008) and (Pigliucci, 2009), for example), it seems necessary to point this out. Indeed, broader social and epistemological issues are being contested here, which call into question what is considered knowledge, or truth, why, and how these can be obtained. These issues are taken up again in Sect. 7 and the interested reader is referred to (Halpern, Mitchell, & Geoghegan, 2017) for an illuminating critique on the pervasive implications of data-driven decision making throughout society.

different data modalities could be used to constrain the interpretation of other modalities. In this way, such work would begin to conduct a broader test of the sensorimotor model.

6.2.2 Investigating Early Attention To Target

On the other hand, some aspects of (Knott, 2012)'s framework could be explored already based on the current LKG corpus. For example, consider *Ordering Constraint 2*, which was:

> **Ordering constraint 2** – during execution of reach-to-grasp action, agent typically attends to target in early stages of the action

Given this hypothesis, examining Fig. 6.3 raises a number of interesting points. For instance, it does appear that, on this trial, the participant looked at the target object prior to moving their hand – see the subplots on the figure's left-hand side represent the TAKE-action, since in this case the participant always described their actions in two parts, TAKE and PUT. Since the segments overlap, the increase in movement velocity that signals action onset in these left-hand subplots of Fig. 6.3 is the second set of spikes (i.e. the one on the right of each subplot). Examining the time prior to the action onsets identified in each subplot, we see that the onset of TAKE-action is, in fact, usually preceded by looking at the object. Breaks between actions likely account for the low velocity in the middle of most segments.

Meanwhile, this pattern is less obvious (see e.g. Fig. 9.7), or even mostly absent for other trials (see e.g. Fig. 9.4; for the complete action segment graphs, see Chapter. 9). This recalls Knott (2012)'s concerns about the persistence of a canonical sequence in the sensorimotor operations of adults. It is difficult to interpret what it means that there is evidence for parts of such a sequence from some participants, but not from others. The finding is compatible with a theory that there is such a structure, but that adults need not execute actions according to it, since they can simulate it instead. Or, the finding could be interpreted as evidence against the sensorimotor model. Else, Webb (2015)'s point also holds that the study design may simply have been chosen in such a way that is unable to provide definitive evidence. In this case though, questions arise about the utility of a study design chosen specific to elicit a canonical action sequence. Part of the reason this study used a less constrained design, was to avoid coercing participants to perform more regimented actions that they would perform outside the laboratory environment.

Or, the variety of action sequences exhibited by different participants can also be seen as a strength of this study. This is since unexpected patterns in participants action may raise important questions about the sensorimotor model. Since, one question is exactly whether adults continue to act according to the canonical sequence, and in which circumstance (if any) they do so. A different, but related question concerns what would be evidence against the sensorimotor model. Given the discussion of the absence of the canonical sequence in adults, a series of studies on children of various ages could be conducted, to see if the sequence in which they perform actions shows any less variety. Across different ages, this might offer clues about the existence of a canonical sequence earlier during development and about a possible time period during which the canonical sequence has less influence on action execution.

As a way of testing the model, Webb (2015) reports the number of trials in which the expected pattern is observed, calculating the probability that this result would occur by chance. In our case, participants looked at the object either prior to moving their hand, or latest as they started to move their hand, in the majority of TAKE-actions. This could be pursued further. Even so, it tests only one aspect of the sensorimotor model - basically one aspect of one operation and state. Whereas, further analysis of the LKG-Corpus might investigate the entire sequence of participants gaze behaviours, similar to the analysis reported in (Webb et al., 2010). However, this is more difficult in the present study, since the scene is the participants physical environment, rather than a video. As a result, three dimensional representations that take into account the changing orientation of each participant's head are required in order to capture the temporal structure of eye-movements, for example.

One approach might be for future analysis on the LKG-corpus to render animations of each trial in a way that includes a trace for the target of the eye-gaze, thereby representing the sequence of gaze behaviours. That might involve defining regions of interest (e.g. "future target location") and grouping fixations to these regions together to simplify the analysis – an first attempt to do so was abandoned due to the complexity of determining the role of different locations at specific points in time. Another approach to deal with the entire sequence would be to analyse the gaze behaviour annotations and attempt to extract a regular structure from them.

A key finding so far is that partial investigations of the sensorimotor model, in terms of the different ordering constraints for example, cannot provide definitive evidence for Knott (2012)'s model. This is also because individual predictions of the model are shared with other proposals (though they may be explained in other terms). Moreover, the existence of a canonical sensorimotor sequence is the foundation, but not the novel contribution of Knott (2012)'s proposal. To test the overall proposal, it is important to investigate this sequence, but that alone only lays the groundwork for a sensorimotor account of linguistic structure. In terms of the sensorimotor basis of linguistic structure, that is what this thesis contributes to.

6.2.3 Investigating Target Identification Prior to Action Onset

Another aspect of (Knott, 2012)'s framework and the sensorimotor basis of linguistic structure that can be explored based on the LKG-corpus is the following different ordering constraint:

> **Ordering constraint 10** – biologically-inspired computational models of action recognition assume that the agent's intended target is identified before the process of action monitoring/-classification begins

This adds a computational interpretation to an observed behaviour, specifically the early fixation (overt or covert) on the target of the upcoming action. That should be a biologically-plausible computational account of the entire action monitoring process (recall the discussion in Sect 3.1.2). At the same time, such an account should be both situated and phenomenologically comprehensible. Together, these demands suggests ways in which a given ordering constraint might be tested. In the case *Ordering Constraint 10*, one approach might be to attempt to infer when target objects were chosen in the LKG-corpus trials. Reaching toward an object is likely to be an accurate predictor, but this motor attention may be a relatively late indication of the participant's intention. Likewise, participants self-report and awareness of their intention to move may occur a considerable time after processes relating to action preparation have begun (consider, for example, the work on lateralised readiness potentials, e.g. Shibasaki & Hallett, 2006).

Meanwhile, fixational eye-movements may, but need not necessarily reflect attention to an object that has been chosen as the target. Visual inspection of videos reveals some participants move their gaze from object to object in between releasing one object and grabbing the next one (some participants cast their eyes about in this way even before completing the current action). Some, but not all participants then reach for the most recently fixated object, indicating different decision processes. The moving gaze from object to object could reflect target selection, or participants scanning the scene, or any other task demand (e.g. thinking of creative ways to describe one's behaviour after saying something similar for the previous n actions). Such behaviours pose a considerable challenge during analysis and call for a principled method, if it is to be avoided that the assumed cognitive process is simply imputed into the data. This may seem like a drawback of the current design, but only because it complicates analysis. However, the goal of this research is not a simple analysis, but a generalisable result. From this perspective, the richness of the recorded data is a strength, difficult to interpret though it might be.

For example, Webb (2015) also raises the issues of the strictness of the sensorimotor sequence, i.e. whether it allow for other attentional actions, related to distracting movements, for example, which might then account for differences in order of attentional actions (specifically, fixational eye movements) across trials. As before, to respond to the challenge certain aspects of participants' behaviour poses for the analysis, specific designs can be developed to complement the original, less constrained design. On the other hand, analysis based on the LKG-corpus could also begin by trying to identify behavioural markers for the onset of action monitoring. Evidence in favour of *Ordering constraint 2* (above) would be a first step. The second, as pointed out at the beginning of this section, calls for a computational explanation (in Marr (1982)'s sense) of the observed behaviour. For example, the LKG-Corpus data could be systematically related to the data on action preparation, discussed in Chap. 3.

Moreover, the LKG-corpus allows a first attempt to compare first-person perspective data to previous data of participants who observed and described the actions of others. Indeed, given the emphasis on the simulation of working memory episodes, and the inherent role of these simulations for action observation according to Knott (2012), it could be that the latter kind of study offers more clear-cut evidence for a canonical sequence. Else, action observation could be more strictly structured than action performance. Thus, the correspondence, and the difference, between these kinds of studies may be important to investigate.

6.2.4 Investigating the Relation Between Actions and Word Order

Departing from Knott (2012), one could investigate the relationship between sensorimotor events and surface characteristics of a sentence, i.e. a sequence of words (comparable to Chomskyan PF, see Sect. 2.2.4). To demonstrate this, consider word-learning. This has been accounted for in terms of sensorimotor experience, which is thought to create mappings between utterances and concepts (Tomasello, 2003).

The principle, in this account, is that sensorimotor experience of objects and their corresponding words is the basis of word-learning. The idea is that this generalises to learning any abstract concept as well. Basically, the approach stipulates that there are correlations between the words infants hear and their experience with the referents of those words. This is the basis of cross-situational word learning, as reported in Krenn, Trapp, Gross, and Neubarth (2017), for example (for seminal computational work, see e.g. Plunkett, Sinha, Møller, & Strandsby, 1992). While early word learning is thought to depend on upon this slow, associative learning, later learning aspect of word learning likely depends upon a variety of social skills, which improve the efficiency of infants cross-situational learning – for a computational approach developed together with "deictic codes"-first author Dana Ballard, see (Yu & Ballard, 2007). [7]

[7] Note, neither account of learning engages directly with the structure of language, though the regular structure of a language almost certainly facilitates word-learning (Suanda, Mugwanya, & Namy, 2014).

Even given the breadth of sensorimotor experience, similar mappings are more difficult to establish between utterances and linguistic structures. Indeed, it appears that some expectation of linguistic structure may be required – a popular approach is the idea of slot-filling Knott (2012, see Chapters 6 and 7 for details). This perspective ingrains aspects of nativist developmental linguistics into the relationship between sensorimotor cognition and linguistic structure. The difference between learning words and learning linguistic structure lies in role of sensorimotor experience in each case. On the one hand, sensorimotor experience appears to drive mappings between sensorimotor processes and phonological and later conceptual word representations. On the other hand, its contribution to linguistic structure may be more like that of a series of constraints. Recall, that the episode representation of "the woman grabbed the cup" is constituted by a replay of a limited number of processes with which to associate AGENT, PATIENT, and OBJECT. It is important to be aware of that investigations of the sensorimotor aspects of language can take these different forms.

Especially in conjunction with motion capture data, the line of research focusing on action-word mappings could be of particular significance for applications in action word learning (see Sect. 6.3.1). This would be work similar to that reported by Schreitter and Krenn (2016); Gross et al. (2018). Here, the focus would shift to surface patterns in language that might be useful for applications of computational linguistics, also in Human-Robot-Interaction (see Sect. 6.3.2). Besides further investigation of the developmental aspects of the sensorimotor model, the LKG-corpus can also be used to investigate action-word mappings (Gross et al., 2018), or the roles of attention and intention for word and concept learning (Tomasello, 2003).

In fact, this line of research raises questions about the multimodality of communication that are not irrelevant to work on the sensorimotor model. For example, one question that arises in Schreitter and Krenn (2016) concerns the role of multimodality during language acquisition (Yu, Smith, & Pereira, 2008; Suanda et al., 2016). These studies suggests that a variety of sensorimotor cues play a central role in word learning. That, in itself, would establish a fundamental role of sensorimotor experience for language learning. Knott (2012) might then be interpreted as taking the next step, by arguing that linguistic structure, too, depends upon sensorimotor experience. [8] Indeed, Knott (2012)'s sensorimotor model may outline a promising series of regularities that could be taken advantage of for learning action-word mappings, for example (as discussed in 6.3.1). In fact, the overlap between these lines of research and the usefulness of their results for the "other side" reiterates the potential of the sensorimotor model to overcome this staunch division into "opposing" camps (for a discussion, see e.g. Ritter, 2005).

6.3 Applications of a Sensorimotor Basis of Language

So, the LKG-corpus is relevant for applied research as well. Indeed, it was created in exchange with two research projects at the Austrian Research Institute for Artificial Intelligence (OFAI). Concerning the first, Sect. 6.3.1 discusses applications to language learning models. Interesting possibilities concern the creation of multimodal action-word mappings and crossmodal language structure learning (Krenn et al., 2019, 2020). The second project deals with human robot collaboration. In this regard, Sect. 6.3.2 engages with measuring and decoding movement intention (Sciutti, Mara, Tagliasco, & Sandini, 2018, is a broad discussion of the project). These applications are only briefly introduced, and one challenge is outlined for each.

[8] However, while they are part of the same broader characterisation of language and language learning in sensorimotor terms, fundamental ideas of the two lines of research should not be confounded. If they were to be assigned a side in the nativist-empiricist debate in contemporary linguistics, Knott (2012) would fall into the former category, and the other work cited in this section into the latter (e.g. Tomasello, 2003; Suanda et al., 2016; Krenn, Tsiourti, Neubarth, Gross, & Hirschmanner, 2019).

6.3.1 RALLI[9]: Sensorimotor-Based Language Learning

RALLI, the first application this thesis had in mind, aimed to address action and word learning in robots by observing the movements and utterances of human tutors. This requires information on the movements and utterances of participants performing a simple action, which the LKG-corpus was designed to contain. For, RALLI, the goal is a robot that has been built to be able to "acquire action concepts, learn and execute actions, and talk about them"[10]. The LKG-corpus provides data that helps to realise this idea.

Action-Word Mappings

RALLI is motivated by the insight that language learning is facilitated by multimodality. In particular, multimodality facilitates learning the surface structure of a particular language, i.e. the appropriate word order. Given their similarity to Gross et al. (2018), the data reported here can be considered a suitable basis for crossmodal and cross-situational word-object and word-action learning. However, while action-word mappings have been addressed productively (e.g. Roy, 2003; Yu & Ballard, 2007; Sadeghi & Scheutz, 2017; Krenn et al., 2020), there is less work on learning the (surface) structure of language. Whereas, structure learning actually looks to be a central part of automated language "understanding". This is because structure is a crucial cue to semantics, since it captures important groups of words, as well as their roles in relation to a certain action. For the phrase "the woman grabbed the cup", for instance, its structure reflects the presence of an action (verb / action), that there is someone who performs that action (subject / agent), and in this case there is the some object on which the action is being performed (object / patient). That information, which is present in the LKG-corpus, is encoded in linguistic structure, which makes this a promising target for future research.

Major Challenge: Embodiment Level

One approach, based on Knott (2012)'s sensorimotor model, would be to move another step closer to human language learning and draw on the potential of the robot's own actions to ground language learning. This recalls the issues of an embodiment level (D. H. Ballard et al., 1997, also see Sect. 3.1.2). The idea of this embodiment is that the current state of the body – for example where the woman is looking while she grabs the cup – is essential for information processing. The approach suggested here speculates about the usefulness for language-learning of the "bodily" processes and states of the robot, such as joint-angle information. This information is already used to control the robot, but could, conceivably, be made available to a language learning module as well. On this basis, a closer parallel could be established between human and robotic learning behaviour. [11]

One aspect to consider is that the embodiment level depends on a variety of attentional processes that would need to be, to some extent at least, implemented in the robot. For a better analogy, robot vision would have to be augmented to include some attentional focus (think of face recognition algorithms in cameras, for example) that orients the robot and adjusts other parameters. In humans, this role of attention likely developed as a result of the need to process information efficiently, and based on the minimal possible representation (Ullman, 1984; Newell, 1994; D. H. Ballard et al., 1997). Meanwhile, it is not clear that similar

[9] Vienna Science and Technology Fund (WWTF) – Robotic Action Language Learning through Interaction (ICT15-045)
[10] https://ralli.ofai.at/home.html [accessed 16.07.2020]
[11] Though, a closer parallel is not necessarily advantageous (for a discussion, see e.g. Beer, Quinn, Chiel, & Ritzmann, 1997; Beer, Chiel, Quinn, & Ritzmann, 1998) or desirable (Coeckelbergh, 2011).

constraints have to be placed on robotic visual information processing, nor that this would necessarily be useful (though see e.g. Marr, 1982; Franceschini, Pichon, & Blanes, 1992; Siagian & Itti, 2007; Franceschini, 2014). Also, whatever the capacity of human working memory robotic memory does not have to be limited in the same way (for a condensed chronology of the discussion, see G. A. Miller, 1956; Shiffrin & Nosofsky, 1994; A. Baddeley, 2003; Ma, Husain, & Bays, 2014). Nonetheless, it appears that multimodal, embodied remembering (see also Sect. 2.2.1) is not just a creative response to a capacity limitation, but actually an efficient means of representation in its own right (for examples of "embodied" memory in robots, see e.g. Dodd & Gutierrez, 0205; Cangelosi & Riga, 2006; Hülse, McBride, & Lee, 2009). That is promising for an embodied approach to learning surface language structures.

On the other hand, the embodiment of sensorimotor processes (real or simulated Ziemke, 2003a) will be specific, in parts, to the body of a particular robot (model). This raises the concern that any attempt to engineer an embodiment level will be difficult to generalise. Meaning, that this embodiment will be specific to the robot model in question. In fact, it may be specific to a particular robot (plausibly best performance will be achieved when the system is calibrated for each individual robot, which is the case in humans, too, after all). Even, since many robots used in research are (partially) modular, this embodiment level will be specific to the particular configuration of components that make up the robot at a particular point in time (i.e. which camera is being used, which gripper, etc.). While work in this direction might begin to implement a variety of constructionist insights (see e.g. Ziemke, 2005), it also appears to counteract the aim of building a widely applicable, general purpose learning architecture. Instead of engineering an embodiment level, future work might look at different implementation of the algorithm and representation outline in D. H. Ballard et al. (1997).

