



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„The Concept of Social Interaction in Cognitive Science:
The Example of Power and Related Affect in the Social
Interaction with Intelligent Tutoring Systems”

verfasst von / submitted by

Katharina Roetzer, BA

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

Wien, 2022 / Vienna 2022

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

UA 066 013

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Joint Degree Programme MEi:CogSci Cognitive
Science UG2002

Betreut von / Supervisor:

Dipl.-Ing. Dr. techn. Paolo Petta

Mitbetreut von / Co-Supervisor:

–

Page intentionally left blank.

Declaration of Originality

I declare that this master's thesis has been composed by myself and confirm that the work contained herein is my own, except where work of others has been explicitly indicated by referencing. This work has not been submitted, in whole or in part, for any other degree or professional qualification.

Parts of this work (the research design reported in part II) have been presented at the International Conference on Intelligent Tutoring Systems 2016 (Zagreb, Croatia, 07.-10.06.2016) as a poster and have been published as an extended abstract in [Roetzer, K.: Social Interaction with Intelligent Tutoring Systems: An Investigation of Power and Related Affect. In: Micarelli, A., et al. (Eds.): Intelligent Tutoring Systems. 13th International Conference, ITS 2016, Zagreb, Croatia, June 7–10, 2016 Proceedings, LNCS, vol. 9684, pp. 521–522. Springer International Publishing, Switzerland (2016). DOI: 10.1007/978-3-319-39583-8].

Furthermore, parts concerning the turns, shifts, and particular conceptions in cognitive science(s) in chapters 2 and 6 (in regard to conceptions and their fruitfulness) have been presented at the VDS CoBeNe PhD Academy 2022 (University of Vienna, Vienna, Austria, 02.-04.02.2022) as a poster.

Acknowledgements

I thank the University of Memphis, in particular the team of the Institute for Intelligent Systems and the staff of the Educational Testing Service, for providing and assisting me with (CSAL) AutoTutor/ASAT, as well as my supervisor, Paolo Petta, for his continuous support and his critical remarks.

Page intentionally left blank.

Table of Contents

1 Introduction.....	1
1.1 Personal Motivation and Background: Why this Thesis?.....	1
1.2 Thesis Topic, Aim, and Scope.....	4
1.2.1 Research Questions, Approach, and Methods.....	5
1.2.2 Thesis Structure and Overview.....	9
Part I Conceptions of Social Interaction in Cognitive Science.....	11
2 Turns and Shifts in Cognitive Science(s).....	11
2.1 Individual and Relational Social Cognition: Methodological Individualism, Body-Social Problem, and Human-Centredness.....	16
2.1.1 Social Interaction and (Some) Related Concepts in the Cognitive Sciences.....	21
2.1.1.1 Philosophy of Mind and Phenomenology: Social and Interactive Mind.....	21
2.1.1.2 Neuroscience and Psychology: Social and Interactive Brains.....	24
2.1.1.3 AI and Computer Science: Social and Interactive Artificial Agents.....	27
3 Social Interaction across Cognitive Science Paradigms.....	33
3.1 Why Investigate Social Interaction on the Level of Cognitive Science Paradigms?.....	33
3.2 Cognitivism and Computational Approaches to Cognition.....	33
3.3 Symbol Processing Approach to Cognition.....	34
3.3.1 (Social) Action and Interaction within the Symbol-Processing Approach.....	36
3.4 Connectionism, Parallel Distributed Processing, and Artificial Neural Networks.....	37
3.4.1 (Social) Action and Interaction in Connectionism.....	38
3.5 A Note on Social Interaction in Computational Paradigms.....	38
3.5.1 Cognitive Architectures and the Implementation of Computational Models in Artificial Agents.....	39
3.6 Dynamical Systems Perspective.....	40
3.6.1 (Social) Action and Interaction in Dynamicism.....	43
3.7 Situated Cognition and 4E-Approaches to Cognition.....	43
3.8 Embodied Cognition.....	47
3.8.1 Variants of Embodied Cognition.....	47
3.8.2 (Social) Action and Interaction in Embodied Cognition: Social Embodiment.....	52
3.8.2.1 Social Embodiment Effects.....	53
3.8.2.2 Metaphors for Social Interaction and Cognition.....	53
3.8.2.3 Four Fundamental Functions of the Body in Social Interaction and Cognition.....	55
3.9 Embedded Cognition.....	58
3.9.1 Embedded vs Embodied vs Extended.....	59
3.9.2 (Social) Action and Interaction in Embedded Cognition.....	60
3.9.2.1 Scaffolding.....	61
3.9.2.2 Social Scaffolding.....	62
3.9.2.3 Affordances.....	63
3.9.2.4 Social Affordances.....	64
3.10 Extended Cognition.....	65
3.10.1 Criteria for Cognitive Extension.....	66
3.10.2 Extended Cognition as a Constitution Thesis.....	67
3.10.3 Alternatives to extension as functional parity.....	69
3.10.4 (Social) Action and Interaction in Extended Cognition: Socially Extended Cognition.....	71

3.10.4.1 Emergent Extended Cognition.....	73
3.11 Distributed Cognition.....	76
3.11.1 (Social) Action and Interaction in Distributed Cognition.....	77
3.11.1.1 Distributed versus Extended versus Embedded.....	79
3.12 Enactivism.....	83
3.12.1 Three Variants of Enactivism.....	86
3.12.1.1 Autopoietic Enactivism.....	87
3.12.1.2 Sensorimotor Enactivism.....	88
3.12.1.3 Radical Enactivism.....	90
3.12.2 Enactivism and other E-approaches.....	91
3.12.2.1 Enactivism as Embedded, Embodied, and (sometimes) Extended.....	92
3.12.2.2 Enactivism as an Independent Theoretical Approach.....	96
3.12.3 (Social) Action and Interaction in Enactivism.....	97
3.12.3.1 Participatory Sense-Making.....	101
Part II Social Interaction with Intelligent Tutoring Systems.....	106
4 Social Interaction with Intelligent Tutoring Systems and the Role of Power and Related Affect – Proposal for a Research Design.....	106
4.1 What are Intelligent Tutoring Systems?.....	107
4.1.1 Brief Historical Background of ITS.....	107
4.1.2 General ITS Types, Architectures, and Behaviours.....	109
4.1.2.1 ITS Types.....	109
4.1.2.2 Four Component Architecture.....	112
4.1.2.3 Two-Loop Framework.....	115
4.1.3 Dialogue-Based ITS incorporating Conversational Agents.....	117
4.1.3.1 Embodied Pedagogical Agents as Human-Like Tutors.....	118
4.1.3.2 Instructional Roles of Pedagogical Agents.....	118
4.1.3.3 Computers as Social Actors and Social Conversational Agents.....	120
4.2 Research Design for Investigating the Role of Power and Related Affect in the Social Interaction with ITS.....	123
4.2.1 Social Interaction with Intelligent Tutoring Systems: Why Investigate Power and Related Affect?.....	125
4.2.1.1 Defining Power.....	127
4.2.1.2 Defining Affect.....	129
4.2.1.3 Power and Affect.....	132
4.3 Methodology.....	135
4.3.1 How to Investigate Power and Related Affect?.....	135
4.3.2 Research Setting: Tutoring Session Design.....	137
4.3.2.1 AutoTutor.....	138
4.3.2.2 Learning Content and Curriculum.....	145
4.3.2.3 Tutoring Session Structure.....	147
4.3.3 Participants.....	152
4.3.4 Means of Data Collection and Analysis: Grounded Theory Methodologies.....	153
Part III Bridging Part I and II: Situated Cognition Perspectives on the Social Interaction with ITS.....	165
5 Discussion across Part I and II.....	165
5.1 Human Learner-ITS Interaction as Social Cognition: Social Actors, Cognitive Agents, or Socio- Cognitive System?.....	165

5.2 Social Interaction with ITS from an Embodied Cognition Perspective.....	166
5.2.1 Social Embodiment Effects.....	167
5.2.2 Socially Embodied Functions and Actions.....	170
5.3 Social Interaction with ITS from an Embedded Cognition Perspective.....	174
5.3.1 Social Scaffolding.....	174
5.3.2 Social Affordances.....	177
5.4 Social Interaction with ITS from an Extended Cognition Perspective.....	179
5.4.1 Socially Extended Cognition.....	180
5.4.2 Social Interaction as Emergent Extended Cognition.....	184
5.5 Social Interaction with ITS from a Distributed Cognition Perspective.....	187
5.5.1 Distributed Cognition ('classical' variant).....	188
5.5.2 Distributed Cognition (Third-Wave Externalism).....	190
5.6 Social Interaction with ITS from an Enactive Cognition Perspective.....	191
5.6.1 Self-Other/Social Contingencies.....	192
5.6.2 Participatory Sense-Making.....	193
6 Conclusions across Part I and II.....	196
6.1 Limitations and Future Work.....	200
7 References.....	204
8 Appendix.....	227
8.1 List of Figures.....	227
8.2 List of Tables.....	227
8.3 Abstract (English).....	228
8.4 Abstract (German).....	229