So, these ideas are still in their infancy. In relation to an embodiment level, at least, the promise of a sensorimotor basis of linguistic structure for applications for word and surface structure learning in robots remains uncertain. Open questions include how accurately human actions and cognitive processes can (and should) be simulated in robots, and how useful it will be to reproduce certain information processes.

6.3.2 CoBot Studio[12]: Human-Robot Interaction

CoBot Studio has a different focus than language-learning, namely the coordination of behaviour between human and robot. In fact, successfully completing the joint task is only one aspect of this coordination, which has also been shown to be central to the user's experience with, and acceptance of the technology (Sciutti et al., 2018). Indeed, not only with respect to task completion is it essential that all participants have some understanding of intentions of their collaborator and the current status of their actions. The LKG-corpus is relevant here, inasmuch as it offers a data set based on which behaviours might be identified that signal human intentions.

Interpreting Actions of Humans and Robots

Basically, the LKG-corpus data could be used as sensory input with which to train a robot to decode human movement intentions, at least in the limited context of the experimental set-up reported here. In this regard, the application is a first step towards coordinating human robot behaviours. In a simple case, consider a robot picking up an object and placing it somewhere else. If this action is intended (by the designers) to take

[12] Austrian Research Promotion Agency (FFG), Ideen Lab 4.0 project CoBot Studio (872590)

place in an environment that the robot shares with humans, this always raises safety concerns (Herrmann & Melhuish, 2010; Hayes & Scassellati, 2013). One way to address these is for the robot to infer human intentions and to avoid the space the human is expected to move through (Bascetta et al., 2011). Various approaches have been taken to model human behaviour, usually in routinised, industrial settings, with the idea of developing representations of human intentions that the robot can use to contribute to the safety of its human collaborators (see e.g. Liu & Wang, 2017; Bi, Guan, et al., 2019). Such work attempts to make humans easier to interpret for robots. The same approach is also used with a focus other than safety, namely the collaborative aspects of human-robot interaction (for reviews, see e.g. Bauer, Wollherr, & Buss, 2008; Green, Billinghurst, Chen, & Chase, 2008). In this context, human motion intentions are likewise calculated (see e.g. Li & Ge, 2013), except with the goal of developing proactive (rather than reactive), and intuitive robot collaborators (see e.g. Awais & Henrich, 2012).

The opposite perspective can also be taken, namely trying to make robots easier to interpret for humans. Much work on this subject focuses on robots' verbal or visual explanations of their actions (see e.g. Perera & Veloso, 2018) – for a theoretically-founded approach see e.g. Sridharan and Meadows (2019). Though, a recent review by Anjomshoae, Najjar, Calvaresi, and Främling (2019) has identified a number of pervasive limitations that characterise the current state of this research. Indeed, Edmonds et al. (2019) – whose paper also relevantly combines an embodied robotic approach with grammar learning – diagnose a frequent tension between robot performance and interpretability.

In fact, the LKG-corpus can likely be used to address this challenge. For, taking inspiration from human motion is not only likely to make robot actions easier to interpret, but may also improve performance, if suitable controllers can be developed for the purpose (Ellis, 2016). [13] Specifically, the data reported in this thesis might be used to allow robots to learn hand-shapes and hand-orientations based on human movement. The idea is, then, to make robot behaviours more accessible to humans by making them intuitively recognisable, based on their similarity to human behaviours. After all, it appears that humans have a special capacity to detect biological motion (see e.g. Troje & Westhoff, 2006) – (for a review of the corresponding neuroanatomy, see e.g. Vaina, Solomon, Chowdhury, Sinha, & Belliveau, 2001). Indeed, G. Hoffman and Ju (2014) explicitly calls upon robot designers to bear expressive movement in mind with respect to interpretability, and to take inspiration from human movements.

One feature of the LKG-corpus data that may make it particularly interesting is that participants were less constrained in their actions than is often the case in other corpora (e.g. Gross et al., 2018). The reason this could be useful, is that tasks that change, or which are unstructured will be the major challenge for human-robot collaboration (for reviews, see e.g. Hayes & Scassellati, 2013; Ajoudani et al., 2018). In such cases, there may not be a routinised and serialised set of actions to accomplish a goal. Instead a dynamic environment may require continuous adaptation, in context of which it becomes more difficult, and potentially more important for the human and robot collaborator to be able to interact seamlessly (for one system developed with this challenge in mind, see Meziane, Otis, & Ezzaidi, 2017).

Major Challenge: Actions, Plans and Intentions

However, the issue of unstructured tasks also recalls the problematic relation between actions and plans (L. A. Suchman, 2007). This is especially clear for unstructured tasks, but relevant to all situations.

[13] In this regard, robot design can also take advantage of other aspects of anthropomorphism (see e.g. Beuter, Spexard, Lutkebohle, Peltason, & Kummert, 2008), which can increase the interpretability of robots (Epley, Waytz, & Cacioppo, 2007).

The fundamental challenge is that how an action is interpreted by a human always depends on the specific situation. Concerning human-robot collaboration, this is closely tied to problems for both robotic calculation of human intentions and the interpretability of robot actions.

For example, without situational knowledge, it is difficult to assess why the robot grabbed a cup, or what the woman is intending to do, when she first looked at the cup – this is related to how it is hard to follow this sentence, without more information. For L. A. Suchman (1987), dealing with this situatedness of knowledge requires a perspective in which the contingencies of an action, for example on artifacts and other actors, is understood as an essential resource. Essential, in the sense that it makes knowledge possible, and as a precondition of the meaning of an action. That means a departure from a traditional "engineering"-mindset to which the ambiguity of action is a problem to be solved. Instead, ambiguity is not only a given, but also a basis of responsive and flexible behaviour.

Drawing on approaches in ethnomethodology (see e.g. Garfinkel, 1967; Heritage, 2013), the insight is that these issues cannot be solved in principle. Rather than implying a reduction of the usefulness of the LKG-corpus (because it is not specific to a certain task, for example), this actually suggests another set of opportunities. Namely, future work can explore the patterns and correspondences in the data, which can be used to constrain and guide future investigations.

In fact, exploring data in this way presents a valuable opportunity. For example, Feitelson and Stefik (1977, in relation to genetics) analysed the ways in which the scientists they observed let their future experimentation be driven by successive insights as they arose – compare this with methodological challenges discussed in Sec.6.2.1.The case study, and ethnomethodological approaches in general, offer a perspective on this explorative approach as a legitimate means of knowledge creation (for another discussion, see e.g. Latour & Woolgar, 2013). Moreover, L. A. Suchman (1987) makes the point that this practice can be seen as a way of ensuring that theory-generation stays in touch with the empirical data on which it is based, rather than becoming the object of a perpetual abstraction away from the situated action.

Indeed, L. A. Suchman (1987) continues, this proximity to empirical data from the situation in question is a crucial aspect of orienting design away from stipulated conditions. Since, by avoiding a continuous abstraction from the behaviours of agent *in situ*, the designers of robots and their actions may be able to move beyond a method of anticipating and constraining the user's actions. Like attempting to solve these issues in principle, straightforward simulation of human communication on the robot's side will likely also be insufficient. Instead, designers will likely need to recognise the asymmetry between human and robot. In that sense, the LKG-corpus remains inadequate, since only future work will address collaborative scenarios (early efforts in the same direction include Schreitter & Krenn, 2016).

From the perspective of the CoBot project, the principle objective is developing an well-founded understanding of the dynamics of human robot collaboration. Only on this basis, can concrete applications be developed. However, this understanding is precisely not supposed to be tied to specific tasks, but reflect an insight into the general dynamics of that collaboration. That is why the targets, at this stage of research, are the communicative signals that are readily understandable to different groups of users. In regard to this, the relation of actions, intentions and plans is a major challenge not only because it is a difficult design/engineering problem, but also because it is fundamental. Since the LKG-corpus does permits an investigation of communicative signals, we can think of it as a contribution to human robot collaboration. This will fully be the case once a collaborative scenario has been implemented in the near future.

Chapter 7

Crossdisciplinary Concerns

A variety of concerns surround applications of the LKG-corpus. While many may appear as practical engineering challenges, all are embedded in a set of practices and presuppositions (Latour & Woolgar, 2013). Indeed, some of these stem from fundamental assumptions and questions at work behind this thesis, which have not been addressed in their own right so far. Now, this chapter adds a reflection on some of these underlying sociopolitical and philosophical concerns. Crucially, it is not an afterthought, but an essential aspect of this thesis. So, the concerns discussed below have been referred to throughout the previous sections, and the discussions in these were conducted with them in mind. While those sections raised a variety of questions (e.g. the gendering of nature in Sect. 2.1.2), only two will be discussed here.

First, Sect. 7.1 will engage with some epistemological concerns that rarely receive the attention they deserve in natural scientific publications they are nonetheless relevant to. In it, I raise questions about how to carry out a responsible and sensitive scientific practice, especially with respect to the situatedness of scientific knowledge production (a related, but different use of the word than in "situated cognition", see Sect. 2.2.2). Significantly, this section also calls to awareness other ways in which the body can be seen as a basis for knowledge and cognition, beyond the scope of linguistic structure and language.

Then, Sect. 7.2 keeps focus on the body, but specifically on the relationship between bodies and data. For, the LKG-corpus exemplifies the reduction of bodies to data that typifies the quantification of human bodies. This practice brings with it a number of challenges that I want to point out, if only to sensitise the reader to the issues they raise. In fact, I have already begun pointing to these issues in the repeated discussion of the problematic relation between plans and actions (recall e.g. Sect. 6.3.2). These discussion will be developed further around the paradox that human bodies tend to disappear as they are turned into data (see e.g. Viseu & Suchman, 2010). For, it appears that human subjectivity is obscured wherever human actions are broken down into measurable features (see e.g. Schüll, 2014) and then abstractly represented (for an excellent discussion in relation to neuroimaging, see Dumit, 2003, 2004).

These two sections address some among a host of tensions in which the present work is embedded. These two issues in particular are addressed as especially salient concerns. Meanwhile, I attempt to address them at all because this is intended as a interdisciplinary, but also transdisciplinary thesis. For, where interdisciplinarity crosses boundaries, transdisciplinarity recognises their artificiality (Klein, 2010). Disciplinary boundaries are constructed, after all, however "organically" they appear to have grown over time (for a discussion, see e.g. Gibbons et al., 1994). [1] In this respect, this section stands for a critical practice, and a coming together of question usually separated into the domains of compartmentalised experts, divided by disciplinary boundaries. Inspired by this, the next section goes directly to a fundamental concern by foregrounding some questions about epistemology and objectivity in science. Then, I discuss the relation between bodies and data, both to problematise the reduction of human subjectivity and to highlight the relevance of this discussion to the question of what constitutes a legitimate source of knowledge.

[1] On the one hand this is a definitional word game. On the other hand, claiming to cross disciplinary boundaries does seem to reinscribe them, and therefore accepts them as given, there to be crossed.

7.1 Objectivity and Scientific Knowledge

Cognition is not all that is situated. Scientific knowledge production is, too. To comment on this, I want to build on feminist epistemological work like D. Haraway (1988) and insist we bear in mind that the knowledge claims made herein are not impartial. They depend on time, place and person (Collins, 1989), as much as on cultural and institutional norms (Kuhn, 1962), as well as practices and tools of knowledge production (Clarke & Fujimura, 2014). Moreover, scientific knowledge itself is rooted in the body (D. Haraway, 1988). It is embedded, and it is not only in the case of false or superseded knowledge that assumptions, equipment and standards of practice contribute to the claims that are produced (C. M. Thompson, 2015). Meaning data do not speak for themselves, and also do not speak only of what they measure. For, in data, the object of the measurement is always inseparable from how it was measured, when, by whom, with which assumptions, and which apparatus. Without a reason to disregard them, each of these aspects of the context of discovery must be borne in mind.

Ideally, this section would do justice to(Agre, 1997b)'s critical practice and reflect on assumptions, ideologies and values underlying the investigation being reported. Unfortunately, one reason for this introduction is to admit omitting such a discussion. Instead I only acknowledge, rather than discuss, the transitory existence of scientific knowledge claims where Kuhn (1962) and others do not (cf. Cohen, 2015). [2] Another point I acknowledge, but will not discuss concerns the parallel knowledges of marginalised groups that are excluded from the typical euro- and male-centric account of the development of science (de Jong et al., 2018). For example, concerning human robot interaction, one of the applications of the LKG-corpus, Pillinger (2019) outlines a number of biases that continue to influence design decisions. These issues are deeply relevant aspects of the situatedness of knowledge production, but outside the present scope.

Instead, where this section picks up the point about the situatedness of knowledge production that this chapter opened with, is in relation to computationalism. Indeed, the computational theory of mind is inseparable from a host of anthropological, computational, neuroscientific, philosophical, psychological, sociological challenges, as well as countless others – which is part of why multidisciplinary research is required (cf. Van den Besselaar & Heimeriks, 2001). One of these challenges reflects a concern raised by fundamental, critical inquiry into scientific knowledge production and scientific epistemology, namely what is a legitimate source of knowledge.

For example, the time it has taken to recognise the role of the body in "higher" cognitive processes (like language) reflects a bias towards disembodied forms of knowledge, or means of creating it (for a discussion, see e.g. Boehner, DePaula, Dourish, & Sengers, 2007; Sengers, 2018). This view is inherited from Enlightenment's championing of the (male, pure, strong) intellect versus the (female, corrupted, weak) body (Outram, 2005). And it can be neatly traced into computationalism's focus on abstract information processing, or into cognitive science generally, with its early, male-dominated focus on chess and theorem-proving.

One place we arrive is back at the issue of embodiment, and the apparent tension with computationalism. We also arrive back at the issue of feminist epistemology, especially thanks to Adam (2006)'s inimical discussion of how gendered ideas at work behind the computational theory of mind are inscribed into computational artefacts and explanations. "Inscribed" means the artefact/explanation carries the computationalist's un-

[2] This refers to the insight that the concept of a 'paradigm' (Kuhn (1962)) is inadequate, both for describing scientific practice, and with respect to its acknowledgement of sociocultural factors that shape the production and narratives of scientific knowledge.

derstanding of the world, for example in the way it defines who can act, and how (Akrich, 1992). This should illustrate again how fundamentally interwoven these issues are.

Again though, too, this cannot be the focus here. Instead, the point I want to take from this for investigating the sensorimotor basis of language and its applications concerns what to learn from these insights about how to conduct "good" science. Now, the concerns listed above indicate that feminist epistemology in particular de-emphasises a variety of forms of objectivity (historically associated with "the mind"), in order to emphasise subjectivity (associated with emotions and the body) instead. [3] So, the re-orientation being conducted in this thesis, one towards the body, can be said to reflect, in a small way, a similar move to the one called for by feminist epistemology. Since, both perspectives call upon a more expansive definition of knowledge, and one that recognises the primacy of the lived body in relation to the intellect. At the same time though, this work holds on to a number of conceptions that would be regarded as problematic from a feminist epistemological perspective (for a discussion, see e.g. E. Anderson, 2020). For example, one conceptions of objectivity criticised in feminist epistemology claims that knowledge of an object is gained by learning to control it, or quantify it (for a discussion of this, see e.g. Schüll, 2014). While the choice of a less constrained task is also a nod that acknowledges that this is problematic, the repeated calls for further and more specific study also perpetuate exactly such an attitude. [4]

Herein lies a major challenge of this thesis, namely reconciling the habits and practices of natural scientific knowledge production with the problematising insights of a critical perspective. Importantly, this is not achieved by changing the content (what research is about), but the process (how research is conducted) (Longino, 1987). That gives this thesis a chance to offer something along the lines of a feminist perspective. To do so, it is empirical adequacy that counts (L. H. Nelson, 1996). In fact, that means one central task is asking whether the frameworks that this work relies upon are likely to produce the best theories (E. Anderson, 2020). Whereas, the distinction between better and worse science, still depends on epistemic virtue (E. Anderson, 2020), a "good" theory also incorporates socio-political concerns, like problematising the gendering of nature, or the construction of sexless bodies (more on that soon) (L. Nelson, 1992). That is what this chapter attempts.

So, on the one hand, the sensorimotor basis of linguistic structure may not, at first, appear to be a socio-politically thick issue – "thick" meaning laden with values (E. Anderson, 2020). However, that appearance is deceiving. Adam (2006) lucidly discusses gendered conceptions of intelligence and knowing, for example, that are at work behind the computationalist perspective adopted in this thesis, too. Whereas, E. Wilson (2016) discusses tensions between neuroscientific perspectives on cognition and feminist notions of subjectivity. And Cameron (1992) has summarised feminist issues in linguistics. Too, Bluhm, Maibom, and Jacobson (2012) have edited a volume of feminist issues in (neuro-) cognitive science, giving voice to transboundary work, like Pitts-Taylor (2015)'s discussion at the intersection of embodiment, enactvism, sociology, and ethnography. Point being, the frameworks that make up this thesis have all received critical attention elsewhere. Moreover, rather than being attempts to defeat it, such attention often aimed to improve future research. This thesis does the same, beginning with a sensitivity to issues of feminist epistemology. Building

[3]Note that every effort is made, here, to avoid reinscribing a distinction between "mind" and "body". The two are exactly not separate in lived experience, and come apart only with difficulty (if at all) in analysis, too.