1 Introduction

1.1 Personal Motivation and Background: Why this Thesis?

During my studies in the MEi:CogSci, I had a strong interest in social cognition from the start, likely due to my bachelor studies in sociology. This interest only grew in my first two semesters of the programme. At that time, my view on cognition and my interest in social cognition were based on the assumption that cognition is always dependent on and shaped by socio-cultural factors, and that it has to be investigated and analysed in socio-cultural contexts. Put simply, anything ‘social’, to me, had to involve interaction; either between two human social actors, or between a social actor and another non-human entity that can be assumed to have agency and qualify as a social (inter)actant. The unit of investigation therefore has to be a minimum of *two* individuals and/or non-human entities that (intentionally) act towards each other (‘micro-level’ perspective). In this sense, the human actor is a *social entity*. In cognitive science, however, it seemed to be the case that the human actor (which was referred to as cogniser, cognitive system, or cognitive agent), is conceptualised as a (detached) *mental entity*. Even if involved in social situations, the unit of investigation was this *one* entity, i.e. one individual or even only the brain. While in sociology anything ‘social’ is a *relational* phenomenon that has to include at least two individuals, in cognitive science ‘social’ could be *individual*. In my understanding, sociology generally focuses on the ‘outer’ of the social, and conceptualised a social phenomenon as something that may result in individual actions or behaviour, but is influenced or even caused by the socio-cultural context (e.g., social rules, norms, expectations, structures). Cognitive science, on the other hand, appeared to focus on the ‘inner’ aspects of social phenomena, i.e. cognitive and/or brain processes and mechanisms that determine the individual’s (social) actions and behaviours towards the environment and others. While a sociological perspective appears to over-emphasise the relational dimension and generally tends to neglect the individual (which I had not realised before starting MEi:CogSci), cognitive science seemed to neglect relational and interactional aspects and to focus on the individual. I found this circumstance intriguing and started to focus on it for the rest of my studies in the MEi:CogSci.

Additionally, the MEi:CogSci sparked an interest in AI and – although I had not planned on doing so when I started the programme – it was clear to me that my master’s thesis had to revolve around social interaction, social cognition, and AI. I had already gotten interested in human-com-

puter interaction (HCI) during my bachelor studies, but had looked at it from an interface and interaction design perspective. Learning the basics of AI opened up a whole new branch of interests. Specifically, I got interested in embodied agents and educational technologies, and, more concretely, Intelligent Tutoring Systems (ITS).

When I started to properly work on this master's thesis in 2016, I had a very different understanding of cognitive science, social cognition, and social interaction – or anything 'social', for that matter – than when I continued working on it in mid/late 2019. When I finished it (after yet another break during the COVID-19 pandemic) in 2022, again, my notion of these terms had changed.

Initially, I had a rather clear and narrow understanding of those terms and was sure about what I planned to do for my master's thesis. My reasoning appeared straight forward. Learning, including the special case of tutoring, is a kind of cognition. Learning is also social, thus, it should be considered social cognition. ITS are designed to promote social aspects of learning and to interact socially with a human learner. This raises questions like: Does the human learner recognise the ITS-agent as an agent/interaction partner? How is the agent/ITS perceived and how does the human learner (inter-)act based on the perception of this ITS-agent? Is the interaction with ITS a case of social cognition? When asking these questions, I was strongly influenced by my bachelor studies in sociology with regard to what 'social' means. I projected this understanding onto the phenomenon of (social) cognition. I automatically and without much reflection incorporated my sociology knowledge and method skills into this topic and used social interaction as a kind of proxy for social cognition. My thinking was based on the assumption that the social science understanding of 'social' was the 'correct' one, and that this notion had to be 'imported' into my work in cognitive science. From a sociological perspective, any kind of social interaction includes aspects of power, as well as affect. From a cognitive science perspective, it seemed that affect was a relevant aspect with regard to social cognition as well, but, interestingly, the phenomenon of power appeared to be under-represented in cognitive science research. Hence, I started off the thesis with the topic of social interaction with ITS, with a focus on the role of power and related affect in this social interaction. This work is reported on in Part II of this master's thesis.

After taking a break from working on my thesis and, literally, gaining some distance, I had to resolve two issues. (1) After taking an almost 4-year break, the ITS software (ASAT/AutoTutor) I had initially worked with was no longer compatible with the current versions. Although the University of Memphis was immensely supportive and offered to provide me with alternative versions of AutoTutor or already finished tutoring sessions they had recently developed to use for my thesis, there were time and ECTS constraints to consider. These constraints did not allow starting over. I was not sure how or if I should continue the empirical work with new versions and/or currently available tutoring sessions from AutoTutor.

(2) After consulting my supervisor regarding (1), he pointed out that finding ways to finish what I had originally planned was not the only option available; I could in fact write a thesis that has two distinct parts. One part would be a reporting of what I had done in 2016, the other could be theoretical. Furthermore, and more importantly, he prompted me to think about what ‘cognition’ and ‘cognitive science’ means to me, as well as how this relates to my initial thesis topic. He also kindly reminded me that I was supposed to write a master’s thesis in cognitive science, and not social science, sociology of technology, or human-computer interaction. Reflecting, I became aware that I had assumed that my thesis topic *was* situated in cognitive science. My reasoning was that the field of ITS is a sub-field of AI that draws upon theories, models, and methods from psychology, cognitive science, educational sciences, and social sciences. If I was to work on the topic of ITS, created a research setting, i.e. tutoring sessions, which involved programming and curriculum design (i.e., computer science and educational sciences, respectively), and applied sociological methods to perform investigations in this setting, then this is an interdisciplinary approach, therefore, it must be cognitive science. However, while the topic may be viewed as being situated in cognitive science, my perspective was not; I had adopted a (mainly) sociological perspective. The way I framed, understood, and approached my thesis topic was informed by sociology, not cognitive science. Formulated in an extreme way: I was doing sociological research on a phenomenon that happened to be part of a cognitive science sub-discipline. I then realised that I actually did not know what exactly ‘social’ means from a cognitive science perspective, what ‘social interaction’ means in cognitive science, if similar notions and conceptions exist, or whether they were relevant concepts in current research. From there on, it got only ‘worse’. I realised that I wasn’t clear on what ‘cognition’ means, or what it *actually* means for cognition to be computational, brain-bound, embodied, embedded, extended, or enactive. Therefore, I did not know what social cognition means exactly, and thus could not derive a conception of social interaction and its role in and for social cognition. This realisation resulted in Part I of the present master’s thesis.

Part I comprises conceptual work meant to address my lack of knowledge regarding the just mentioned concepts, and is an attempt to systematically work on those topics.

Now, that the thesis is finished, I am still not 100% clear on the mentioned terms and concepts, but have come to understand that their definition and meaning is also *not* clear within cognitive science. I now understand that a plurality of context-dependent – i.e. dependent on a specific paradigm and/or discipline – conceptualisations and notions exist within cognitive science. ‘Navigating’ them in order to apply them to a specific phenomenon means to recognise when they are compatible, complementary, or contradictory. In my case, the phenomenon those conceptualisations and notions were to be applied to was the social interaction with ITS, which is discussed in Part III of this thesis.

Furthermore, I now consider work to be a master’s thesis in cognitive science (and not social science, sociology of technology, or human-computer interaction). Admittedly, I still perceive myself as somewhere in-between social science and cognitive science, but have definitely gained a profound understanding of what it means to practice and adopt a cognitive science perspective.

1.2 Thesis Topic, Aim, and Scope

This master’s thesis is concerned with the phenomenon of social interaction and aims to investigate conceptualisations of social interaction within cognitive science.

It furthermore investigates a specific case of human-computer interaction as a social kind of social interaction; the interaction between a human learner and a type of educational technology called ‘Intelligent Tutoring System’ (ITS). ITS are computer-based learning environments that include computational modelling and research insights stemming from educational science, artificial intelligence and computational linguistics, psychology, and other disciplines (Graesser, Conley, et al., 2012a). ITS are designed to track the learner's psychological states, such as subject matter knowledge, skills, strategies, motivation, emotions, or other attributes, and provide adaptive instructions (ibid). Most ITS also incorporate animated software agents that communicate with the learner in natural language (embodied conversational agents) and perform tutoring along specific instructional roles (Baylor, Amy L, Kim, 2005; Graesser, Conley, et al., 2012a). A concrete research design for investigating human-ITS interaction – i.e. with the ITS AutoTutor – as a kind of social interaction is proposed. This research design is informed by social science theory and is particularly concerned with the dimensions of power and related affect in social interac-

tions.

Additionally, conceptions of social interaction from a cognitive science perspective – specifically, situated cognition paradigms – are discussed with respect to the social interaction with ITS.

1.2.1 Research Questions, Approach, and Methods

Three research questions (RQ) formed the basis of this master’s thesis.

- RQ1: What conceptions of social interaction currently exist in classical/computational and situated/4E cognitive science paradigms?
- RQ2: How do power and related affect influence and shape the social interaction with ITS regarding: i) social roles; ii) interaction and the (situational) structures it is embedded in; and iii) social relations?
- RQ3: Which of the conceptions from RQ1 are relevant for and can be fruitfully applied to the case of social interaction with ITS?

RQ2 was formulated already in 2016, RQ2 and RQ3 were included in 2019/20.

The approach employed to address *RQ1* is a conceptual one and comprises a systematic review and discussion of the literature in the sense of conceptual discourse. This conceptual work is understood as a theoretical, i.e. non-empirical, approach that does not aim at constructing new theories, but rather means “to bridge existing theories in interesting ways, link work across disciplines, provide multi-level insights, and broaden the scope of our thinking” (Gilson & Goldberg, 2015, S. 128). Concretely, RQ1 addresses conceptions of social interaction across classical cognitive paradigms (i.e. computationalism/cognitivism, comprising symbol processing approaches and connectionism), as well as situated or 4E approaches to cognition (i.e. embodied, embedded, extended, distributed, and enactive cognition).

The initial definition of social interaction used in this master’s thesis, i.e. in Part II and when starting Part I, is oriented on work on social action by the sociologist Max Weber (1978). This definition has also been used to model social actions and interactions among socially situated artificial agents (e.g. in (Barnes et al., 2019)). Additionally, this notion has been extended to explicitly include non-human entities or agents (e.g. as in the work of (Cerulo, 2009, 2011; Clarke, 2018; L. Suchman et al., 2002; L. Suchman, 2006)).

In the context of this master’s thesis:

Any interaction between at least two social agents – which can be two humans and/or non-human entities that can be assumed to have agency and/or qualify

as a social (inter)actant – who

- 1) intentionally orient their actions towards each other, i.e. they show reciprocal actions and reactions, during which
 - 2) the actions of the one influence the actions of the other, and
 - 3) these actions of the other are considered throughout the course of the interaction (in contrast to automatic or unintentional social behaviours),
- is a *social interaction*.

Note, however, that the focus of the present master’s thesis lies on instances of social interaction as dyadic interaction, i.e. micro-level interactions between *exactly* two agents.

RQ2 addresses social interaction as a specific case of dyadic interaction involving a human learner and an ITS. For the purposes of this thesis, the conversational ITS AutoTutor (see chapter 4.3.2.1) – that also features a conversational embodied agent as a tutor – is used to create three consecutive tutoring sessions about the history of the University of Vienna (see chapter 4.3.2.2). These sessions comprise the research setting. A concrete research design is proposed (see part 2, chapter 4.3), which employs the qualitative empirical methodology of Grounded Theory (GT). GT is a qualitative methodology for systematically collecting, analysing, synthesising, and conceptualising qualitative data for constructing mid-range theories that are ‘grounded’ in those data (Charmaz, 2001a, 2006, 2008). Specifically, Constructivist Grounded Theory, (CGT, (Charmaz, 2006)) and Situated Analysis (SA, (Clarke, 2018; Clarke et al., 2015)) are proposed as a methodological framework for data analysis, as well as collection (i.e. observation of tutoring sessions and associated interviews with the participants, see chapter 4.3.3 and 4.3.4 for details).

In addition to the definition provided with regard to RQ1, RQ2 furthermore assumes that social interaction always involves aspects of power and affect to a certain degree, i.e. any interaction that is of a social kind involves power and affect. Within the context of the present master’s thesis *power* is broadly conceived of as a relation and social circumstance; it is a configurational property of social relations and structures (see chapter 4.2.11). Power only exists in relation with and to other individuals. It is present in all social relationships and shapes social interactions by enabling or closing off possibilities for certain behaviours, actions, and interactions.

Affect is, in the context of this thesis, used as an umbrella term that encompasses emotions and feelings (see chapter 4.2.1.2). Affect is considered to be empirically accessible and (at least) partially observable from the outside. Thus, affect encompasses both positive and negative affect as observable aspects of emotions and feelings in social interactions, whereby observable means either third-person observations (e.g. of emotional expression in form of facial expressions, gestures) or via subjective first-person reporting in interviews (e.g. reflection on feelings).

The formulation of the three levels of RQ2 are informed by sociological theory (e.g. work of Goffman (1956), as well as Kemper's relational reading of Goffman and Collins (Kemper, 2011)). With regard to *level i*), social actors are assumed to act out or 'play' specific roles, and to organise their actions and behaviours in accordance with the social norms, rules, or expectations that are associated with those social roles (Goffman, 1956; Kemper, 2011). Additionally, social roles are of interest because pedagogical agents are intentionally designed to follow instructional roles (see chapter 4.1.3.1, 4.1.3.2). It is assumed that humans may apply (role) expectations – and associated actions and behaviours – they would have of human tutors to these artificial agents (A. Baylor & Kim, 2005; Haake & Gulz, 2006; Y. Kim & Baylor, 2016; Nass et al., 1994a). The level of the social role is necessary to investigate what kind of actions or behaviours are (not) expected, as well as what happens – in terms of power and related affective dynamics – when such role expectations are (not) met.

Level ii) of social structure accounts for the circumstance that social action and social structure are assumed to be intertwined (Giddens, 1984, 2006). It addresses the structure in which the agents' moment-to-moment interaction is embedded (i.e. in the sense of a micro-structural approach). At this intersection of action and structure, power and related affect should structure the possibilities (or lack of possibilities) for particular actions and interactions. Relative power between two actors along interpersonal hierarchies and power dynamics during the course of tutoring, as well as how this relates to affect, might be revealed.

Finally, *level iii*) addresses social relations. Past social interaction are assumed to influence present interactions (Boiger & Mesquita, 2012; Burr, 2015). Thus, the change or development of social roles, actions, interaction (and embedded structures) is of interest. This aspect is – in the context of this thesis – referred to as social relationship in the sense of ongoing social interactions. Focusing on relations may reveal insight regarding the change and development of the social interaction over time (i.e. across the three tutoring sessions) and how power and related affective dynamics are established, changed, or (de-)stabilised.

RQ3 aims to address the conceptualisations described in respect to RQ1 and whether they can be fruitfully applied to the phenomenon addressed in RQ2. Note that the word 'fruitfully' has a specific meaning within the context of this master's thesis. Like for RQ1, the approach to RQ3 is a conceptual one, but it is additionally guided by the idea of a 'fruitful concept' (Flemmen, 2017; Sohlberg & Leiulfstrud, 2018) which stems from conceptual constructionist, social constructionist, and philosophy of science research. In a nutshell, social interaction is understood as a

concept. Concepts are the building blocks of theories, which are bundles of interconnected and interrelated concepts (ibid). Concepts are also what bridges or mediates theory and data (Bulmer, 1979; Flemmen, 2017; Merton, 1968, p. 19) in the sense that theoretical concepts inform the formulation of research questions and their operationalisations, as well as interpretations of data. However, the meaning of a specific concept is often not clear and depends on local lexical usages (e.g. in a specific discipline or research field), which result in debates about definitions and conflicts regarding the interpretations of the concept (Sohlberg & Leiulfsrud, 2018).