[4]Note that questioning this notion of objectivity does not constitute a rejection of science, but an attempt to improve it (E. Anderson, 2020)In fact, several works within feminist epistemology reflect an effort to promote criticism towards a more inclusive scientific practice (see e.g. D. Haraway, 1988). For, the feminist criticism is that much research is oriented towards maintaining material control and inadvertently or others entrenches social hierarchies (Harding, 2006). Attention to values is not supposed to replace evidence, but create a basis on which to assess it (E. Anderson, 2020).

on this, the next section recognises and engages with one particular issue that arose in relation to this thesis, namely the relationship between bodies and data.

7.2 Bodies and Data

Indeed, this section picks up an issue that the alert reader may have noticed already. Namely, throughout the discussion of the LKG-corpus data especially, there is a tendency to reduce the actions of human participants to quantifiable measures. Hand movement velocity is one example, or the coordinates towards which the participant's eyes are directed. Here, actions are quantified. And not just actions, but bodies, and persons. However, the argument that follows is not normative, not about the "right" or "wrong" of this issue. Instead, I want to establish an example for the point that studying the basis of linguistic structure does not just mean studying sensorimotor processes, it also means studying socio-political processes.

One aspect of this is that the questions is not so much what is being studied, as how (cf. Longino, 1987). This line of thought was already introduced above. There, it was raised in relation to how scientific knowledge is constructed. Now, this section looks ahead at what happens as a result of certain constructions, and at how this knowledge can be applied.

To do so, consider the first application discussed in Sect. 6.3: language-structure learning. There, I suggested that the LKG-corpus data could be used as the basis for better language comprehension (and production) algorithms. These, when they have been written, will be implementable in robots. Robots that will be able to deal with human language better than before, which is an ability that is, in the current *Zeitgeist*, may lead to increased automation (Rifkin, 1995; Strangleman, 2007). One example is the customer service sector (for early work, see e.g. Cox, Kamm, Rabiner, Schroeter, & Wilpon, 2000). So researching the sensorimotor basis of linguistic structure can seem innocuous, but it may not be far from such research to actions with far-reaching implications.

That establishes possible implications of this work. It also reflects a subtle equivocation inherent in this thesis. Namely, it blurs the boundary between humans and machines. This can be traced to the reconceptualisation of human thought and behaviour in terms of information (for a discussion, see Orr, 2006). In this one conceptual manoeuvre, the constructed boundary between human and machine is called into question. For, information is an abstract quantity and the cybernetic principles of its manipulation do not know any difference between human and machine (cf. Wiener, 2019). The equivocation, then, suggests human and machine are interchangable. And in doing so, it paves the way concerning major social challenges like increased automation (irrespective of the warnings of one of the founders of cybernetics, see Wiener, 1949).

Moreover, the tendency equivocate humans and machines also lies in the quantifying approach taken here. Undoubtedly, the approach of this thesis is indebted to cybernetics (Boden, 2008). And, the objective of cybernetics is to control flows and information, and in doing so to control behaviour (cf. Wiener, 2019). This perspective, and the values with which it is associated, necessarily leaks through into the present work (compare this statement to the feminist epistemological positions discussed in E. Anderson, 2020). Not just from cybernetics itself, but from the computationalism it inspired. In this case, that blurs the boundary between human and machine. This is not problematic in itself, but always already suggests that the two are interchangable (Sengers, 2018). Indeed, part of the difficulty with coming to terms with the question of what (not) to automate, and why, is the consequence of that equivocation between humans and machines (see also Boehner et al., 2007). And this, fundamentally, depends on the quantification of the human, much like it is practised here.

In fact, the same issues arises concerning measuring movement intention. In the design of industrial robots discussed in Sect. 6.3.2, this is very much a control problem. Human intentions are understood to exist in measurable form, there to be identified and decoded by the robot (Boehner, DePaula, Dourish, & Sengers, 2005). This is a paradigmatic example of the usefulness of cybernetic, which turns everything there is into code (D. J. Haraway, 1989; Orr, 2006). However, this raises similar concerns as those outline above.

For example, part of what is being controlled is the human user. Indeed, scientific management has moved off the factory floor and become a totalising influence on workplace planning (Sengers, 2018, problematises it in both places). For, only bodies that can be measured can be controlled (Schüll, 2014). On the one hand, that means this issue arises directly out of the quantification of worker's bodies (Moore & Robinson, 2016). On other hand, this imparts a responsibility to at least pose some relevant questions. These include: whom does that control ultimately serve? And to what end? Indeed, where the LKG-corpus could facilitate human-robot collaboration, there has not been sufficient attention to the costs involved, nor to who is really likely to benefit. That does not need to be discussed here, but it needs to be acknowledged. And so, too, does one root of this issue in a particular relation between bodies and data.

That leaves open the question of what to do about the issues raised above, apart from acknowledging and reflecting on them. Nor is that the object of this thesis. Indeed, the omission is intentional. The responsibility of this thesis has to end with raising these issues. For, in the end, the topic of this thesis is the sensorimotor basis of linguistic structure. So, while this work is an example of the conversion of bodies into data, and while it acknowledges this issue, it was not designed as a basis to discuss the issue. Instead, an effort was made to avoid the pitfalls of a quantifying gaze as much as possible. Plus, the challenges this issues poses for two applications of the current effort have been outlined. So, no future work should be able to draw on the LKG-corpus without considering these latter questions as well, for example:

1. what assumptions are at work behind the framework used in our research?
2. how do these assumptions influence and shape the entire research process?
3. how can responsible research be conducted, that bears these issues in mind?

The last question is particularly relevant considering some research cannot be conducted without a reduction to data in certain ways. The applications discussed in Sect. 6.3 rely on a reduction of human bodies to data, at least to make them interpretable to the artificial systems, including robots, involved. From the perspective of these systems, this is necessary in order to interact with humans. Concerning human intentions, for example, L. A. Suchman (2007) discusses one promising approach to deal with this. Namely, the idea of Wei (2002)'s work is remain agnostic about the intentionality of movements, by focusing on tracking and responding to users to engage them in an embodied dialogue, as opposed to insisting upon recognition and the inference of internal states. This facilitates a less limited human-machine interaction that recognises the agency of humans and machine, and which can engage with the kind of culturally and materially situated being that humans are. One next step for LKG-corpus is to build these concerns into the design of future work as a way of implementing a critical practice.

Chapter 8

Summary: Contributions of This Thesis

As it stands, this thesis makes a variety of contributions. In light of this, it concludes with an inventory of what these are. Table 8.1 contains an overview that notes the novel contributions, as well as aspects of the thesis that are more indebted to previous work. Briefly, the novel contributions of the thesis span the breadth of a new corpus (Reinboth et al., 2020), a software tool, a discussion of its conceptual foundations, suggestions for future work, and consideration of socio-political issues.

In summary, this thesis began with an introduction that focused on the work reported in Knott (2012) (see Sect. 1.1) and then on explaining how I have developed this approach further (see Sect. 1.2). Next, Chapter 2 established a foundation for the remainder of the thesis. To do so, it clarified a number of basic terms and concepts (see Sect. 2.1). It also discussed the key theoretical framework on which this thesis is built (see Sect. 2.2). At this point, I introduced and defended central ideas from a 'weak' embodied perspective (see Sect. 2.2.1), and outlined a compatible interpretation of computationalism (see Sect. 2.2.3).

Having addressed these fundamental questions, the following chapter was dedicated to a detailed introduction to Knott (2012)'s sensorimotor model. Sections 3.1 outlined some of the basic ideas on which the model is based, including the central role of attentional actions. Building on this introduction, I then presented the sequence of processes that constitutes an exemplary sensorimotor sequence, namely grabbing a cup (see Sect. 3.2). This detailed the sensorimotor processes involved. To make the bridge to language, Sect. 3.3 introduced the working memory mechanisms that take the role of a shared mechanism between sensorimotor cognition and linguistic structure. Then Sect. 3.4 briefly evaluated Knott (2012)'s proposal.

With the theoretical foundation laid, Chapter 4 summarised the participants involved and explained the data collection procedure. The chapter also pointed out this effort's correspondences with and differences to Knott (2012) (see Sect. 4.4). Moreover, the chapter included a reflection on the lessons learned from data collection and outlined improvements (see Sect. 4.5). A similar reflection concludes Chapter 5 as well (see Sect. 5.4). That chapter also includes an explanation of the how data is represented for the LKG corpus (see Sect. 5.1) and an overview of the data modalities recorded for this thesis (see Sect. 5.3), including the introduction of a novel annotation scheme (see Sect. 5.3.2.2).

Arrived at the results, I then introduced a software tool to visualise the LKG corpus data (see Sect. 6.1). This was designed to facilitate the further development of a model of the sensorimotor basis of linguistic structure. Based on visualisation generated using this tool, Sect. 6.2 suggested how future studies could continue to investigate specific predictions based on a sensorimotor model of linguistic structure inspired by Knott (2012). In addition to these comments on basic research, Sect. 6.3 outlined two possible applications of this work and mentioned a central challenge for each.

Finally, concerning challenges, Chapter 7 considered a relevant sociopolitical issue. Namely, I took a feminist epistemological position on scientific knowledge creation in order to problematise abstraction and quantification (see Sect. 7.1). Specifically, I discussed the relation of bodies and data, as an example of the importance of being sensitive to assumptions that are at work in the background of this thesis (see Sect. 7.2). This rounded out a transdisciplinary perspective of the sensorimotor basis of linguistic structure.

Table 8.1: Summary of the contributions of this thesis in contrast to prior work on the topics of this thesis. The areas/topics roughly reflect the chapters of the thesis. Much other prior work is summarised or referred to in Knott (2012), which is why it is not referred to explicitly here.

Area / Topic	**Prior Work**	**Novel contribution**
language resource	other corpora, e.g. (Gross et al., 2018)	LKG-corpus (Reinboth et al., 2020) set-up design, data collection (see Chapters 4) development of annotation tiers, manual annotation (see Chapter 5)
conceptual backdrop	discussion of minimalism and shared mechanisms hypothesis in Knott (2012)	conceptual clarification, including discussion of embodied cognition, defence of computationalism (see especially Chapter 2)
sensorimotor basis of language	sensorimotor model (Knott, 2012)	relation to linguistic structure (as opposed to syntax; see Chapter 3 in particular) suggestions for investigating relation of sensorimotor processes to surface structure, i.e. word ordering (see Sect. 6.2.4)
empirical basis for future work	suggestions in e.g. Knott (2012)	lessons learned from data collection and processing (see Sect. 4.5 and Sect. 5.4)
	prior empirical work, e.g. conclusions drawn by Webb et al. (2010); Webb (2015)	visualisations and software tool applicable to future data (see Sect 6.1)
broader epistemological / social concerns	–	discussion of sociopolitical issues relevant to quantification of human bodies (see Chapter 7) references throughout thesis to relevant issues, including feminist perspectives and questions concerning philosophy of science

References

Adam, A. (2006). *Artificial knowing: Gender and the thinking machine.* New York, USA (NY): Routledge.

Agre, P. E. (1997a). *Computation and human experience.* Cambridge, UK: Cambridge University Press.

Agre, P. E. (1997b). Toward a critical technical practice: Lessons learned in trying to reform ai. In G. Bowker, L. Gasser, L. Star, & B. Turner (Eds.), *Bridging the great divide: Social science, technical systems, and cooperative work.* Mahwah, USA (NJ): Lawrence Erlbaum Associates.

Agre, P. E., & Chapman, D. (1987). Pengi: An implementation of a theory of activity. In *Proceedings of the association for the advancement of artificial intelligence AAAI* (pp. 286–272).

Aizawa, K. (2007). Understanding the embodiment of perception. *The Journal of Philosophy*, *104*(1), 5–25.

Ajoudani, A., Zanchettin, A. M., Ivaldi, S., Albu-Schäffer, A., Kosuge, K., & Khatib, O. (2018). Progress and prospects of the human–robot collaboration. *Autonomous Robots*, *42*(5), 957–975.

Akrich, M. (1992). The de-scription of technical objects. In W. Bijker & J. Law (Eds.), *Shaping technology / building society.* Cambridge, USA (MA): MIT Press.

Allen, M., & Friston, K. J. (2018). From cognitivism to autopoiesis: towards a computational framework for the embodied mind. *Synthese*, *195*(6), 2459–2482.

Allport, A. (1989). Visual attention. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 631 – 682). Cambridge, USA (MA): MIT press.

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired Magazine*, *16*(7), 16–07.

Anderson, E. (2020). Feminist epistemology and philosophy of science. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Spring 2020 ed.). Metaphysics Research Lab, Stanford University.

Anderson, J. R. (1983). *Rules of the mind.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, M. L. (2007, March). Evolution of cognitive function via redeployment of brain areas. *Neuroscientist*, *13*(1), 13–21.

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences*, *33*(4), 245–266.

Andrade, J. (2001). The working memory model: Consensus, controversy, and future. In J. Andrade (Ed.), *Working memory in perspective* (p. 322-352). London, UK: Psychology Press.

Anjomshoae, S., Najjar, A., Calvaresi, D., & Främling, K. (2019, May). Explainable agents and robots: Results from a systematic literature review. In *18th international conference on autonomous agents and multiagent systems (AAMAS 2019)* (pp. 1078–1088). Montreal, CA.

Arbib, M. (1981). Perceptual structures and distributed motor control. In V. Brooks (Ed.), *Handbook of physiology: Section II the nervous system. Part 1: Motor control.* American Physiological Society.

Arbib, M. A. (1992). Schema theory. *The Encyclopedia of Artificial Intelligence*, *2*, 1427–1443.

Arbib, M. A. (2019). From action to typology? A neuro-evolutionary perspective. *Language and Linguistics Compass*, *9*(2), 102–117.

Arbib, M. A., Gasser, B., & Barrès, V. (2014). Language is handy but is it embodied? *Neuropsychologia*, *55*, 57–70.

Averbeck, B. B., Chafee, M. V., Crowe, D. A., & Georgopoulos, A. P. (2002). Parallel processing of serial movements in prefrontal cortex. *Proceedings of the National Academy of Sciences*, *99*(20), 13172–13177.

Averbeck, B. B., & Lee, D. (2007). Prefrontal neural correlates of memory for sequences. *Journal of Neuroscience*, *27*(9), 2204–2211.

Awais, M., & Henrich, D. (2012). Proactive premature intention estimation for intuitive human-robot collaboration. In *2012 IEEE/RSJ International conference on intelligent robots and systems* (pp. 4098–4103).

Bach-y Rita, P. (2004). Tactile sensory substitution studies. *Annals of the New York Academy of Sciences*, *1013*(1), 83–91.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423.

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, *36*(3), 189–208.

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of psychology*, *63*, 1–29.

Baddeley, A., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General*, *130*(4), 641.

Baddeley, A. D., & Andrade, J. (2000). Working memory and the vividness of imagery. *Journal of Experimental Psychology: General*, *129*(1), 126.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47–89). Elsevier.

Baggs, E., & Chemero, A. (2018, Nov). Radical embodiment in two directions. *Synthese*.

Ballard, D., & Rao, R. (1995). An active vision architecture based on iconic representations. *Artificial Intelligence*, *78*, 461 – 505.

Ballard, D. H. (1986). Cortical connections and parallel processing: Structure and function. *Behavioral and Brain Sciences*, *9*(1), 67–90.

Ballard, D. H. (1991). Animate vision. *Artificial Intelligence*, *48*, 57–86.

Ballard, D. H., Hayhoe, M. M., Li, F., & Whitehead, S. D. (1992). Hand-eye coordination during sequential tasks. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *337*(1281), 331–339.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*(4), 723–742.

Barad, K. (1999). Agential realism: Feminist interventions in understanding scientific practices. In M. Biagioli (Ed.), *The science studies reader* (pp. 1–11). New York, USA (NY): Routledge.

Barsalou, L. W. (1999, aug). Perceptual symbol systems. *Behavioural Brain Sciences*, *22*(4), 577–609.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *364*(1521).

Barsalou, L. W. (2010). Grounded cognition: Past, present, and future. *Topics in Cognitive Science*, *2*(4), 716–724.

Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7*(2), 84–91.

Bascetta, L., Ferretti, G., Rocco, P., Ardö, H., Bruyninckx, H., Demeester, E., & Di Lello, E. (2011). Towards safe human-robot interaction in robotic cells: An approach based on visual tracking and intention estimation. In *2011 IEEE/RSJ International conference on intelligent robots and systems* (pp. 2971–2978).

Bauer, A., Wollherr, D., & Buss, M. (2008). Human–robot collaboration: a survey. *International Journal of Humanoid Robotics*, *5*(01), 47–66.