In conceptual discourse, a commonplace practice that is often – implicitly or explicitly – employed in philosophy of science is to assess what *the* concept is in its essence. This is called the ‘convergent essentialist approach’ (cf. (Sohlberg & Leiulfsrud, 2018)). That is, identifying all concepts and their local usages in order to systematically define the essence or common denominator of all local concepts (ibid). But:

finding a common denominator of all concepts, this exercise is not particularly informative. [...] Although these themes are basic for gaining a general understanding of conceptual constructions, they are not very useful in scientific practice [emphasis added], i.e., in the application [emphasis added] and understanding of concepts. (Sohlberg & Leiulfsrud, 2018, p. 4)

Concepts should thus not be analysed or assessed with regard to their essence, i.e. what the concepts are about and/or their truth value, but rather, they should be assessed in relation to their *fruitfulness* (Hagen, 2017; Sohlberg & Leiulfsrud, 2018). Furthermore, concepts cannot always be properly assessed or understood within their original theoretical context. Although a concept in itself does not have explanatory power, a partial de-contextualisation may be employed as a way of gaining new insights (Sohlberg & Leiulfsrud, 2018).

Within the context of the present master’s thesis, a *fruitful concept* is conceived of as a concept that may lack truth value (because it is not the truth value that is being assessed), but is, in reference to Merton (1968) and Sohlberg and Leiulfsrud (2018), generating new insights or research questions that may inform future theoretical models or empirical research. Hagen (2017) similarly points out: “Concepts that make the empirical analysis difficult are less fruitful, whereas concepts that solve problems encountered during the empirical analysis are more fruitful”(Hagen, 2017, p. 135). These notions of fruitfulness appear to be aligned with the comparative practice of conceptual work in order to link and bridge concepts across disciplines since a “comparative perspective on specific instances of the studied phenomena may uncover new insights that are otherwise concealed by their different contexts [...] it is certainly revealing insights into the potential *applications* [emphasis added] of concepts” (Sohlberg & Leiulfsrud, 2018, p. 10).(Sohlberg and Leiulfsrud 2018, p.10). The approach to RQ3 is thus a conceptual one (as in RQ1), but particu-

larly focuses on the question of fruitfulness. That is, assessing a concept's usefulness and applicability for furthering new research insights via partial de-contextualisation and de-construction. In the case of this thesis, this means to abstract the conceptions of social interaction to the level of cognitive science paradigms (which generally are concerned with human cognition and human-human interaction, see chapter 2.1) and to re-apply them to the case of human-ITS interaction in the context of social cognition. RQ3 is thus not primarily concerned with whether a conception of social interaction is 'correct' or 'true', but instead assesses concepts in terms of fruitfulness.

1.2.2 Thesis Structure and Overview

This master's thesis comprises three parts; one for each RQ. Part I corresponds to RQ1, Part II to RQ2, and Part III to RQ3.

In *part I*, RQ1 is addressed in the course of conceptual work that comprises theoretical review and literature research. First, a brief historical perspective on cognitive science, including turns and shifts within the field, is elaborated (chapter 2). Then, social cognition and some related conceptual and methodological issues are introduced (chapter 2.1). Social interaction – with regard to human-human interaction, as well as the interaction with or of artificial agents – and related concepts across cognitive science disciplines are described (chapter 2.1.1). This is followed by a conceptual investigation of social interaction across cognitive science paradigms. Classical paradigms (computational paradigms) and situated approaches (situated/4EA cognition) will be examined, whereby the focus of part I lies on situated cognition paradigms. Concretely, the classical/computational paradigms of symbol-processing (chapter 3.3) and connectionism (chapter 3.4), as well as the dynamical systems approach (chapter 3.5), and the situated approaches encompassing embodied (chapter 3.7), embedded (chapter 3.8), extended (chapter 3.9), and distributed (chapter 3.10) cognition, as well as enactivism (chapter 3.11), are described. Each chapter addressing one cognitive science paradigm encompasses a review of the core terms and concepts, as well as their meaning, of the respective paradigms. Then, theoretical considerations regarding the role of social interaction in (social) cognition – and related conceptions of social interaction – are described. Finally each chapter on a cognitive science paradigm discusses specific 'social variants' of the paradigm that contain notions of or focus on social interaction.

Part II encompasses a comprehensive reporting of the research design proposal for addressing RQ2, which was developed in 2016. First, the field of ITS (chapter 4.1), including a brief history of the field (chapter 4.1.1), as well as general ITS types (chapter 4.1.2.1), architectures (chapter 4.1.2.2), and behaviours (chapter 4.1.2.3) is described. Then, the ITS type of conversational or

dialogue-based ITS is focused on (chapter 4.1.3). AutoTutor – which is the conversational ITS that has been used within the context of this master’s thesis – is described in detail (chapter 4.3.2.1). A detailed description of the proposed research design, including a statement on the role and relevance of social interaction as a phenomenon within ITS research (chapter 4.2.1), as well as a theoretical review of social science conceptions and definition of power (chapter 4.2.1.1) and affect (chapter 4.2.1.2). Then, the methodology section reports the concrete research setting, including how AutoTutor was used to create tutoring sessions to be investigated (chapter 4.3.2.1), the concrete design and content of these sessions (chapter 4.3.2.2 and chapter 4.3.2.3), as well as methods for investigating the sessions, i.e. participants (chapter 4.3.3) and means of data collection and analysis (chapter 4.3.4).

Part III, titled “Bridging Part I and II: Situated Cognition Perspectives on the Social Interaction with ITS”, is meant to address RQ3. This comprises a discussion of social interaction with ITS from an embodied (chapter 5.2), embedded (chapter 5.3), extended (chapter 5.4), distributed (chapter 5.5), and enactive cognition (chapter 5.6) perspective. Specifically, each of the ‘social variants’ of the cognitive science paradigms and the associated conceptions of social interaction identified in part I will be systematically discussed with regard to their applicability (in the sense of fruitfulness) to the phenomenon of social interaction with ITS (part II). Finally, conclusions across part I and II are drawn (chapter 6), including limitations and future work (chapter 6.1).

Part I Conceptions of Social Interaction in Cognitive Science

2 Turns and Shifts in Cognitive Science(s)

From a historical perspective, cognitive science has undergone several shifts or turns with respect to its research programmes, resulting in different research areas, foci, as well as several competing notions of what – or where – cognition is (see figure 1).

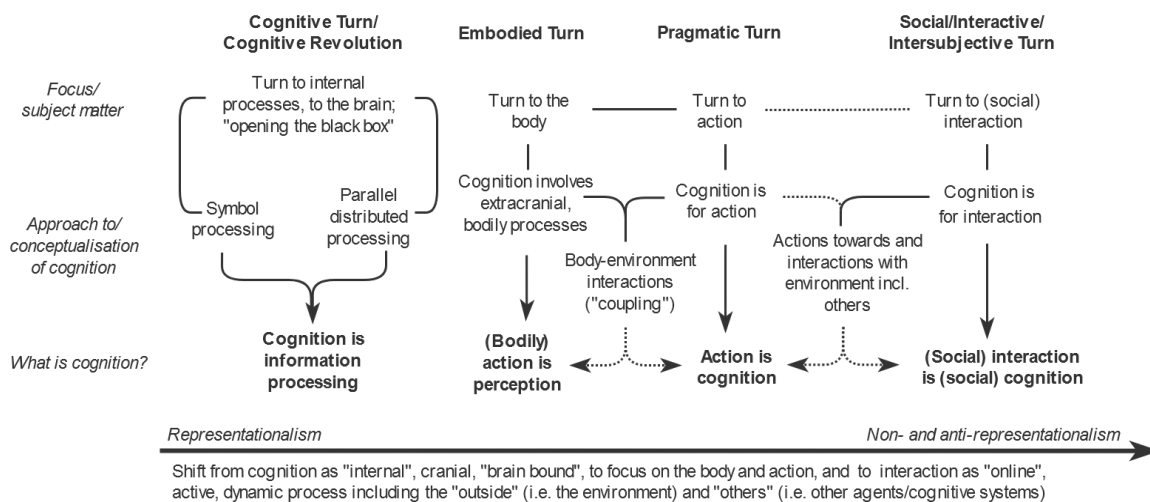


Figure 1: Simplified context map of paradigm shifts and turns in cognitive science

When cognitive science began to establish itself as a discipline, research from different areas was marked by an increased interest in phenomena such as the mind, consciousness, and cognition, which were, in preceding behaviourist research, assumed to be scientifically inaccessible. Such subjects were a ‘black box’, which cannot be studied directly, i.e. research needed to focus on directly observable behaviours (Bermúdez, 2019; Miller, 2003; Sturm & Gundlach, 2013; Thagard, 2005). In the course of the ‘*cognitive revolution*’, several researchers started to ‘open the black box’ and to shift their attention to the brain, so to speak.

In his personal account of how he experienced the formation of cognitive science, Miller (2003) states:

Cognitive science is a child of the 1950s, the product of a time when psychology, anthropology and linguistics were redefining themselves and computer science and neuroscience as disciplines were coming into existence. Psychology could not participate in the cognitive revolution until it had freed itself from behaviorism, thus restoring cognition to scientific respectability. By then, it was becoming clear in several disciplines that the solution to some of the problems depended crucially on solving problems traditionally allocated to other disciplines. Collaboration was called for [...]. (Miller, 2003, p. 141).

The ‘birth’ of cognitive science is often pinpointed to 1956 (Gardner, 1988; Miller, 2003; Sturm & Gundlach, 2013). On 10-12 September 1956, a symposium organised by the ‘Special Interest Group in Information Theory’ at the MIT, USA, took place, at which several seminal papers were presented (Miller, 2003). Allen Newen and Herbert Simon – who are today, together with Marvin Minsky and John McCarthy, considered to be central figures in establishing the discipline of AI – presented their work on the ‘Logic Theory Machine’, which was the first proof of a theorem carried out via a computing machine (Gardner, 1988). Additionally, the linguist Noam Chomsky presented his ‘Three Models of Language’, and the psychologist George Miller outlined his work on short-term memory; ‘The Magical Number Seven, Plus or Minus Two’ (Gardner, 1988; Miller, 2003). While the initial ‘contributors’ to cognitive science came from the fields of AI, psychology, and linguistics, neuroscience and anthropology too established their perspectives on the phenomenon of the mind. During the late 1950ies and 1960ies, neuroscientists started to record single neurons and its responses in the nervous system in animals, and anthropology started to study language-related cognitive practices in different cultures (Gardner, 1988). Starting in 1960 with the foundation of the Harvard Center for Cognitive Studies, and the group dedicated to information-processing psychology at Carnegie-Mellon University, several research groups, labs, and departments dedicated to cognition were formed and started to ‘institutionalise’ the study of the mind (Gardner, 1988; Miller, 2003). Due to an increased scientific interest in cognition and the newly established discipline of neuroscience, the Alfred P. Sloan Foundation established and funded several programmes to study the nervous system, including the fields of neuropsychology, neurophysiology, and neurochemistry in 1976 (Gardner, 1988; Miller, 2003). The Foundation also initiated the Sloan Special Programme in Cognitive Science, thereby finalising the ‘institutionalisation’ of cognitive science as a research field (Gardner, 1988).

Looking at those developments, it appear almost ‘logical’ that cognitive science was strongly informed by the concept of computation, as well as the investigation of the – finally accessible – brain. In the beginning of cognitive science, i.e. its establishment as a discipline, it has been driven by computational-representational and symbol-processing-based approaches to cognition intended to account for a naturalistic understanding of the mind and its processes (Bermúdez, 2019; Besold & Kühnberg, 2013; S. Walter, 2014a). In essence, the brain, and thus cognition, was understood to function in the same way as a computer, i.e. by processing information in terms of input-output relations, grounded in representations (*ibid*). Representation-based paradigms to cognition (see chapter 3.2) understood the mind and its processes as computation of mental representations within a highly modular cognitive architecture, including context-invariant pro-

cessing in subsystems (Engel et al., 2016; S. Walter, 2014a). Such computation was assumed to occur on an independent level of analysis (functionalism, see chapter 2.1, 3.2)) and this model of cognition solely included the cognitive system itself (individualism, see chapter 2.1) (Engel et al., 2016). These approaches to cognition comprise what is today generally referred to as ‘standard’, ‘classical,’ or ‘mainstream’ cognitive science, as well as ‘computationalism’, ‘cognitivism’, and ‘representationalism’ (see chapter 3.2).

The first major shift within cognitive science was caused by heavy criticism of the computational paradigm (see chapter 3.3, 3.3.1), and alternative approaches started to emerge during the 1980ies, initially in robotics, later also in cognitive psychology, philosophy, and neuroscience (ibid). A milestone indicating this shift was the book ‘The Embodied Mind’ by the philosopher and neuroscientist Francisco J. Varela, the psychologist Eleanor Rosch, and the philosopher Evan Thompson published in 1991. Embodied approaches to cognition claim that cognition is not merely a kind of information-processing machinery that receives input (perception) from the world and ‘maps’ it to a corresponding output (action) (Newen et al., 2018; Shapiro, 2014; S. Walter, 2014a), or at least that this does not only include the brain. Rather, cognition is assumed to be body-based: a brain requires a body to perceive information and produce output, like movement, and a specific body furthermore only allows for very specific kinds of perceptions and actions (ibid). The body and experiences lived through this body, as well as the environment and interactions with it contribute or even constitute cognitive processes (Varela et al., 1991; S. Walter, 2014a). Thus, cognition involves non-cranial or ‘more-than-brain’ processes, meaning the body and the environment (ibid). In a sense, cognitive science turned away from the brain and towards the body, resulting in a broader – and more complex – conception of cognition as a brain-body unit. This shift from cognition to involving (only) the brain to involving the entire body is referred to as ‘*the embodied turn*’.

Associated with the turn to the body is the ‘*turn to action*’ or ‘*pragmatic turn*’: “The pragmatic turn goes hand in hand with an embodied turn.” (Dominey et al., 2016a, p. 343). In their introductory chapter “Where’s the Action?” to the book “The Pragmatic Turn”, Engel and colleagues (Engel et al., 2016), state that cognitive science has undergone a change in approaches to cognition within the last two decades. They refer to this change towards action-oriented perspectives on cognition as ‘pragmatic shift’. This shift is characterised by an understanding of cognition as ‘enactive’, i.e. as a form of practice that serves action, which essentially means that cognition is

grounded in sensorimotor skills (ibid). Action-oriented approaches are currently still evolving in different cognitive sciences, e.g. in robotics, neuroscience, psychology, and philosophy of mind (ibid). While these approaches still lack strong conceptual ties across disciplines, the perspective of dynamical systems (see chapter 3.5) does provide a unifying language, terminology, methods, and descriptions of action-oriented approaches. Those approaches are – different to brain-centred approaches – understood in relational terms. An enactive approach conceives of cognition as being spread out in an interrelated agent-environment-system and aims to examine cognition in terms of the relations between brain, body, and environment (Dominey et al., 2016a). The key premisses of action-oriented approaches are that i) cognition is a capacity to generate structures through action and ii) that the cognitive system is immersed in its task domain, i.e. the specific environment. Meaning is acquired through a relational cognitive system in the context of action and this cognitive system is embodied and extended in its environment (Dominey et al., 2016a; Engel et al., 2016; Menary, 2016). Rejecting computationalism and representationalism entirely, a cognising agent is an agent that is actively engaged in and interacting with the environment as well as other agents in a closed loop (Dominey et al., 2016a). In pragmatic terms, cognition exists to serve action in the context of interaction with the environment and other agents (ibid).

Within debates about cognition in the fields of social cognition and philosophy of cognitive science, this focus on interaction is addressed even more explicitly as a ‘*social turn*’ (Kyselo, 2014), and ‘*interactive turn*’ or ‘*intersubjective turn*’ (De Jaegher et al., 2010a; De Jaegher, 2018) within (social) cognition research. Researchers associated with this turn argue that approaching social cognition with a focus on social interaction is needed to push the field forward and to expand scientific research beyond the individual (ibid). The social turn calls for a change in focus: away from the individual and towards social interaction and its role in and for (social) cognition. Rather than addressing cognition with a focus on the ‘gap’ between brain and body (as embodied approaches do), social and interactional approaches mean to address the gap between individual and others (Kyselo, 2014). Although the interactive/intersubjective turn is still concerned with cognition in the context of bodily actions and interaction with the environment, it stresses the relevance of interaction with others, i.e. social interaction. Its proponents (e.g. (De Jaegher, 2018; De Jaegher et al., 2010a; Froese & Gallagher, 2012; Froese & Ziemke, 2009)) argue that social cognition is not reducible to an individual mind and/or individual cognitive mechanisms. An even stronger claim associated with the intersubjective turn is that interactive processes are not a mere context for social cognition, but can in fact complement, constitute, or even replace them (De Jae-

gher et al., 2010a; Froese et al., 2015). This holds for direct and indirect instances of social cognition: “Social cognition has noninteractive aspects (e.g. remote observations of social scenes, reflection on others’ actions). But even here, social interaction is likely to play an enabling role in the processes involved” (De Jaegher et al., 2010 p. 445). In a nutshell, cognition does not only serve action, but also social interaction.