Beer, R. D., Chiel, H. J., Quinn, R. D., & Ritzmann, R. E. (1998). Biorobotic approaches to the study of motor systems. *Current Opinion in Neurobiology*, *8*(6), 777–782.

Beer, R. D., Quinn, R. D., Chiel, H. J., & Ritzmann, R. E. (1997). Biologically inspired approaches to robotics: What can we learn from insects? *Communications of the ACM*, *40*(3), 30–38.

Beiser, D. G., & Houk, J. C. (1998). Model of cortical-basal ganglionic processing: encoding the serial order of sensory events. *Journal of Neurophysiology*, *79*(6), 3168–3188.

Benedek, M., Stoiser, R., Walcher, S., & Körner, C. (2017). Eye behavior associated with internally versus externally directed cognition. *Frontiers in Psychology*, *8*, 1092.

Benton, A. L. (1991). The prefrontal region: Its early history. In H. S. Levin, H. M. Eisenberg, & A. L. Benton (Eds.), *Frontal lobe function and dysfunction* (pp. 3–33). New York, NY: Oxford University Press.

Berlucchi, G., & Aglioti, S. M. (2010). The body in the brain revisited. *Experimental Brain Research*, *200*(1), 25.

Beuter, N., Spexard, T., Lutkebohle, I., Peltason, J., & Kummert, F. (2008). Where is this?-gesture based multimodal interaction with an anthropomorphic robot. In *Humanoids 2008-8th IEEE-RAS International conference on humanoid robots* (pp. 585–591).

Bi, L., Guan, C., et al. (2019). A review on emg-based motor intention prediction of continuous human upper limb motion for human-robot collaboration. *Biomedical Signal Processing and Control*, *51*, 113–127.

Bishop, L., Cancino-Chacón, C., & Goebl, W. (2019). Eye gaze as a means of giving and seeking information during musical interaction. *Consciousness and Cognition*, *68*, 73–96.

Bishop, L., & Goebl, W. (2018, 08). Performers and an active audience: Movement in music production and perception. *Jahrbuch Musikpsychologie*, *28*. doi: 10.5964/jbdgm.2018v28.19

Bluhm, R., Maibom, H. L., & Jacobson, A. J. (2012). *Neurofeminism: Issues at the intersection of feminist theory and cognitive science*. New York, USA (NY): Springer.

Boden, M. A. (2008). *Mind As Machine: A History Of Cognitive Science*. Oxford, UK: Oxford University Press.

Boehner, K., David, S., Kaye, J., & Sengers, P. (2005). Critical technical practice as a methodology for values in design. In *Chi 2005 workshop on quality, values, and choices* (pp. 2–7).

Boehner, K., DePaula, R., Dourish, P., & Sengers, P. (2005). Affect: from information to interaction. In *Proceedings of the 4th decennial conference on critical computing: between sense and sensibility* (pp. 59–68).

Boehner, K., DePaula, R., Dourish, P., & Sengers, P. (2007). How emotion is made and measured. *International Journal of Human-Computer Studies*, *65*(4), 275–291.

Bonini, C. P. (1963). *Simulation of information and decision system in the firm*. Upper Sadie River, USA (NJ): Prentice-Hall.

Borghi, A., & Pecher, D. (2011). Introduction to the special topic embodied and grounded cognition. *Frontiers in Psychology*, *2*, 187.

Böschen, S. (2019). Processing issues in science policy: emerging epistemic regimes. In J. S. Dagmar Simon Stefan Kuhlmann & W. Canzler (Eds.), *Handbook on science and public policy* (pp. 317 – 335). Cheltenham, UK: Edward Elgar Publishing.

Bowker, G. C. (2014). Big data, big questions – the theory/data thing. *International Journal of Communication*, *8*, 1795–1799.

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

Bruner, J. S. (1990). *Acts of meaning* (Vol. 3). Boston, USA (MA): Harvard University Press.

Bryant, D. J., & Geoffrey Wright, W. (1999). How body asymmetries determine accessibility in spatial frameworks. *The Quarterly Journal of Experimental Psychology: Section A*, *52*(2), 487–508.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Annals of the New York Academcy of Science*, *1124*, 1 – 38.

Cameron, D. (1992). *Feminism and linguistic theory.* New York, USA (NY): Springer.

Cangelosi, A., & Harnad, S. (2002). The adaptive advantage of symbolic theft over sensorimotor toil: Grounding language in perceptual categories. *Evolution of Communication*, *4*(1).

Cangelosi, A., & Riga, T. (2006). An embodied model for sensorimotor grounding and grounding transfer: Experiments with epigenetic robots. *Cognitive Science*, *30*(4), 673–689.

Carandini, M., Demb, J. B., Mante, V., Tolhurst, D. J., Dan, Y., Olshausen, B. A., ... Rust, N. C. (2005). Do we know what the early visual system does? *Journal of Neuroscience*, *25*(46), 10577–10597.

Castiello, U. (2005). The neuroscience of grasping. *Nature Reviews Neuroscience*, *6*(9), 726–736.

Censor, N., Sagi, D., & Cohen, L. G. (2012). Common mechanisms of human perceptual and motor learning. *Nature Reviews Neuroscience*, *13*(9), 658–664.

Chemero, A. (2009, 01). Radical embodied cognitive science. *Philosophical Psychology*, *17*.

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

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use.* Santa Barbara, USA (CA): Greenwood Publishing Group.

Chomsky, N. (2014). *The minimalist program.* Boston, USA (MA): MIT Press.

Chun, M. M. (2011). Visual working memory as visual attention sustained internally over time. *Neuropsychologia*, *49*(6), 1407 – 1409.

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology*, *62*, 73–101.

Clark, A. (2006a). Language, embodiment, and the cognitive niche. *Trends in cognitive sciences*, *10*(8), 370–374.

Clark, A. (2006b). Vision as dance? three challenges for sensorimotor contingency theory. *Psyche*, *12*(1).

Clark, A. (2014). *Mindware: An introduction to the philosophy of cognitive science.* Oxford, UK: Oxford University Press.

Clarke, A. E., & Fujimura, J. H. (2014). *The right tools for the job: At work in twentieth-century life sciences* (Vol. 149). Princeton, USA (NJ): Princeton University Press.

Coeckelbergh, M. (2011). Are emotional robots deceptive? *IEEE Transactions on Affective Computing*, *3*(4), 388–393.

Cohen, M. (2015). *Paradigm shift: How expert opinions keep changing on life, the universe, and everything.* Exeter, UK: Imprint Academic.

Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience*, *22*(1), 319–349.

Cole, M., Hood, L., & McDermott, R. P. (1997). Concepts of ecological validity: Their differing implications for comparative cognitive research. In M. Cole, Y. Engestrom, & O. Vasquez (Eds.), *Mind, culture, and activity: Seminal papers from the laboratory of comparative human cognition* (pp. 49–56). New York, NY: Cambridge University Press.

Collins, P. H. (1989). The social construction of black feminist thought. *Signs: Journal of Women in Culture and Society*, *14*(4), 745–773.

Connor, C. E., Egeth, H. E., & Yantis, S. (2004). Visual attention: bottom-up versus top-down. *Current Biology*, *14*(19), R850–R852.

Cox, R. V., Kamm, C. A., Rabiner, L. R., Schroeter, J., & Wilpon, J. G. (2000). Speech and language processing for next-millennium communications services. *Proceedings of the IEEE*, *88*(8), 1314–1337.

Craik, K. J. W. (1943). *The nature of explanation.* Cambridge, UK: Cambridge University Press.

Croft, W. (2007). The origins of grammar in the verbalization of experience. *Cognitive Linguistics*, *18*(3), 339–382.

Damasio, A. R. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness.* Boston, USA (MA): Houghton Mifflin Harcourt.

Daugman, J. (1990). Brain metaphor and brain theory. In E. L. Schwartz (Ed.), *Computational neuroscience* (pp. 9 – 18). Boston, USA (MA): MIT Press.

Degenaar, J., & O'Regan, J. K. (2017). Sensorimotor theory and enactivism. *Topoi*, *36*(3), 393–407.

de Jong, M., Zhang, K., Roth, A. M., Rhodes, T., Schmucker, R., Zhou, C., ... Veloso, M. (2018). Towards a robust interactive and learning social robot. In *Proceedings of the 17th international conference on autonomous agents and multiagent systems* (pp. 883–891).

Desmurget, M., Reilly, K. T., Richard, N., Szathmari, A., Mottolese, C., & Sirigu, A. (2009). Movement intention after parietal cortex stimulation in humans. *Science*, *324*(5928), 811–813.

D'Esposito, M., Ballard, D., Zarahn, E., & Aguirre, G. (2000). The role of prefrontal cortex in sensory memory and motor preparation: An event-related fmri study. *NeuroImage*, *11*(5), 400 - 408.

Dijkstra, K., Kaschak, M. P., & Zwaan, R. A. (2007). Body posture facilitates retrieval of autobiographical memories. *Cognition*, *102*(1), 139–149.

Dijkstra, K., & Post, L. (2015). Mechanisms of embodiment. *Frontiers in psychology*, *6*, 1525.

Dodd, W., & Gutierrez, R. (0205). The role of episodic memory and emotion in a cognitive robot. In *ROMAN 2005. IEEE international workshop on robot and human interactive communication, 2005.* (p. 692-697).

Dominey, P., Arbib, M., & Joseph, J.-P. (1995). A model of corticostriatal plasticity for learning oculomotor associations and sequences. *Journal of Cognitive Neuroscience*, *7*(3), 311–336.

Driver, J., & Spence, C. (1998). Cross–modal links in spatial attention. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *353*(1373), 1319–1331.

Dumit, J. (2003). Is it me or my brain? Depression and neuroscientific facts. *Journal of Medical Humanities*, *24*(1-2), 35–47.

Dumit, J. (2004). Travelling Images, Popularizing Brains. In *Picturing personhood: Brain scans and biomedical identity* (pp. 139–169). Princeton, USA (NJ): Princeton University Press.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433.

Edmonds, M., Gao, F., Liu, H., Xie, X., Qi, S., Rothrock, B., ... Zhu, S.-C. (2019). A tale of two explanations: Enhancing human trust by explaining robot behavior. *Science Robotics*, *4*(37).

Ellis, S. R. (2016). *Anthropomorphic robot design and user interaction associated with motion* (techreport Nos. NASA/TM-2016-219426, ARC-E-DAA-TN35939). Moffett Field, USA (CA): National Aeronautics and Space Administration Ames Research Center.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211.

Engelkamp, J. (1998). *Memory for actions*. Hove, UK: Psychology Press/Taylor & Francis (UK).

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: a three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864.

Eriksen, C. W., & Hoffman, J. E. (1972). Temporal and spatial characteristics of selective encoding from visual displays. *Perception & Psychophysics*, *12*(2), 201–204.

Evans, V., & Green, M. (2006). *Cognitive linguistics: An introduction*. Edinburgh, UK: Edinburgh University Press.

Fagg, A. H., & Arbib, M. A. (1998). Modeling parietal–premotor interactions in primate control of grasping. *Neural Networks*, *11*(7-8), 1277–1303.

Fecher, B., & Friesike, S. (2014). Open science: One term, five schools of thought. In S. Bartling & S. Friesike (Eds.), *Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing*. Cham: Springer International Publishing.

Feitelson, J., & Stefik, M. (1977). A case study of the reasoning in a genetics experiment. *Heuristic Programming Project Memo HPP-77-18 (working paper)(May 1977)*.

Feldman, J., & Narayanan, S. (2004). Embodied meaning in a neural theory of language. *Brain and Language*, *89*(2), 385–392.

Fillmore, C. J. (1969). Types of lexical information. In D. Steinberg & L. Jacobovitz (Eds.), *Studies in syntax and semantics* (pp. 109–137). Springer.

Findlay, J. M., & Gilchrist, I. D. (2001). Visual attention: The active vision perspective. In *Vision and attention* (pp. 83–103). Springer.

Flanagan, J. R., & Johansson, R. S. (2003, August). Action plans used in action observation. *Nature*, *424*(6950), 769–771.

Fodor, J. A. (1983). *The modularity of mind*. Boston, USA (MA): MIT press.

Fogassi, L., & Gallese, V. (2004). Action as binding key to multisensory integration. In G. Calvert, C. Spence, & B. Stein (Eds.), *Handbook of multi-sensory processing*. Cambridge, USA (MA): MIT press.

Franceschini, N. (2014). Small brains, smart machines: from fly vision to robot vision and back again. *Proceedings of the IEEE*, *102*(5), 751–781.

Franceschini, N., Pichon, J.-M., & Blanes, C. (1992). From insect vision to robot vision. *Philosophical Transactions of The Royal Society Of London. Series B: Biological Sciences*, *337*(1281), 283–294.

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301.

Fum, D., Missier, F. D., & Stocco, A. (2007). The cognitive modeling of human behavior: Why a model is (sometimes) better than 10,000 words. *Cognitive Systems Research*, *8*(3), 135–142.

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Sciences*, *4*(1), 14–21.

Gallese, V., Gernsbacher, M. A., Heyes, C., Hickok, G., & Iacoboni, M. (2011). Mirror neuron forum. *Perspectives on Psychological Science*, *6*(4), 369–407.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: the role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, *22*(3-4), 455–479.

Garfinkel, H. (1967). What is ethnomethodology. In *Studies in ethnomethodology.* Englewood Cliffs, USA (NJ): Prentice Hall.

Gelman, R., & Baillargeon, R. (1983). Review of some piagetian concepts. *Handbook of child psychology: Formerly Carmichael's Manual of child psychology/Paul H. Mussen, editor*.

Georgieff, N., & Jeannerod, M. (1998). Beyond consciousness of external reality: a "who" system for consciousness of action and self-consciousness. *Consciousness and Cognition*, *7*(3), 465–477.

Gershkoff-Stowe, L., & Goldin-Medow, S. (2002). Is there a natural order for expressing semantic relations? *Cognitive Psychology*, *45*(3), 375–412.

Gibbons, M., Limoges, C., Nowotny, H., Schwartzmann, S., Scott, P., & M, T. (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies.* London/Thousand Oaks/New Delhi/Singapore: SAGE Publications.

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

Gibson, J. J. (1986). The theory of affordances. In *The ecological approach to visual perception* (pp. 127 – 136). Mahwah, USA (NJ): Lawrence Erlbaum Associates.

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, *57*(4), 544–569.

Glenberg, A. M., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. *cortex*, *48*(7), 905–922.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558–565.

Goldinger, S. D., Papesh, M. H., Barnhart, A. S., Hansen, W. A., & Hout, M. C. (2016, Aug). The poverty of embodied cognition. *Psychonomic Bulletin & Review*, *23*(4), 959–978.

Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences*, *105*(27), 9163–9168.

Goodwin, A. W., & Wheat, H. E. (2004). Sensory signals in neural populations underlying tactile perception and manipulation. *Annual Review of Neuroscience*, *27*, 53–77.

Grafton, S. T. (2010). The cognitive neuroscience of prehension: recent developments. *Experimental Brain Research*, *204*(4), 475–491.

Green, S. A., Billinghurst, M., Chen, X. Q., & Chase, J. G. (2008). Human-robot collaboration: A literature review and augmented reality approach in design. *International Journal of Advanced Robotic Systems*, *5*(1), 1.

Grezes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Human Brain Mapping*, *12*(1), 1–19.

Gross, S., Hirschmanner, M., Krenn, B., Neubarth, F., & Zillich, M. (2018, May). Action verb corpus. In *Proceedings of the 11th international conference on language resources and evaluation (LREC 2018)* (pp. 2147–2151). Miyazaki (Japan).

Grossberg, S. (1978). Behavioral contrast in short-term memory: Serial binary memory models or parallel continuous memory models. *Journal of Mathematical Psychology*, *17*(3), 199–219.

Grüsser, O.-J. (1995). On the history of the ideas of efference copy and reafference. In *Essays in the history of the physiological sciences* (pp. 35–55). Brill Rodopi.

Haggard, P. (2008). Human volition: towards a neuroscience of will. *Nature Reviews Neuroscience*, *9*(12), 934.

Haggard, P., Rossetti, Y., & Kawato, M. (2008). *Sensorimotor foundations of higher cognition* (No. 22). Oxford University Press, USA.

Hajnal, A., Fonseca, S., Kinsella-Shaw, J. M., Silva, P., Carello, C., & Turvey, M. (2007). Haptic selective attention by foot and by hand. *Neuroscience Letters*, *419*(1), 5–9.

Halpern, O., Mitchell, R., & Geoghegan, B. D. (2017). The smartness mandate: Notes toward a critique. *Grey Room*, 106–129.

Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies*, *14*(3), 575–599.

Haraway, D. J. (1989). *Primate visions: Gender, race, and nature in the world of modern science.* London, UK: Psychology Press.

Harding, S. (2006). *Science and social inequality: Feminist and postcolonial issues.* Chicago, USA (IL): University of Illinois Press.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1-3), 335–346.