Paradigms and research programmes associated with the embodied, pragmatic, and interactive/intersubjective/social turn are generally referred to as ‘situated cognition’ (see chapter 3.6). From a situated cognition perspective, cognition cannot be examined in isolation, i.e. without considering the body and its environment. This potentially includes other cognitive entities, agents and/or systems in this environment. In contrast to classical approaches in cognitive science, in which cognition is mainly conceptualised as (neural) ‘offline’ processing of sensory input and motor output, situated cognition foremost takes place ‘online’ through continuous perception-action cycles in reciprocal interaction with the environment. Within situated cognition, the body becomes more than an ‘output vehicle’ (cf. (S. Walter, 2014a) controlled by a central processing unit (i.e. the brain). The body itself – including its biological, physiological, and morphological specifics – can act as meaningful cognitive resource (ibid). The environment can act as cognitive resource to minimise cognitive load, as cognitive systems can keep their internal cognitive effort low by accessing already available resources in the environment (Lyre & Walter, 2013; S. Walter, 2014a). Therefore, the focus of analysis in situated approaches shifts from the previously rather ‘abstract’ problems or phenomena that have been central in classical cognitive science – like planning, reasoning, or language comprehension – to more ‘pragmatic’ action-oriented phenomena happening in real-time, where cognitive systems in their specific bodily/physical constitution interact with a dynamic and changing environment, including others (Dominey et al., 2016a; Engel et al., 2016). Due to the interest in such phenomena, the primary unit of analysis for the study of cognition must change; from the ‘inner’ (and potentially inaccessible) workings of an individual and its brain to the observable ‘outside’, i.e. the body and its environment, as well as resources in this environment, including other agents or cognitive systems. It is this shift in focus towards action and (dynamic) interactions, that makes situated approaches to cognition a promising framework for the analysis of social interaction within cognitive science, and particularly social cognition (see chapter 2.1).

2.1 Individual and Relational Social Cognition: Methodological Individualism, Body-Social Problem, and Human-Centredness

Broadly conceived, social cognition is any cognitive process that involves other people (Frith & Blakemore, 2006). Note that in the literature ‘social cognition’ is sometimes used to describe a subfield in psychology, i.e. social cognition as a specific research field or area within psychology. Within the context of this master’s thesis, the term social cognition is used to describe any research that is concerned with social cognition in the sense of a phenomenon. As a phenomenon, social cognition corresponds to research that is concerned with this phenomenon, regardless of its disciplinary origin. In this sense, social cognition is investigated *not* only by psychology, but also philosophy, neuroscience, AI, and other disciplines. Research on this phenomenon, i.e. social cognition research, encompasses a wide range of topics, such as Theory of Mind (also called mentalising or mind-reading), discerning agency and self-other distinction, recognising faces, producing, recognising and interpreting affective expression, joint attention, eye-gaze detection and following, imitation, coordination, cooperation, and joint action, empathy, and others (Frith & Blakemore, 2006; Holdnack, 2019; Marchant & Frith, 2009; Vernon, 2014).

With reference to the previous chapter, particularly, the embodied, pragmatic, and intersubjective turn, three recent debates or issues regarding social cognition research appear to be of interest for the present master’s thesis: i) methodological individualism, ii) the body-social problem, and iii) human-centredness.

The term ‘*methodological individualism*’ pertains to the methodological approach of focussing on an individual as the explanatory unit within social cognition (De Jaegher & Di Paolo, 2007; S. Gallagher, 2018a, 2020). According to Gallagher (2020) methodologically individualistic approaches are driven by the following goal:

to identify mechanisms found within an individual, and specifically within an individual brain, to explain how that individual understands other people. This assumption closes off the possibility that what explains social cognition is something that involves more than cognitive processes in one individual. (S. Gallagher, 2020, p. 3)

Opponents of individualistic conceptions of social cognition suggest that the individual has to be investigated within the context of interaction with others. Employing social interaction as the analytical unit for social cognition is supposed to be a means of overcoming methodological individualism (De Jaegher et al., 2010a; Froese & Gallagher, 2012; Herschbach, 2012). This may be achieved by different means, e.g. through enactivist conceptions (cf. De Jaegher and Di Paolo 2007; Gallagher 2017, 2018), embodied conceptions (Lindblom 2020), interactionist conceptions

(Froese and Gallagher 2012), or mechanistic conceptions¹ (Herschbach, 2012) of social cognition that focus on social interaction.

This debate is in part related to an – implicit and explicit – distinction between individual social cognition or individual sense-making and ‘proper’ social cognition that involves intersubjective processes in terms of a participatory and relational interaction (i.e. conceptualisations in which social cognition is explained by the interactions between individuals and the effects this has on the individuals involved). This is also reflected in work from different cognitive science disciplines, e.g. re-conceptualisations of social cognition as a second-person cognition (de Bruin et al., 2012; Schilbach et al., 2013b) and online social cognition (Schilbach, 2014) in neuroscience and psychology, or interactional (Froese & Gallagher, 2012; S. Gallagher, 2001) and narrative (S. Gallagher, 2006; S. Gallagher & Hutto, 2008) accounts of ToM in philosophy (see next chapter for details).

In relation to the embodied and the intersubjective turn – which means to overcome methodological individualism – the philosopher Kyselo (2014) – identified an interesting conceptual problem in cognitive science: the ‘*body-social problem*’. This problem addresses conceptual discrepancies with regard to social cognition in terms of the concepts of self-other distinction, self, and identity when taking the embodied *and* the intersubjective social turn into account:

Both approaches are mutually exclusive. Without due conceptual clarification, adopting either version, i.e., a primacy of embodiment or a primacy of the social, reduces the other. The assumption that the body individuates the self while the social remains merely context puts into doubt the second disciplinary development in cognitive science, the social turn, and would re-invite accusations of methodological individualism. (Kyselo, 2014, p. 986:4)

A simplified interpretation of this issue is: if the body is focussed on, one may focus on the individual and equate the body with the self. If one instead focuses on the social, the role of the self and individual in and for social cognition may be downplayed. However, the self is assumed to be important in social cognition, since it is the individual which interacts with and distinguishes itself from others.

How body and social dimensions relate may be answered in two ways: 1) there can be a self without the social since the individual in terms of the self is equated with the living organism, i.e. the body. Although embodied and social aspects of the self are interrelated, the self also remains independent of sociality to some degree (‘social as contextual claim’, cf. (Kyselo, 2014)). Or: 2) the self as a distinguishable unity is primarily constituted not by the body, but by the social (‘so-

¹ Note that mechanistic accounts are not discussed within the context of this master’s thesis, but see (Herschbach, 2012) for a detailed account on mechanistic conceptions as a response and/or alternative to enactive conceptions.

cial constitutive claim’, cf. (ibid)). If the body is equated with the (embodied) self – in tension with a social self – then there is “nothing social about the organismic or the moving body *per se* [original emphasis]” (Kyselo, 2014, p. 986:4), or at least not as long as what counts as a body is not clarified (see chapter 3.7.1 on that matter). Kyselo (2014) – whose work is best classified as enactivist – further suggests that embodied and social approaches to social cognition – and thereby, indirectly, to the self – entail a re-conceptualisation of the *boundaries* of the individual. Embodied and social conceptualisation entail different claims about what it means to be an individual or a ‘whole’ self. Each provide different principles of individuation since: “The self as a whole can either be embodied or social, but it cannot be both”(Kyselo, 2014, p. 986:3). Embodiment entails an individuation to an autonomous individual identity (body), while the social approach pertains to a group identity of at least two individuals in the sense of an autonomy in or through interaction (ibid).

Like other pragmatic or action-oriented approaches, enactivist conceptions of (social) cognition are generally focussing on action and interaction, and do explicitly – to a greater extent than other situated approaches – conceptualise social cognition by means of social interaction. That is, enactive (or radically embodied and body-enactivistic, see chapter 3.7.1) perspectives on social cognition use social interaction – rather than the individual or its body – as an analytical and/or explanatory unit (for details on the enactive conception of social interaction and cognition, see chapter 3.11.3). Proponents of this position do, however, argue that their approach is neither interactionist, nor individualistic, and that the individual is indeed relevant in social interaction and cognition (E. Di Paolo & De Jaegher, 2017). Concerning the potential ‘dissolving’ of identity and self, and of neglecting the individual in enactivist conceptions of social cognition by over-emphasising social interactional aspects (e.g. in the concept of participatory sense-making, see chapter 3.11.3.1), De Jaegher and colleagues (De Jaegher & Di Paolo, 2013; E. Di Paolo & De Jaegher, 2017) do indeed address the role of the individual.

[...] enactivists already do not think that all that matters happens only in interaction. We criticize methodological individualism but do not thereby hold true its exact opposite (the irrelevance of individual cognition). The enactive stance attempts to supersede such a dichotomy. (De Jaegher & Di Paolo, 2013, p. 345:2).

Different approaches to overcome the body-social problem have been formulated. E.g. that the self has to be conceptualised as being socially enacted with the body mediating the self (Kyselo, 2014), or that the human body needs to be re-conceptualised as being irrevocably socially constituted, while the human social world is irrevocably bodily constituted (Higgins, 2018). Alternatively, De Jaegher and Di Paolo (2013; 2017) argue that this conceptual debate is, in part, due to a

misinterpretation of enactivist accounts of social interaction as being (purely) interactionist, which would indeed entail problematic claims regarding self, other, and interaction (E. Di Paolo & De Jaegher, 2017). An enactive perspective on social cognition:

is *neither* [original emphasis] what goes on in people's brains or in their self-affective bodies, *nor* [original emphasis] what occurs in social interaction processes—if we consider each alternative on its own. On the contrary, intersubjective phenomena emerge only as a *dynamic relation* [original emphasis] between these two [...] To see the enactive approach as defending an interactionist position is to repeat the mistakes of methodological individualism we criticize in others. (Di Paolo & De Jaegher, 2017, p. 88).

Furthermore, there may be misconceptions of the underlying framework or model employed to social interaction and cognition. That is, a dynamical systems model (see chapter 3.5). In a dynamical systems framework, social interaction (and cognition) 'comprises' individual systems/agents that are 'coupled' (see chapter 3.9.2 for the concrete meaning of this term). This coupling is co-regulated and constrained by internal processes in each of the individual systems/agents, and the relational dynamics between them become autonomous (De Jaegher & Di Paolo, 2013). Essentially, social interactions depend on individuals but are, at the same time, not completely determined by them (*ibid*).

Note that the conceptual debate revolving around the body-social problem, as well as methodological individualism, is still ongoing within cognitive science and philosophy of (cognitive) science. Although several positions and 'solutions' have been proposed (see above), it does not appear to be resolved.

The issue of '*human-centredness*' pertains to the circumstance that the majority of theories, models, and concepts about social cognition, social interaction, or generally understanding and interacting with others – including theoretical debates associated with the above mentioned issue of methodology individualism and the body-social problem – adopts a human-centred perspective, i.e. social cognition as it occurs in (adult) human-human interaction (Strasser, 2015, 2018a). Some authors explicitly mention that their conceptions of social interaction and/or cognition is not limited to humans, and may be applied to animals that show social behaviours like coordination (e.g. primates and specific insects like bees, ants, or fireflies; for a detailed discussion see e.g. (De Jaegher and Di Paolo 2007; De Jaegher, Di Paolo, and Gallagher 2010; Lindblom and Ziemke 2003)). Many standard conceptualisations of social cognition aspects, such as e.g. acting jointly, or coordination and cooperation during joint problem solving, do however exclude artificial agents from the start (Strasser, 2018a).

Strasser (2018a) explicitly addresses this issue and states that standard notions “have a tendency to characterize socio-cognitive abilities as if they were unique to sophisticated human beings. However [...] it is important to develop a conceptual framework providing notions that are able to account for various types of social agents.” (Strasser, 2018a, p. 106). To expand conceptual frameworks with regard to social cognition from living beings to artificial agents “some of the standard conditions should rather be seen as biological constraints specific to human beings and should not be used to exclude artificial agents from the start.” (Strasser, 2015, p. 108).

For example, research in the fields of artificial life, robot epigenetics, or social and humanoid robotics in AI aims to create agents capable of social interaction with each other or with humans, as well as social learning (potentially over several developmental phases or generations of agents, see e.g. (Froese and Ziemke 2009; Seel 2012; Smith and Gasser 2005; Stapleton and Froese 2015)). These agents generally do have a (physical or virtual) body capable of performing and/or learning to perform actions towards and interactions with its environment, including other (human or artificial) agents. However, if social cognition – or, more generally, sociality – is conceived of as something that is so uniquely sophisticated that it can only be found in (adult) humans, such agents can simply never fulfil the necessary criteria to qualify as social agents (Strasser, 2015, 2018a). Similar problems arise if the conception of embodiment is based on the assumption that only a ‘living’ body may be embodied. That means, a body possessing certain biological – and only biological – factors is necessary for embodiment. A notion that is based on a conceptualisation of the body as a living body, i.e. a living being, like a human, in a biological sense (see chapter 3.7.1 for more information), does automatically excluding artificial agents. Hence, if social cognitions requires certain biological conditions to be met, then an artificial agent cannot qualify as a social agent and remains a tool or artefact in any case, no matter how sophisticated or complex the behaviours, actions, or interactions it produces may be.

If, however, such aspects are instead viewed as human-specific biological constrains (cf. (Strasser, 2018a)) that are, furthermore, not the only type of conditions required for social cognition, artificial agents may in fact qualify as social agents. One way of achieving this is to re-conceptualise standard notions that are regularly used in the context of social cognition (embodiment, sociality, joint action, cooperation, and the like) in the sense of ‘minimal approaches’. That is, a ‘minimal’ definitions or version of the original concept (Dautenhahn, Ogden, et al., 2002; Froese et al., 2012; Strasser, 2015, 2018a). Essentially, formulating a minimal version of an established concept means to question the necessity of certain conditions associated with the concept given a specific context (Strasser, 2018a). Then this notion is re-conceptualised in a minimal sense, i.e.

by excluding or re-formulating criteria that were previously assumed to be necessary. This way, such minimal approaches (for concrete examples, see next chapter) can account for a wider range of social and/or cognitive abilities and do include non-living entities like artificial agents (ibid).

2.1.1 Social Interaction and (Some) Related Concepts in the Cognitive Sciences

Social interaction (in the context of social cognition) is investigated by several disciplines in cognitive science including psychology, neuroscience, AI, and philosophy. In social interaction research, studies seemingly tend to involve concepts, models, or techniques from different disciplines. For example, a model of social cognition in AI may be based on psychological research or philosophical theories, and a neuroscientific study may aim to investigate whether there is a neural basis for a philosophical theory. However, each discipline appears to have a local meaning of what social interaction or a related, but differently named concept, has. There are multiple concepts, in various disciplines, that appear to either be closely related or to have – in essence – a similar meaning as ‘social interaction’. While several disciplines study the same phenomenon – and may directly or indirectly cooperate in their studies – they have a different approach to, understanding of, and explanatory interest in it (Przyrembel et al., 2012).

In the following, some concepts² that appear to be closely related to the concept of social interaction are described. Described are those concepts that appear to be of interest with respect to the embodied, pragmatic, and social turn in cognitive science, as well as the mentioned associated issues of methodological individualism, body-social problem, and human-centredness.

2.1.1.1 Philosophy of Mind and Phenomenology: Social and Interactive Mind

Philosophy generally approaches a phenomenon by conceptual means. With regard to social cognition, philosophers (primary in philosophy of mind, philosophy of cognitive science, and phenomenology) are mostly concerned with the question if social cognition requires or produces specific kinds of knowledge (Przyrembel et al., 2012), e.g. conceptual or normative aspects regarding how we understand others (see e.g. intersubjectivity and ToM below). The result of philosophical studies are usually conceptual frameworks or theories (ibid).

² Note that some of the concepts listed in this chapter are used in multiple cognitive science disciplines, e.g. there are conceptual debates of intersubjectivity and ToM in philosophy, the study of the neural basis of intersubjectivity and ToM in neuroscience, or behavioural studies on intersubjectivity and ToM in psychology. Similarly, the concepts of joint action and cooperation are used by multiple disciplines. For the classification of concepts as belonging to – primarily – one discipline, concepts have been attributed to the discipline in which they appeared to be primarily used, applied, or employed during the literature research for the present master's thesis. Hence, the attribution of concepts to a certain discipline is specific to the context and process of writing this master's thesis; other authors with another focus may classify them differently. Furthermore, the list of concepts described here is in no way exhaustive; the concepts presented have been selectively chosen with respect to the topic of this master's thesis.