Harré, R. (2002). *Cognitive science: A philosophical introduction.* London, UK: SAGE.

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

Haselager, P., van Dijk, J., & van Rooij, I. (2008). A lazy brain? embodied embedded cognition and cognitive neuroscience. In *Handbook of cognitive science* (pp. 273–290). Elsevier.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2), 301–307.

Hayes, B., & Scassellati, B. (2013). Challenges in shared-environment human-robot collaboration. In (Vol. 8).

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*(4), 188 - 194.

Heine, B. (1997). *Cognitive foundations of grammar.* London, UK: Oxford University Press.

Heritage, J. (2013). *Garfinkel and ethnomethodology.* Hoboken, USA (NJ): John Wiley & Sons.

Herrmann, G., & Melhuish, C. (2010). Towards safety in human robot interaction. *International Journal of Social Robotics*, *2*(3), 217–219.

Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of Cognitive Neuroscience*, *21*(7), 1229–1243.

Hoffman, G., & Ju, W. (2014). Designing robots with movement in mind. *Journal of Human-Robot Interaction*, *3*(1), 91–122.

Hoffman, J. E., & Subramaniam, B. (1995, Jan). The role of visual attention in saccadic eye movements. *Perception & Psychophysics*, *57*(6), 787–795.

Holmes, N. P., & Spence, C. (2004). The body schema and multisensory representation (s) of peripersonal space. *Cognitive Processing*, *5*(2), 94–105.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (tec): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*(5), 849–878.

Hornstein, N. (2009). *A theory of syntax: Minimal operations and universal grammar.* Cambridge, UK: Cambridge University Press.

Horton, T. E., Chakraborty, A., & Amant, R. S. (2012). Affordances for robots: a brief survey. *AVANT. Pismo Awangardy Filozoficzno-Naukowej*, *2*, 70–84.

Houghton, G., & Hartley, T. (1995). Parallel models of serial behavior: Lashley revisited. *Psyche*, *2*(25), 1–25.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(5), 1217.

Hülse, M., McBride, S., & Lee, M. (2009). Implementing inhibition of return: embodied visual memory for robotic systems. In *9th International conference on epigenetic robotics: Modeling cognitive development in robotic systems* (pp. 213–214).

Hutchins, J. (1999, September). Retrospect and prospect in computer-based translation. In *Proceedings of machine translation summit VII* (pp. 30–44). Kent Ridge Labs, Singapore.

Irigaray, L. (1989). Is the subject of science sexed (C. M. Bové, Trans.). In N. Tuana (Ed.), (pp. 58 – 68). Bloomington, USA (IN): Indiana University Press.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*(3), 194–203.

Jackendoff, R. (1992). *Semantic structures* (Vol. 18). Boston, USA (MA): MIT press.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution.* New York, USA (NY): Oxford University Press.

Jackendoff, R. (2003). Précis of foundations of language: brain, meaning, grammar, evolution. *Behavioral and Brain Sciences*, *26*(6), 651–665.

Jeannerod, M. (1981). Intersegmental coodination during reaching at natural visual objects. In A. B. M Jeannerod J Long (Ed.), *Attention and performance* (pp. 153–169). Hillsboro, NJ: Lawrence Erlbaum Associates.

Jeannerod, M. (1996). Reaching and grasping. parallel specification of visuomotor channels. In *Handbook of perception and action* (Vol. 2, pp. 405–460). Elsevier.

Jeannerod, M. (2001). Neural simulation of action: a unifying mechanism for motor cognition. *Neuroimage*, *14*(1), S103–S109.

Jeannerod, M., Paulignan, Y., & Weiss, P. (1998). Grasping an object: one movement, several components. In *Novartis foundation symposium* (pp. 5–15).

Jirak, D., Menz, M. M., Buccino, G., Borghi, A. M., & Binkofski, F. (2010). Grasping language–a short story on embodiment. *Consciousness and cognition*, *19*(3), 711–720.

Johansson, R. S., Westling, G., Bäckström, A., & Flanagan, J. R. (2001). Eye–hand coordination in object manipulation. *Journal of Neuroscience*, *21*(17), 6917–6932.

Jugnet, A., & Lhôte, E. (2019). Looking at 'unexpectedness': a corpus-based cognitive analysis of surprise and wonder. In N. Depraz & A. Celle (Eds.), *Surprise at the intersection of phenomenology and linguistics* (pp. 140–169). Amsterdam, NL: John Benjamins Publishing Company.

Kent, C., & Lamberts, K. (2008). The encoding–retrieval relationship: retrieval as mental simulation. *Trends in Cognitive Sciences*, *12*(3), 92–98.

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: theoretical developments, current evidence and future directions. *cortex*, *48*(7), 805–825.

Kilner, J. M. (2011). More than one pathway to action understanding. *Trends in Cognitive Sciences*, *15*(8), 352–357.

Kitcher, P. (1984). 1953 and all that. a tale of two sciences. *The Philosophical Review*, *93*(3), 335–373.

Klein, J. T. (2010). A taxonomy of interdisciplinarity. In C. Mitcham (Ed.), *The oxford handbook of interdisciplinarity* (Vol. 15, pp. 15–30). Oxford, UK: Oxford University Press.

Knott, A. (2010). A sensorimotor characterisation of syntax, and its implications for models of language evolution. In S (Ed.), *The evolution of language* (pp. 214–221). World Scientific.

Knott, A. (2012). *Sensorimotor Cognition and Natural Language Syntax.* Boston, USA (MA): MIT Press.

Knott, A. (2014). Syntactic structures as descriptions of sensorimotor processes. *Biolinguistics*, *8*, 1–52.

Knott, A. (2017a). *An extended model of deictic routines, supporting a wider-coverage sm interpretation of syntax* (techreport No. OUCS-2017-02). Dunedin, NZ: Department of Computer Science, University of Otago, New Zealand.

Knott, A. (2017b). *Road map for an embodied model of language and cognition* (techreport No. OUCS-2017-01). Dunedin, NZ: Department of Computer Science, University of Otago, New Zealand.

Knott, A. (2017c). *Sensorimotor cognition and natural language syntax part ii* (techreport No. OUCS-2017-04). Dunedin, NZ: Department of Computer Science, University of Otago, New Zealand.

Knott, A., & Takac, M. (2016). *A simulationist model of episode representations in working memory: syntactic interpretation, nested episodes and storage requirements* (techreport No. OUCS-2016-04). Dunedin, NZ: Department of Computer Science, University of Otago, New Zealand.

Knuuttila, T. (2011). Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science: Part A*, *42*(2), 262 - 271.

Koch, C., & Ullman, S. (1987). Shifts in selective visual attention: towards the underlying neural circuitry. In *Matters of intelligence* (pp. 115–141). Springer.

Kolers, P. A., & Roediger III, H. L. (1984). Procedures of mind. *Journal of Verbal Learning and Verbal Behavior*, *23*(4), 425–449.

Kornhuber, H. H., & Deecke, L. (2016). Brain potential changes in voluntary and passive movements in humans: readiness potential and reafferent potentials. *Pflügers Archiv – European Journal of Physiology*, *468*(7), 1115–1124.

Krenn, B., Sadeghi, S., Neubarth, F., Gross, S., Trapp, M., & Scheutz, M. (2020). Models of cross-situational and crossmodal word learning in task-oriented scenarios. *IEEE Transactions on Cognitive and Developmental Systems*.

Krenn, B., Trapp, M., Gross, S., & Neubarth, F. (2017, Sep). Crossmodal cross-situational learning with attention. In *Seventh joint IEEE international conference on development and learning and on epigenetic robotics: Workshop on computational models for crossmodal learning.* Lisbon, Portugal.

Krenn, B., Tsiourti, C., Neubarth, F., Gross, S., & Hirschmanner, M. (2019). Active language learning inspired from early childhood information seeking strategies: Workshop on cognitive architectures for human-robot interaction. In *Proceedings of the 18th international conference on autonomous agents and multiagent systems.* Montreal, Canada.

Kuhn, T. S. (1962). *Structure of scientific revolutions* (1st ed.). Chicago, USA (IL): Chicago University Press.

Kusunoki, M., Gottlieb, J., & Goldberg, M. E. (2000). The lateral intraparietal area as a salience map: the representation of abrupt onset, stimulus motion, and task relevance. *Vision Research*, *40*(10-12), 1459–1468.

Labov, W. (1971). The notion of 'system'in creole studies. In *Pidginization and creolization of languages* (pp. 447 – 472). Cambridge, UK: Cambridge University Press.

Lakatos, I. (1976). *Proofs and refutations: The logic of mathematical discovery.* Cambridge, UK: Cambridge University Press.

Lakatos, I., & Feyerabend, P. (1999). *For and against method: including lakatos's lectures on scientific method and the lakatos-feyerabend correspondence.* Chicago, USA (IL): University of Chicago Press.

Lakoff, G. (1973). Fuzzy grammar and the performance/competence terminology game. *TRansactions of the Chiacgo Linguistic Society*, 271 – 291.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago, USA (IL): University of Chicago Press.

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*(11), 1311–1328.

Land, M. F. (2009). Vision, eye movements, and natural behavior. *Visual Neuroscience*, *26*(1), 51–62.

Land, M. F., & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *352*(1358), 1231–1239.

Latour, B., & Woolgar, S. (2013). *Laboratory life: The construction of scientific facts.* Princeton, USA (NJ): Princeton University Press.

Lederman, S. J., & Klatzky, R. L. (2009). Haptic perception: A tutorial. *Attention, Perception, & Psychophysics*, *71*(7), 1439–1459.

Li, Y., & Ge, S. S. (2013). Human–robot collaboration based on motion intention estimation. *IEEE/ASME Transactions on Mechatronics*, *19*(3), 1007–1014.

Lighthill, J. (1973). A report on artificial intelligence. *UK Science and Engineering Research Council*.

Liu, H., & Wang, L. (2017). Human motion prediction for human-robot collaboration. *Journal of Manufacturing Systems*, *44*, 287–294.

Longino, H. E. (1987). Can there be a feminist science? *Hypatia*, *2*(3), 51–64.

Longuet-Higgins, H. C. (1973). Comments on the lighthill report and the sutherland reply. In *Artificial intelligence: A paper symposium* (pp. 35–37).

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), 347–356.

Machery, E. (2007, jun). Concept empiricism: a methodological critique. *Cognition*, *104*(1), 19–46.

Mahon, B. Z. (2015). What is embodied about cognition? *Language, Cognition and Neuroscience*, *30*(4), 420–429.

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris*, *102*(1), 59–70. doi: 10.1016/j.jphysparis.2008.03.004

Maravita, A., Spence, C., & Driver, J. (2003). Multisensory integration and the body schema: close to hand and within reach. *Current Biology*, *13*(13), R531–R539.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* Boston, USA (MA): MIT Press.

Maturana, H., & Varela, F. (1980). *Autopoiesis and cognition.* Dordrecht, NL: D. Reidel.

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

McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. (1956). *A proposal for the dartmouth summer research project on artificial intelligence* (Tech. Rep.). Hanover, USA (NH): Darmouth College.

McClamrock, R. (1991). Marr's three levels: A re-evaluation. *Minds and Machines*, *1*(2), 185–196.

Meziane, R., Otis, M. J.-D., & Ezzaidi, H. (2017). Human-robot collaboration while sharing production activities in dynamic environment: Spader system. *Robotics and Computer-Integrated Manufacturing*, *48*, 243–253.

Miłkowski, M. (2013). Limits of computational explanation of cognition. In *Philosophy and theory of artificial intelligence* (pp. 69–84). Springer.

Miłkowski, M. (2017). Situatedness and embodiment of computational systems. *Entropy*, *19*(4), 162.

Miłkowski, M. (2018). Objections to computationalism: A survey. *Roczniki Filozoficzne*, *66*(3), 57–75.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*(1), 167–202.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97.

Miller, G. A., & Chomsky, N. (1963). Finitary models of language users. In D. Luce (Ed.), *Handbook of mathematical psychology* (pp. 2–419). John Wiley & Sons.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior.* New York, USA (NY): Rinehart and Winston.

Milner, D., & Goodale, M. (2006). *The visual brain in action* (Vol. 27). Oxford, UK: Oxford University Press.

Milroy, L. (1985). What a performance! some problems with the competence-performance distinction. *Australian Journal of Linguistics*, *5*(1), 1–17.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134–140.

Moore, P., & Robinson, A. (2016). The quantified self: What counts in the neoliberal workplace. *new media & society*, *18*(11), 2774–2792.

Mossio, M., & Taraborelli, D. (2008). Action-dependent perceptual invariants: From ecological to sensorimotor approaches. *Consciousness and Cognition*, *17*(4), 1324–1340.

Mountcastle, V. B. (2005). *The sensory hand: neural mechanisms of somatic sensation.* Boston, USA (MA): Harvard University Press.

Murata, A., Gallese, V., Luppino, G., Kaseda, M., & Sakata, H. (2000). Selectivity for the shape, size, and orientation of objects for grasping in neurons of monkey parietal area aip. *Journal of Neurophysiology*, *83*(5), 2580–2601.

Myachykov, A., Garrod, S., & Scheepers, C. (2009). Attention and syntax in sentence production: A critical review. *Discours. Revue de Linguistique, Psycholinguistique et Informatique. A journal of Linguistics, Psycholinguistics and Computational Linguistics*(4).

Nagel, S. K., Carl, C., Kringe, T., Märtin, R., & König, P. (2005). Beyond sensory substitution—learning the sixth sense. *Journal of Neural Engineering*, *2*(4), R13.

Nelson, L. (1992). *Who knows: From quine to a feminist empiricism.* Temple University Press.

Nelson, L. H. (1996). Empiricism without dogmas. In *Feminism, science, and the philosophy of science* (pp. 95–119). Springer.

Newell, A. (1980). Physical symbol systems. *Cognitive Science*, *4*(2), 135–183.

Newell, A. (1994). *Unified theories of cognition.* Boston, USA (MA): Harvard University Press.

Newen, A., De Bruin, L., & Gallagher, S. (2018). *The oxford handbook of 4e cognition.* London, UK: Oxford University Press.

Nobre, A. C., Gitelman, D., Dias, E., & Mesulam, M.-M. (2000). Covert visual spatial orienting and saccades: overlapping neural systems. *Neuroimage*, *11*(3), 210–216.

Nowakowski, P. R. (2017). Bodily processing: the role of morphological computation. *Entropy*, *19*(7), 295.

Orr, J. (2006). *Panic diaries: a genealogy of panic disorder*. Durham, USA (NC): Duke University Press.

Ortner, S. B. (1972). Is female to male as nature is to culture? *Feminist Studies*, *1*(2), 5–31.

Outram, D. (2005). *The enlightenment* (2nd ed.). London, UK: Cambridge University Press.

Oztop, E., & Arbib, M. A. (2002). Schema design and implementation of the grasp-related mirror neuron system. *Biological Cybernetics*, *87*(2), 116–140.

Peirce, C. S. (1995). Philosophical writings of peirce. In B. Justus (Ed.), (p. 98-119). New York, USA: Dover Publications Inc.

Perera, V., & Veloso, M. (2018). Interpretability of a service robot: Enabling user questions and checkable answers. In D. Lee, A. Steen, & T. Walsh (Eds.), *Proceedings of the fourth global conference on artificial intelligence* (pp. 176–187).

Pezzulo, G., Barsalou, L., Cangelosi, A., Fischer, M., McRae, K., & Spivey, M. (2013). Computational grounded cognition: A new alliance between grounded cognition and computational modeling. *Frontiers in Psychology*, *3*(612), 1-11.

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., Spivey, M., & McRae, K. (2011). The mechanics of embodiment: A dialog on embodiment and computational modeling. *Frontiers in psychology*, *2*, 5.

Piaget, J. (1971). *Biology and knowledge*. Chicago, IL: Chicago University Press.

Piaget, J. (1977). The role of action in the development of thinking. In *Knowledge and development* (pp. 17–42). Springer.

Pigliucci, M. (2009, June). The end of theory in science? *EMBO reports*, *10*, 534–534.

Pillinger, A. (2019). *Gender and feminist aspects in robotics* (techreport No. European Commission topic SwafS-03-2016-2017). Vienna, AUT: Technical University of Vienna.

Pinto, Y., van der Leij, A. R., Sligte, I. G., Lamme, V. A., & Scholte, H. S. (2013). Bottom-up and top-down attention are independent. *Journal of Vision*, *13*(3), 16–16.

Pitts-Taylor, V. (2015). A feminist carnal sociology?: Embodiment in sociology, feminism, and naturalized philosophy. *Qualitative Sociology*, *38*(1), 19–25.

Plunkett, K., Sinha, C., Møller, M. F., & Strandsby, O. (1992). Symbol grounding or the emergence of symbols? vocabulary growth in children and a connectionist net. *Connection Science*, *4*(3-4), 293–312.

Poline, J.-B., Breeze, J., Ghosh, S., Gorgolewski, K., Halchenko, Y., Hanke, M., … Kennedy, D. (2012). Data sharing in neuroimaging research. *Frontiers in Neuroinformatics*, *6*, 9.

Pollard, C., & Sag, I. A. (1994). *Head-driven phrase structure grammar*. Chicago, IL: University of Chicago Press.

Pöppel, E. (1994). Temporal mechanisms in perception. *International Review of Neurobiology*, *37*, 185–202.

Popper, K. (1959). *The logic of scientific discovery*. London, UK: Hutchison.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25.

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, *22*(2), 253–279.

Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the finst spatial-index model. *Cognition*, *32*(1), 65–97.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506.

Reinboth, T., Gross, S., Bishop, L., & Krenn, B. (2020, May). Linguistic, kinematic and gaze information in task descriptions: The lkg-corpus. In *Proceedings of the 12th international conference on language resources and evaluation (LREC 2020).* Marseille (France).

Rhodes, B. J., Bullock, D., Verwey, W. B., Averbeck, B. B., & Page, M. P. (2004). Learning and production of movement sequences: Behavioral, neurophysiological, and modeling perspectives. *Human Movement Science*, *23*(5), 699–746.

Richardson, M. J., Shockley, K., Fajen, B. R., Riley, M. A., & Turvey, M. T. (2008). Ecological psychology: Six principles for an embodied–embedded approach to behavior. In *Handbook of cognitive science* (pp. 159–187). Elsevier.

Rifkin, J. (1995). *The end of work.* New York, USA (NY): Putnam.

Ritter, N. (2005). The role of linguistics in cognitive science. In N. Ritter (Ed.), *Linguistic review [special issue].* Amsterdam, NL: De Gruyter Mouton.

Rivas, C. (2018). Finding themes in qualitative data. In C. Seale (Ed.), *Researching society and culture* (pp. 431–453). Sage Publications Sage CA: Los Angeles, CA.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, *21*(5), 188–194.

Rodríguez-Algarra, F., Sturm, B., & Dixon, S. (2019). Characterising confounding effects in music classification experiments through interventions. *Transactions of the International Society for Music Information Retrieval*, 52–66.

Roediger, H. L. (1980). Memory metaphors in cognitive psychology. *Memory & Cognition*, *8*(3), 231–246.

Roy, D. (2003). Grounded spoken language acquisition: Experiments in word learning. *IEEE Transactions on Multimedia*, *5*(2), 197–209.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592.

Rucci, M., McGraw, P. V., & Krauzlis, R. J. (2016). Fixational eye movements and perception. *Vision Research*, *118*, 1 - 4.

Rupert, R. D. (2009). *Cognitive systems and the extended mind.* Oxford University Press.

Rushworth, M., Johansen-Berg, H., Göbel, S. M., & Devlin, J. (2003). The left parietal and premotor cortices: motor attention and selection. *Neuroimage*, *20*, 89–100.

Rushworth, M. F., Nixon, P. D., Renowden, S., Wade, D. T., & Passingham, R. E. (1997). The left parietal cortex and motor attention. *Neuropsychologia*, *35*(9), 1261 - 1273. doi: https://doi.org/10.1016/S0028-3932(97)00050-X

Sabaté, M., González, B., & Rodríguez, M. (2004). Brain lateralization of motor imagery: Motor planning asymmetry as a cause of movement lateralization. *Neuropsychologia*, *42*(8), 1041-1049.

Sadeghi, S., & Scheutz, M. (2017). Joint acquisition of word order and word referent in a memory-limited and incremental learner. In *2017 8th IEEE international conference on cognitive infocommunications (coginfocom)* (pp. 000357–000362).

Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., & Leventhal, A. G. (1998). Signal timing across the macaque visual system. *Journal of Neurophysiology*, *79*(6), 3272–3278.

Schneider, W. X., & Deubel, H. (2002). Selection-for-perception and selection-for-spatial-motor-action are coupled by visual attention: A review of recent findings and new evidence from stimulus-driven saccade control. *Attention and performance XIX: Common mechanisms in perception and action*, *19*, 609–627.

Schoner, G., & Spencer, J. P. (2006). An embodied approach to cognitive systems: A dynamic neural field theory of spatial working memory. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 28).

Schreitter, S., & Krenn, B. (2016, May). The OFAI multi-modal task description corpus. In *Proceedings of the tenth international conference on language resources and evaluation (lrec'16)* (pp. 1408–1414). Portorož, Slovenia.

Schüll, N. D. (2014). *Addiction by design: Machine gambling in las vegas.* Princeton, USA (NJ): Princeton University Press.

Schwarz, A., Ofner, P., Pereira, J., Sburlea, A. I., & Mueller-Putz, G. R. (2017). Decoding natural reach-and-grasp actions from human eeg. *Journal of Neural Engineering*, *15*(1), 016005.

Schwoebel, J., & Coslett, H. B. (2005). Evidence for multiple, distinct representations of the human body. *Journal of Cognitive Neuroscience*, *17*(4), 543–553.

Sciutti, A., Mara, M., Tagliasco, V., & Sandini, G. (2018). Humanizing human-robot interaction: On the importance of mutual understanding. *IEEE Technology and Society Magazine*, *37*(1), 22–29.

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

Segalowitz, S. (1980). Piaget's achilles' heel: a safe soft spot? *Human Development*, *23*(2), 137–140.

Sengers, P. (2018). The engineering of experience. In *Funology 2* (pp. 287–299). Springer.

Shapiro, L. (2011). Embodied cognition. In M. E, R. Samuels, & S. Stich (Eds.), *Oxford handbook of philosophy and cognitive science.* Oxford, UK: Oxford University Press.

Shibasaki, H., & Hallett, M. (2006). What is the bereitschaftspotential? *Clinical Neurophysiology*, *117*(11), 2341–2356.

Shiffrin, R. M., & Nosofsky, R. M. (1994). Seven plus or minus two: a commentary on capacity limitations. *Psychological Review*, *101*(2), 357–361.

Siagian, C., & Itti, L. (2007). Rapid biologically-inspired scene classification using features shared with visual attention. *IEEE transactions on pattern analysis and machine intelligence*, *29*(2), 300–312.

Smith, D. T., & Schenk, T. (2012). The premotor theory of attention: time to move on? *Neuropsychologia*, *50*(6), 1104–1114.

Smolensky, P. (1987). Connectionist ai, symbolic ai, and the brain. *Artificial Intelligence Review*, *1*(2), 95–109.

Smulders, F. T., & Miller, J. O. (2012). The lateralized readiness potential. In E. S. Kappenman & S. J. Luck (Eds.), *The oxford handbook of event-related potential components* (pp. 209–229). Oxford University Press New York, NY.

Soylu, F. (2016). An embodied approach to understanding: Making sense of the world through simulated bodily activity. *Frontiers in Psychology*, *7*, 1914.

Sridharan, M., & Meadows, B. (2019). Towards a theory of explanations for human–robot collaboration. *KI-Künstliche Intelligenz*, *33*(4), 331–342.

Steels, L. (2003). Social language learning. In M. Tokoro & L. Steels (Eds.), *The future of learning* (pp. 133–162). Amsterdam, The Netherlands: IOS Press.

Steels, L. (2008). The symbol grounding problem has been solved. so what's next. In de Vega M., A. Glenberg, & A. Graesser (Eds.), *Symbols and embodiment: Debates on meaning and cognition* (pp. 223–244). New York, NY: Oxford University Press USA.

Stodden, V., Guo, P., & Ma, Z. (2013, 06). Toward reproducible computational research: An empirical analysis of data and code policy adoption by journals. *PLOS ONE*, *8*, 1-8.

Strangleman, T. (2007). The nostalgia for permanence at work? the end of work and its commentators. *The Sociological Review*, *55*(1), 81–103.

Suanda, S. H., Mugwanya, N., & Namy, L. L. (2014). Cross-situational statistical word learning in young children. *Journal of Experimental Child Psychology*, *126*, 395–411.

Suanda, S. H., Smith, L. B., & Yu, C. (2016). The multisensory nature of verbal discourse in parent–toddler interactions. *Developmental Neuropsychology*, *41*(5-8), 324–341.

Suchman, L. (1993). Response to vera and simon's situated action: A symbolic interpretation. *Cognitive Science*, *17*(1), 71 – 86.

Suchman, L. A. (1987). *Plans and situated actions: The problem of human-machine communication.* Cambridge, UK: Cambridge University Press.

Suchman, L. A. (2007). *Human-machine reconfigurations: Plans and situated actions* (2nd ed.). Cambridge, UK: Cambridge University Press.

Sutton, J., & Williamson, K. (2014). Embodied remembering. In L. Shapiro (Ed.), *The routledge handbook of embodied cognition.* London, UK: Routledge.

Svensson, H., Lindblom, J., & Ziemke, T. (2007). Making sense of embodied cognition: Simulation theories of shared neural mechanisms for sensorimotor and cognitive processes. In R. M. F. Tom Ziemke Jordan Zlatev (Ed.), *Body, language and mind* (Vol. 1, pp. 241–270). Berlin, DE: Mouton de Gruyter.

Takac, M., Benuskova, L., & Knott, A. (2012). Mapping sensorimotor sequences to word sequences: A connectionist model of language acquisition and sentence generation. *Cognition*, *125*(2), 288–308.

Takac, M., & Knott, A. (2015, Oct). A neural network model of episode representations in working memory. *Cognitive Computation*, *7*(5), 509–525.

Takac, M., & Knott, A. (2018). A neural network model of event representations: Sensorimotor sequencing, place coding, self-organization, and bayesian inference. In *Computational neuroscience of event cognition (connect).* Dunedin, NZ: University of Otago.

Tettamanti, M., Buccino, G., Saccuman, M., Gallese, V., Danna, M., Scifo, P., ... Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, *17*(2), 273–281.

Thagard, P. (2005). *Mind: Introduction to cognitive science.* Boston, USA (MA): MIT Press.

Thelen, E., & Smith, L. B. (1996). *A dynamic systems approach to the development of cognition and action.* Boston, USA (MA): MIT Press.

Thompson, C. M. (2015). Situated knowledge, feminist and science and technology studies perspectives. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences (second edition)* (Second Edition ed., p. 1 - 4). Oxford: Elsevier.

Thompson, E. (2010). *Mind in life: Biology, phenomenologym and the sciences of the mind.* Boston, USA (MA): Harvard University Press.

Thompson, K. G., & Bichot, N. P. (2005). A visual salience map in the primate frontal eye field. *Progress in Brain Research*, *147*, 249–262.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Boston, USA (MA): Harvard University Press.

Tomasello, M. (2009). *The cultural origins of human cognition.* Boston, USA (MA): Harvard University Press.

Tomlin, R. S. (1986). *Basic word order: Functional principles.* London, UK: Croom Helm.

Toulmin, S. (1953). *The philosophy of science.* London, UK: Hutchison.

Touryan, J., Lance, B. J., Kerick, S. E., Ries, A. J., & McDowell, K. (2016). Common eeg features for behavioral estimation in disparate, real-world tasks. *Biological Psychology*, *114*, 93–107.

Treisman, A. (1991). Search, similarity, and integration of features between and within dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(3), 652.

Trevena, J. A., & Miller, J. (2002). Cortical movement preparation before and after a conscious decision to move. *Consciousness and Cognition*, *11*(2), 162–190.

Troje, N. F., & Westhoff, C. (2006). The inversion effect in biological motion perception: Evidence for a "life detector"? *Current Biology*, *16*(8), 821–824.

Turella, L., & Lingnau, A. (2014). Neural correlates of grasping. *Frontiers in Human Neuroscience*, *8*, 686.

Ullman, S. (1984). Visual routines. *Cognition*, *18*, 97–159.

Vaina, L. M., Solomon, J., Chowdhury, S., Sinha, P., & Belliveau, J. W. (2001). Functional neuroanatomy of biological motion perception in humans. *Proceedings of the National Academy of Sciences*, *98*(20), 11656–11661.

Van den Besselaar, P., & Heimeriks, G. (2001, 01). Disciplinary, multidisciplinary, interdisciplinary -concepts and indicators. In *8th Conference on scientometrics and informetrics.* Sydney, AUS.

Van Elk, M., Slors, M., & Bekkering, H. (2010). Embodied language comprehension requires an enactivist paradigm of cognition. *Frontiers in Psychology*, *1*, 234.

Varela, F., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience.* Boston, USA (MA): MIT Press.

Vera, A. H., & Simon, H. A. (1993). Sitiated action: A symbolic interpretation. *Cognitive Science*, *17*, 7-48.

Viseu, A., & Suchman, L. (2010). Wearable augmentations. In J. Edwards, P. Harvey, & P. Wade (Eds.), *Technologized images, technologized bodies* (pp. 161 – 184). Oxford/New York: Berghahn Books.

Viviani, P. (1990). Eye movements in visual search: Cognitive, perceptual, and motor control aspects. In E. Kowler (Ed.), *Eye movements and their role in visual and cognitive processes* (pp. 353–383). Elsevier Science Publishing Company.

von Uexküll, J. (1928). *Theoretische biologie.* Springer Verlag.

Walcher, S., Körner, C., & Benedek, M. (2017). Looking for ideas: Eye behavior during goal-directed internally focused cognition. *Consciousness and Cognition*, *53*, 165 – 175.

Walles, H., Robins, A. V., & Knott, A. (2013). *A neural network model of visual attention and object classification: technical details.* (techreport No. OUCS-2013-09). Dunedin, NZ: Department of Computer Science, University of Otago, New Zealand.

Watterson, B. (2005). *The complete calvin and hobbes.* Kansas City, USA (MO): Andrews McMeel Publishing.

Webb, A. (2015). *Actions of Attention, and Attention to Action: Investigating the Relationship Between Visual Attention, Episodic Representation, and Language* (PhD Thesis). University of Otago, Dunedin, NZ.

Webb, A., Knott, A., & MacAskill, M. R. (2010). Eye movements during transitive action observation have sequential structure. *Acta Psychologica*, *133*(1), 51–56.

Weber, R. P. (1990). *Basic content analysis.* Los Angeles, USA (CA): Sage Publications Sage.

Wei, S. X. (2002). Resistance is fertile: Gesture and agency in the field of responsive media. *Configurations*, *10*(3), 439–472.

Weichselgartner, E., & Sperling, G. (1987). Dynamics of automatic and controlled visual attention. *Science*, *238*(4828), 778–780.

Weston, C., Gandell, T., Beauchamp, J., McAlpine, L., Wiseman, C., & Beauchamp, C. (2001). Analyzing interview data: The development and evolution of a coding system. *Qualitative Sociology*, *24*(3), 381–400.

Whaley, L. J. (1996). *Introduction to typology: the unity and diversity of language.* New York, USA (NY): SAGE Publications.

Wiener, N. (1949). *Letter to uaw president walter reuther (august 13, 1949).* https://libcom.org/history/father-cybernetics-norbert-wieners-letter-uawpresident-walter-reuther.

Wiener, N. (2019). *Cybernetics or control and communication in the animal and the machine.* Cambridge, USA (MA): MIT press.

Wilson, E. (2016). *Neural geographies: Feminism and the microstructure of cognition.* London, UK: Taylor & Francis.

Wilson, E. A. (1996). Loving the computer' cognition, embodiment and the influencing machine. *Theory & Psychology*, *6*(4), 577–599.

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

Wilson, R. A., & Foglia, L. (2017). Embodied cognition. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Spring 2017 ed.). Metaphysics Research Lab, Stanford University. `https://plato.stanford.edu/archives/spr2017/entries/embodied-cognition/`.

Winges, S. A., Weber, D. J., & Santello, M. (2003). The role of vision on hand preshaping during reach to grasp. *Experimental Brain Research*, *152*(4), 489–498.

Winograd, T., & Flores, F. (1986). *Understanding computers and cognition: A new foundation for design.* Norwood, NJ: Ablex.

Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, *70*(13-15), 2149–2165.

Yu, C., Smith, L. B., & Pereira, A. F. (2008). Grounding word learning in multimodal sensorimotor interaction. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 30).

Ziemke, T. (2001). Are robots embodied. In *First international workshop on epigenetic robotics modeling cognitive development in robotic systems* (Vol. 85).

Ziemke, T. (2003a). On the role of robot simulations in embodied cognitive science. *AISB journal*, *1*(4), 389–399.

Ziemke, T. (2003b). What's that thing called embodiment? In *Proceedings of the annual meeting of the cognitive science society* (Vol. 25).

Ziemke, T. (2005). Cybernetics and embodied cognition: on the construction of realities in organisms and robots. *Kybernetes*, *34*(1/2), 118–128.

Zlatev, J. (2010). Phenomenology and cognitive linguistics. In D. Schmicking & S. Gallagher (Eds.), *Handbook of phenomenology and cognitive science* (pp. 415–443). Springer.

Zlatev, J. (2016). Turning back to experience in cognitive linguistics via phenomenology. *Cognitive Linguistics*, *27*(4), 559–572.

Zlatev, J., Blomberg, J., & David, C. (2010). Translocation, language and the categorization of experience. In *Space in language and cognition: The state of the art and new directions* (pp. 389–418). London, UK: Equinox.

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, *135*(1), 1.

Chapter 9

Appendixes

Appendix 1: German Abstract

Diese Arbeit bietet eine verkörperte Perspektive auf die Nahtstelle zwischen sensomotorischer Kognition und sprachlicher Struktur. Der Ansatz ist von Knott (2012) inspiriert, der vorschlägt, dass die zugrunge liegende Struktur des Satzes "die Frau greift eine Tasse" als sequentielle Repräsentation bestimmter sensomotorischer Prozesse verstanden werden kann, welche in einer vorgegeben Reihenfolge das Greifen einer Tasse ausmachen. Die Arbeit erweitert dabei bestehende empirische Erforschungen dieses Ansatzes, in dem sie eine neue multimodale Studie durchführt die sich an Korpus-basierten Methoden der Erforschung von Regelmäßigkeiten der menschlichen Sprache orientiert. In dem sie das tut, wird die Grundlage für detailierte Nachforschungen zu den sensomotorischen Grundlagen sprachlicher Struktur gelegt. Zu diesem Zweck trägt diese Arbeit folgende Forschritte bei, nämlich i) eine Diskussion der konzeptuellen Grundlagen und des theoretischen Rahmens auf die sie aufbaut, ii) eine eigenständige empirische Methode und Datenverarbeitung, welche einen neuen Korpus ausmachen (Reinboth et al., 2020), iii) eine Software um Visualiserungen der erhobenen Daten zu erstellen, plus iv) Vorschläge für zuküunftige Nachforschungen zu und Anwendungen des beschriebenen Ansatzes, aufbauend auf diesen Visualiserungen, und v) eine Diskussion grundsätzlicher soziopolitischer Themen in die diese Arbeit eingebettet ist. Das bedeutet, dass diese Arbeit Aspekte der künstlichen Intelligenz, Linguistik, Philosophie, Psychologie und der Wissenschafts- und Technikforschung zu einer transdisziplinären Perspektive zum Thema einer sensomotorischen Grundlage der sprachlichen Strukture verbindet.

Appendix 2: Action Segment Graphs

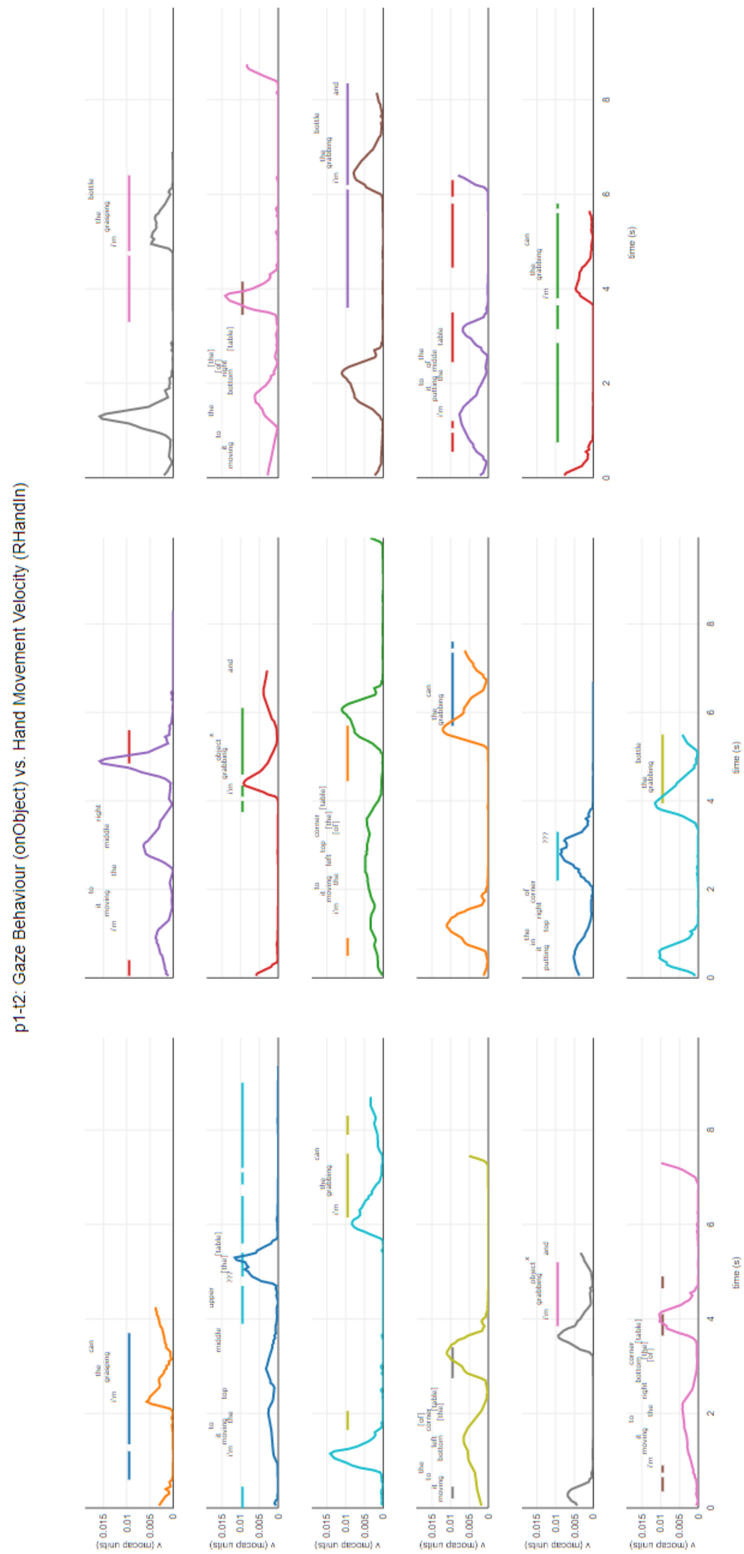

Figure 9.1: The second trial of participant p1-t1, added for comparison to Fig. 6.3. Notice the similar decomposition into TAKE AND PUT ACTIONS.

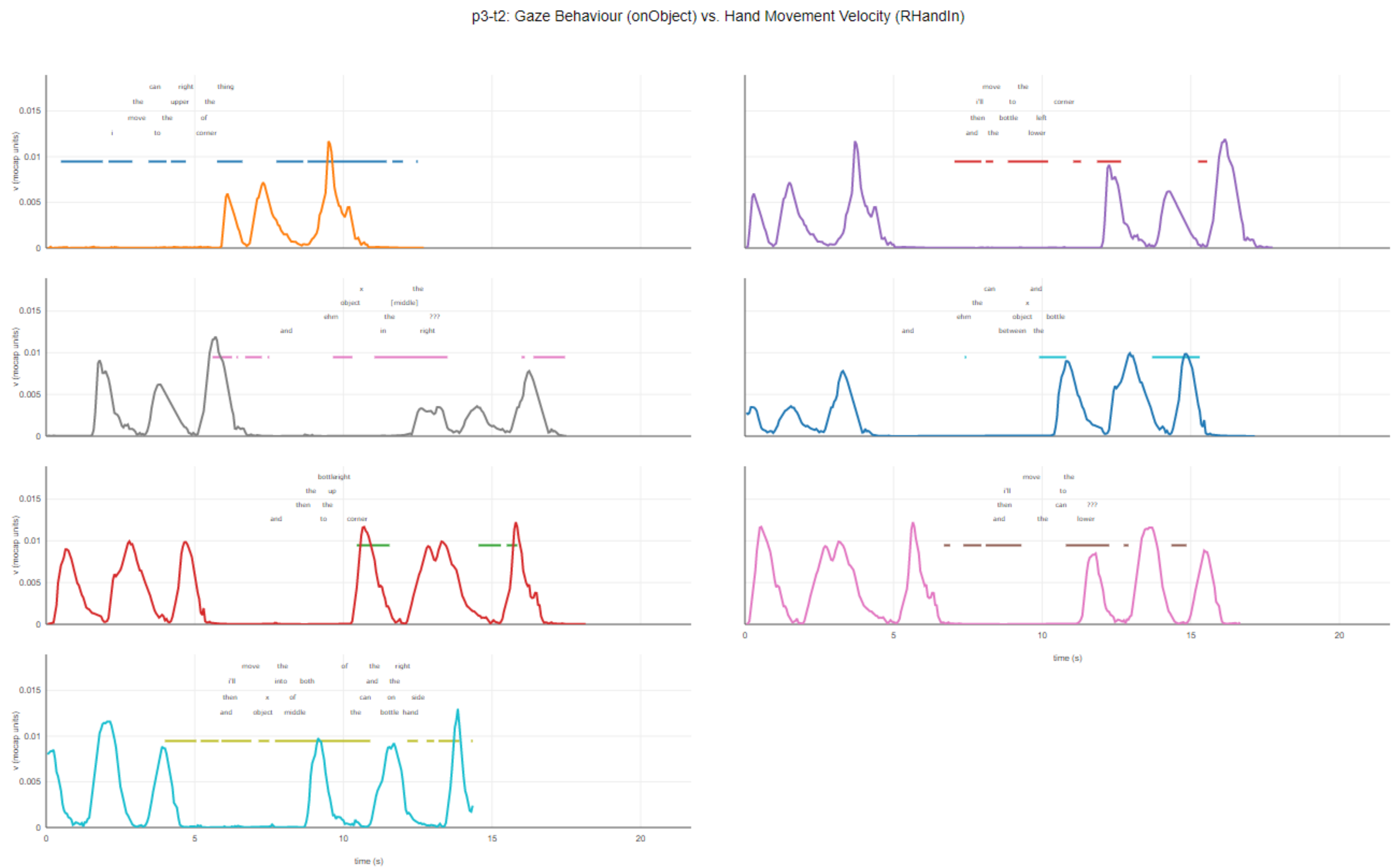

Figure 9.2: A representative trial for a very regular pattern across TAKE AND PUT ACTIONS.

p5-t2: Gaze Behaviour (onObject) vs. Hand Movement Velocity (RHandIn)

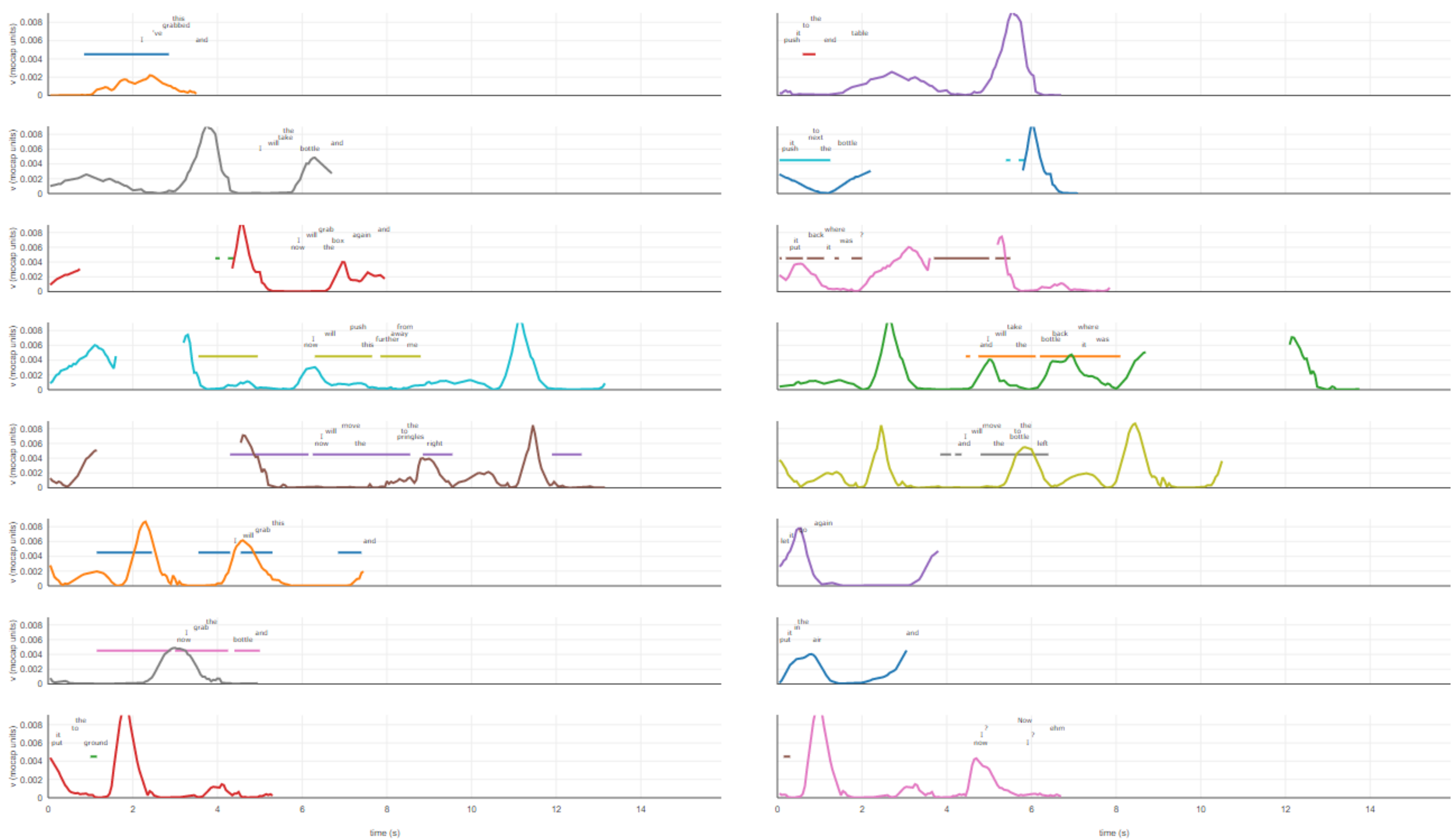

Figure 9.3: A trial indicating the effect of gaps in the motion capture recordings, were data lost during recording proved irretrievable.

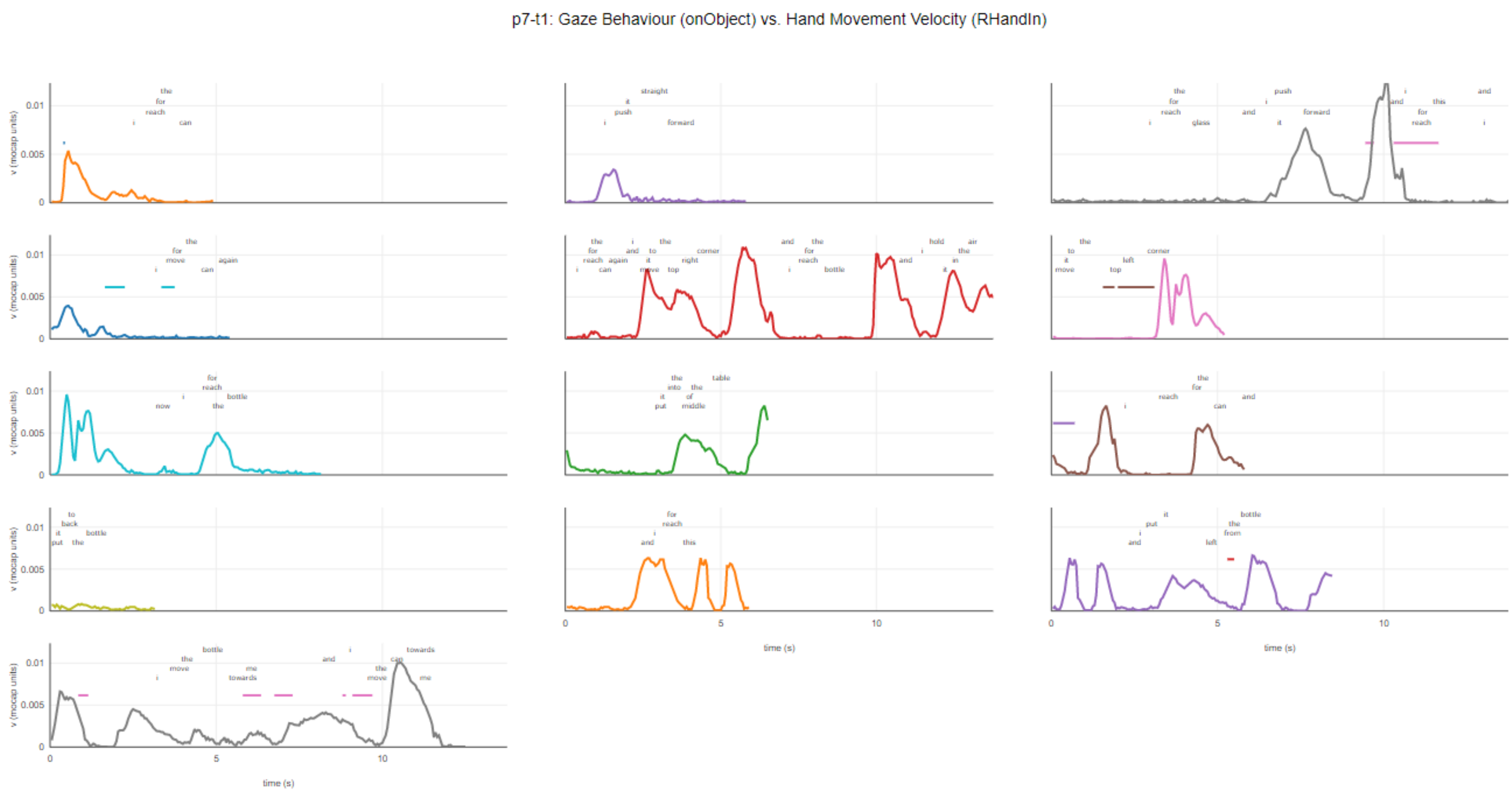

Figure 9.4: Trial of a participant with unusually long utterances.

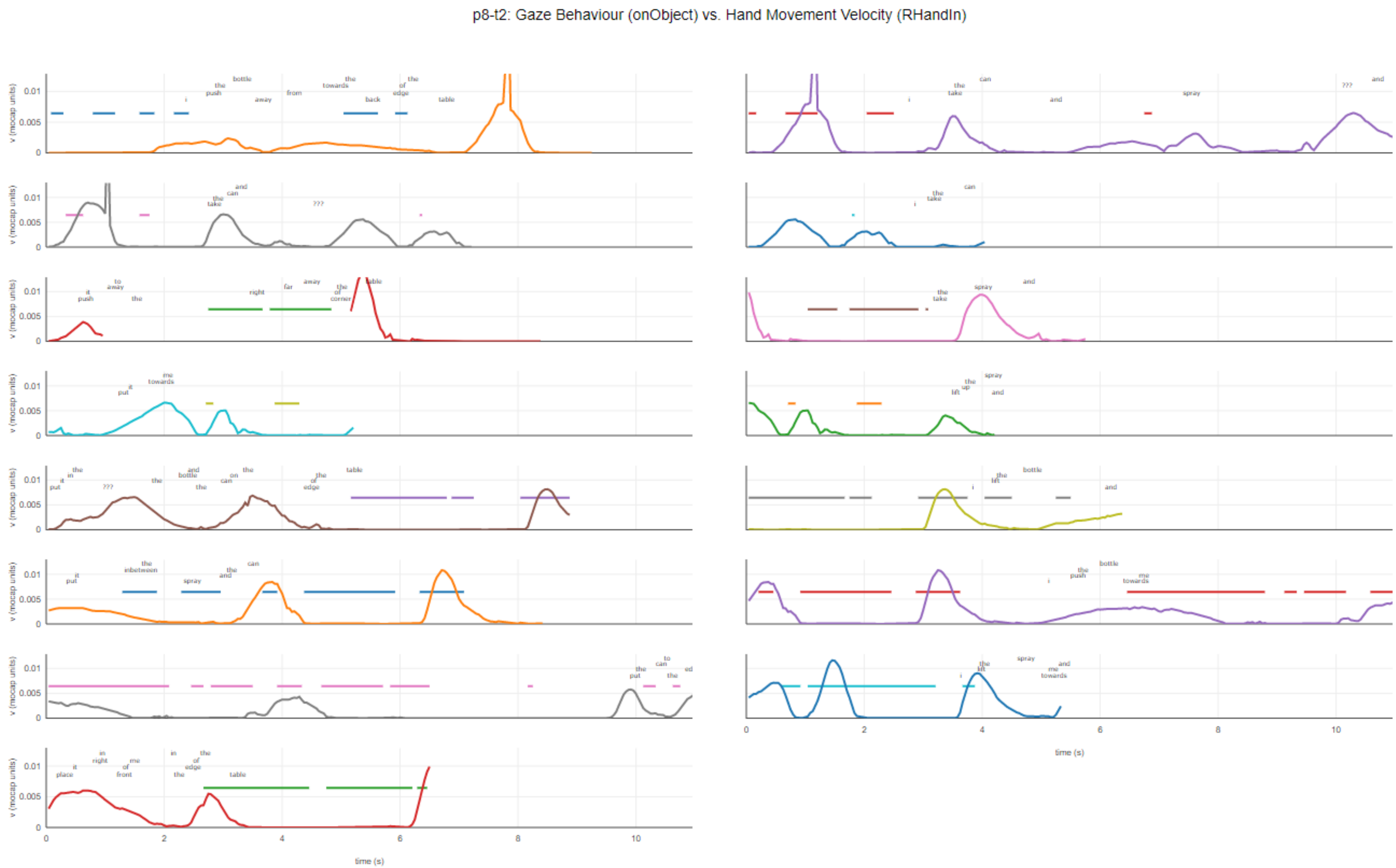

Figure 9.5: Different example of a more consistent pattern in actions and utterances.

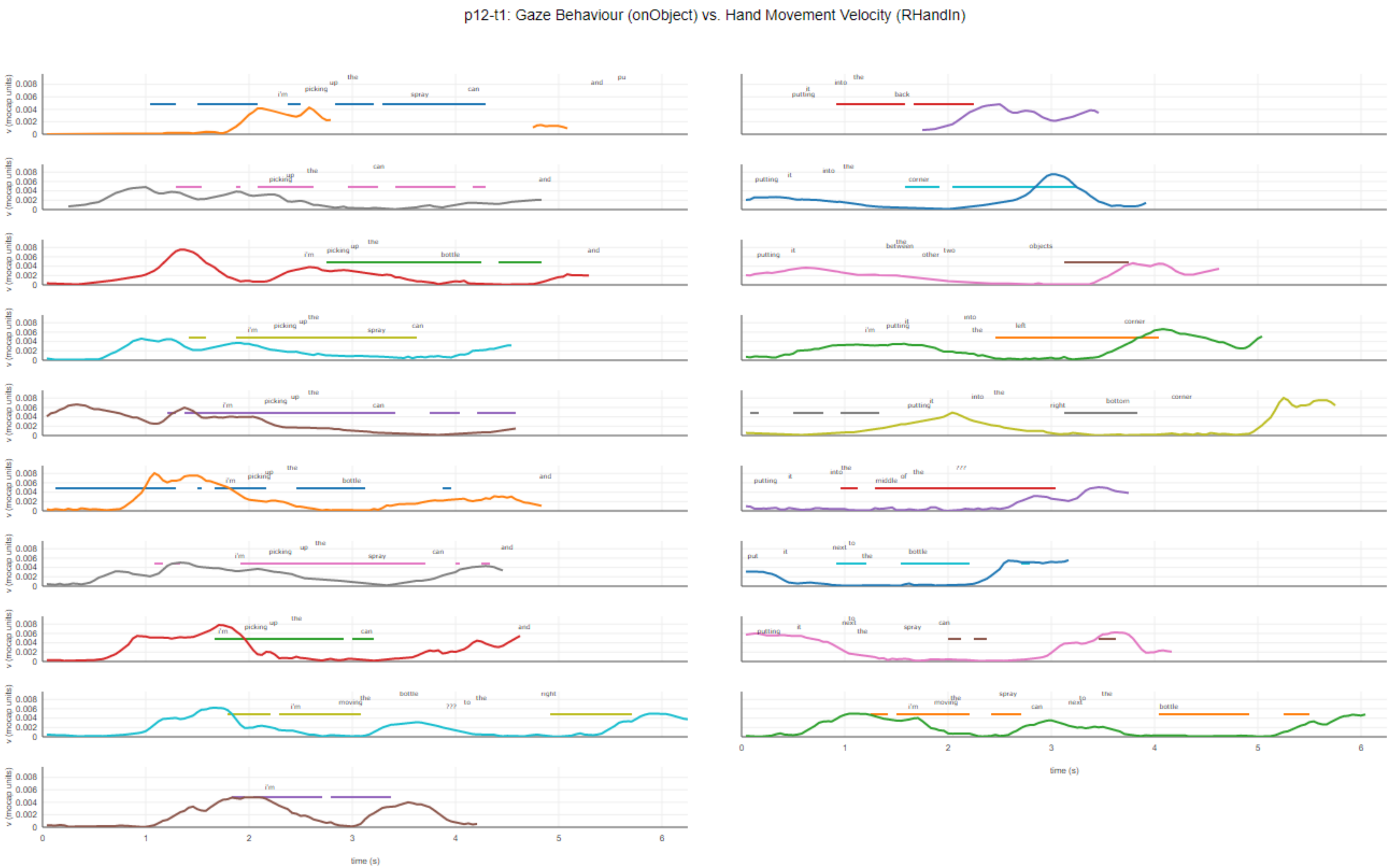

Figure 9.6: A trial with differences in episode lengths and gaze behaviour, but broadly similar movement dynamics.

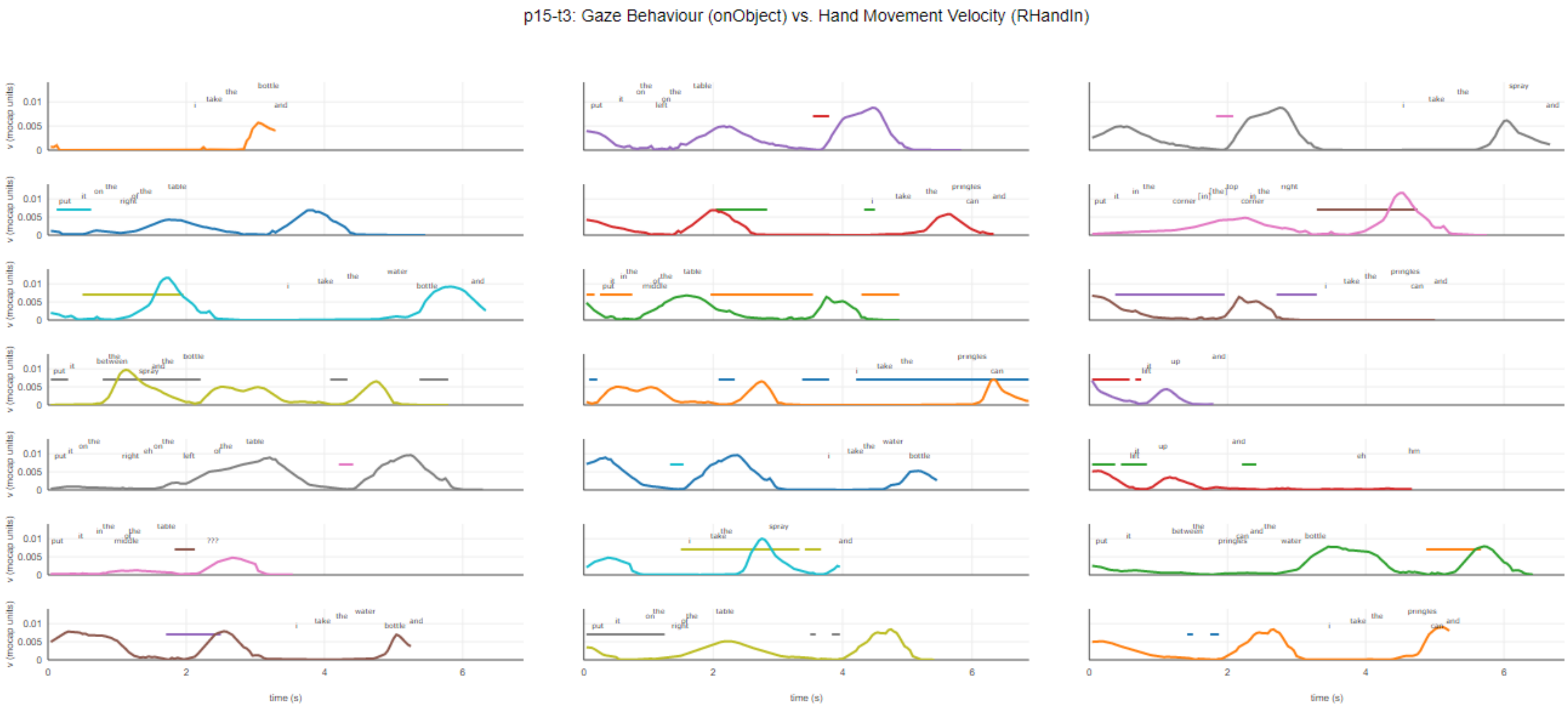

Figure 9.7: A different trial characterised by a mostly consistent pattern in aspects of movement and gaze behaviour.

TIM REINBOTH

COGNITIVE SCIENTIST

tim.reinboth@univie.ac.at, (+43) 06702043204

PUBLICATIONS

(CO-)AUTHOR

Reinboth, T., Gross, S., Bishop, L. and Krenn, B. (2020) Linguistic, Kinematic and Gaze Information in Task Descriptions : The LKG-Corpus. *Proccedings of the 12th International Conference on Language Resources and Evaluation (2020)*, Marseille, France.

Reinboth, T. (2018) Labour, Its Coercion, and How Cognitive Science Has Engaged with Them. In Hochenauer, P., Rötzer, K., Schreiber, C. Zimmermann, E. and Farkaš, I. (eds.) *Proceedings of the MEi:CogSci Conference 2018*, Bratislava, Slovakia. Comenius University in Bratislava, Slovakia.

Reinboth, T. (2017) Investigating the Structuring Role of Multimodality in Adult Verbal Discourse. In Hochenauer, P., Rötzer, K., Schreiber, C. Zimmermann, E. and Farkaš, I. (eds.) *Proceedings of the MEi:CogSci Conference 2017*, Budapest, Hungary. Comenius University in Bratislava, Slovakia.

Reinboth, T (2016) Influence of Peer Recommendation on the Neural Dynamics of Preference-Based Decision Making. *MMU Psychology Journal (Dissertations) UK* (unpublished) *URL:* http://e-space.mmu.ac.uk/id/eprint/617762 [accessed 26/04/2020]

(CO-)EDITOR

Geschwandtner, L., Jeay-Bizot, L., Makai, A., Reinboth, T., Zimmermann, E. and Farkaš, I. (eds.) Proceedings of the MEi:CogSci Conference 2020, Vienna, Austria(online). Comenius University in Bratislava, Slovakia.

Hochenauer, P., Rötzer, K., Reinboth, T., Tacha, J. and Farkaš, I. (eds.) Proceedings of the MEi:CogSci Conference 2019, Ljubljana, Slovenia. Comenius University in Bratislava, Slovakia.

EXPERIENCE

Austrian Research Institute for Artificial Intelligence (OFAI)

Vienna, Austria
10.2019 – present

RESEARCHER

CoBot Studio – Austrian Research Promotion Agency (FFG) (Project 872590)

Data collection | Annotation | Data analysis | Programming | Study design

UNIVERSITY OF VIENNA

Vienna, Austria
09.2019 – present

PROGRAM MANAGER

Co-coordinator of Middle European Interdisciplinary Master's in Cognitive Science (MEi:CogSci)

SOCIAL, COGNITIVE AND AFFECTIVE NEUROSCIENCE UNIT (SCAN)

University of Vienna
Vienna, Austria
03.2018 – 07.2019

SCIENTIFIC PROJECT STAFF

- Effects of Acute Stress on Social Behaviour
- Emotion Ego/Centric Bias Over the Lifespan

Project Lead: Prof. Dr. C. Lamm

Data Analysis | MRI Operator | Lead Experimenter | Study Piloting

UNIVERSITY OF VIENNA

Vienna, Austria
03.2020 – 08.2020
03.2019 – 08.2019

TUTOR

Master's Seminar in Cognitive Science

Feedback to students on their master's thesis concepts	Workshop on feedback and the peer review process	Implementation of digital seminar in response to COVID-19

NEUROPSYCHO-PHARMACOLOGY & BIOPSYCHOLOGY RESEARCH UNIT (NBU)

University of Vienna
Vienna, Austria
03.2017 – 03.2018

SCIENTIFIC PROJECT STAFF

- Neurobiology of Beliefs in Social Interaction

Project Lead: Prof. Dr. C. Lamm

Data Analysis	MRI Operator	Lead Experimenter

EDUCATION

UNIVERSITY OF VIENNA

Austria, Vienna

10.2020 – ongoing

MASTER OF ARTS IN SCIENCE, TECHNOLOGY AND SOCIETY (STS)

20 ECTS

UNIVERSITY OF VIENNA

Austria, Vienna
10.2016 – 09.2020

MASTER OF SCIENCE IN COGNITIVE SCIENCE

Thesis: Theoretical and Corpus-Based Investigation of the Relation Between Sensorimotor Processes and Linguistic Structure

UNIVERSITY OF GLASGOW

Glasgow, Scotland
09.2012 – 06.2016

BACHELOR OF SOCIAL SCIENCE WITH HONOURS IN PSYCHOLOGY

First Class Honours With Distinction
Callum Neil Airth Award for Best Dissertation

INTERNATIONAL SCHOOL OF STUTTGART

Stuttgart, Germany
graduated 06.2011

INTERNATIONAL BACCALAUREATE

Higher Level: Maths, Physics, History
Standard Level: English, German, French

SKILLS

SOFTWARE SKILLS

Basic Skills in easytranscript, ELAN, LaTeX, MATLAB, python, PsychoPy, RStudio, Tensorflow

TECHNICAL SKILLS

MRT Qualified User (Siemens Magnetom Skyra) at University of Vienna

Tim Reinboth, Vienna – 26/04/2020