One of the most central concepts with regard to social interaction and social cognition is the Theory of Mind (ToM). Broadly conceived, ToM is concerned with how (human) cognitive agents are able to understand others.

Related Concept	Concept in a nutshell	Authors/source publications
Theory of Mind (ToM) / Mentalising / Mind-Reading	<p>Theory-Theory (TT): There are different variants of TT (e.g. empirical variant, normative variant). In general, TT claims that the understanding of others relies on a theoretical stance that humans adopt. That is, a particular theory, i.e. folk psychology, that provides the individual with a ‘common-sense’ or ‘everyday’ understanding and explanations of why people behave or act in a certain way. The understanding of oneself or others is theoretical, quasi-scientific, and inferential. Mental states of others are unobservable and theoretical entities that are inferred to best predict behavioural data.</p>	<p>(Froese & Gallagher, 2012; S. Gallagher, 2001, 2006; S. Gallagher & Hutto, 2008; S. Gallagher & Zahavi, 2008; Hutto, 2007)</p>
	<p>Simulation Theory (ST): There are different variants of ST (e.g. radical/emphatic variant, analogy-based variant). In general, ST states that the ability to understand others stems from the ability of self-simulation. That is, using one’s own mind as a model to simulate the mind of others. Individuals simulate beliefs, intentions, desires, or emotions by putting themselves in the other’s place. This means to simulate what oneself would be thinking or feeling, and to then project the results of this simulation on the other.</p>	
	<p>Interaction Theory of Mind (I-ToM): I-ToM is a critical response to TT and ST, claiming that the former are based on a third-person or ‘spectator’ perspective. Essentially, I-ToM claims that humans directly perceive and understand others; without relying on theories or simulations. Phenomenological views of understanding others (during interaction), involve non-mentalising, embodied and perceptual approaches. Whenever an individual encounters another individual, i.e. its body, then this body is provided in a meaningful situation and context. This involves a direct perceptions of expressions and actions of this body.</p>	
	<p>Narrative ToM (N-ToM), Narrative Practice Hypothesis (NHP): N-ToM is sometimes considered to be an extension or alternative to TT. N-ToM proposes that narratives allow humans to understand others and their mental states. Narratives are a particular mode of thinking; they relate to the concrete and particular in order to understand meaning, intentions, and the like. The difference to theories (in TT) is that theories are concerned with abstract concepts, e.g. principles or generalisations across particular contexts. The associated NPH claims – in essence – that being told stories, i.e. narratives, from an early age is what enables humans to think about others' mental states and the related contexts. The ability to understand narratives starts around age 2.</p>	

Table 1: Concepts related to social interaction identified in philosophy: ToM

Another concept that appears to closely related to social interaction is the one of intersubjectivity. Intersubjectivity, broadly conceived, means to understand others, e.g. their intentions, emotions, or actions. Note that this concept is here listed under philosophy, but is also used (developmental) psychology.

Related Concept	Concept in a nutshell	Authors/source publications
Inter-subjectivity	<p>Primary: Primary intersubjectivity is a perceptive and proprioceptive sense of the own body, and the distinction of self and other (in direct dyadic interactions). Babies are socially oriented from birth onwards. They can distinguish human from non-human entities. They are capable of identifying other humans as agents performing certain movements, expressions, or actions in the sense of goal-directed and intentional behaviours, which are imitated. This involves affective coordination; a basic ability for interacting with others. According to phenomenologists, this does not involve mentalising in the sense of taking a theoretical stance or a simulation of some inner state, but instead is perception-, action-, and body-based.</p>	(S. Gallagher, 2011b; S. Gallagher & Hutto, 2008; S. Gallagher & Zahavi, 2008; Lindblom, 2015a; Morganti, 2008; Stevanovic & Koski, 2018)
	<p>Secondary: After 9-14 months of age, infants learn to not only recognise someone as an agent performing an intentional action, but start to grasp the meaning of and conceptual relations between intentional movements, expressions, or actions. This second level of intersubjectivity involves coordinated forms of triadic interactions, e.g. between infant and caretaker and a shared object like a toy. This includes joint attention, gaze-following, and forms the basis for empathy. Phenomenologists claim that this does not involve a process of inferring intentions as if the child had the intention him/herself, or as if these intentions were ‘hidden away’ in the mind of the other. The child perceives others as intentional agents and learns to understand what the other wants in a certain situation/contexts.</p>	
	<p>Tertiary: At 20-24 months of age, children develop tertiary intersubjectivity. This involves the ability to not only distinguish oneself from others, but also to gain an understanding that one is simultaneously perceived by others. Children then also recognise themselves in a mirror, which indicates self-recognition. Joint actions become more complex, like joint pretense play. Joint actions also start to involve language and complex forms of communication and interactions, which can now be understood in relation to oneself and others in a broader social context. Note that by 6-7 years of age not only joint or shared, but also collective intentionality is developed. That is, understanding social rules and norms and how to apply them to oneself. (This is sometimes referred to as an advanced tertiary or a quaternary level of intersubjectivity.)</p>	

Table 2: Concepts related to social interaction identified in philosophy: Intersubjectivity

2.1.1.2 Neuroscience and Psychology: Social and Interactive Brains

From a neuroscientific and psychological perspective, investigations of social cognition aim to identify specific underlying cognitive mechanisms that are required or involved in social cognition (Przyrembel et al., 2012). Neuroscience (primarily the sub-field of social neuroscience) and psychology (primarily social psychology and developmental psychology) generally investigate social cognition with experimental methods, which can be behavioural tests and experiments, biological and neurological measures and imaging techniques, or a combination thereof (e.g. neural activity measured through fMRI or EEG during a certain task that involve ToM, (ibid)).

Related Concept	Concept in a nutshell	Authors/source publications
Second-person perspective/ cognition	<p>Rather than investigating one individual when focussing on or experiencing a task and/or reflecting on oneself (first person), or when observing someone else performing a certain task (third person), second person cognition involves the investigation of (at least) two individuals during their interactions.</p> <p>Social cognition is different when in direct interaction with another, i.e. when being in a second person perspective, in which social interaction is a constitutive element of an individuals social cognition.</p> <p>In psychological and/or neuroscientific experiments, second-person cognition/perspective means to</p> <ul style="list-style-type: none"> • study an individual (one-brain study) during interaction with another subject whose biological/neural or behavioural responses are not measured, or • studying/measuring responses of two individuals (two-brain study/hyperscanning) during a prolonged time of interacting with each other. 	(de Bruin, van Elk, and Newen 2012; Pérez and Gomila 2021; Schilbach 2010; Schilbach et al. 2013a, 2013b)
Online social cognition / online interaction	<p>Similar to and sometimes used synonymous with ‘second-person’ perspective/cognition. Online social cognition/interaction pertains to instances of reciprocal/interpersonal interactions in which two or more agents are actively involved in some kind of real-life(-like) social engagement.</p> <p>In contrast, offline social cognition/interaction refers to two or more agents who passively observe each other during social interactions.</p> <p>Note that the terms ‘online’ and ‘offline’ <i>social</i> cognition have a different meaning than ‘online’ and ‘offline’ cognition (e.g. as in embodied approaches to cognition, see chapter 3.8.1)</p>	(Auvray & Rohde, 2012; Pfeiffer et al., 2013; Przyrembel et al., 2012; Schilbach, 2014)

<p>Interpersonal/ reciprocal interaction</p>	<p>Describes the direct interaction between (at least) two individuals, in which the actions of one directly influence the actions of the other and vice versa.</p> <p>Experimental paradigms of interpersonal/reciprocal interaction include game-theory-based cooperative or competitive games, gaze-following or sharing in terms of joint attention, and others.</p> <p>This may be done in two general ways:</p> <ul style="list-style-type: none"> • Scanning/measuring one person while interacting with someone outside the scanner (e.g. in case of fMRI), or • simultaneously scanning (at least) two subject when they are involved in a social task (e.g. via hyperscanning methods; EEG, looking for brain activation and/or synchronisation across subjects). 	<p>(Konvalinka & Roepstorff, 2012; McCabe et al., 2001; Montague et al., 2002; Schilbach, 2015)</p>
<p>ToM network system, mirror neurons system (MNS)</p>	<p>Assumed to be the neural basis, or at least closely associated with, the theoretical concept ToM. Specific brain areas have been identified as being active during social tasks or interactions assumed to involve (some type of) ToM. This ToM network/system typically includes the areas of the medial prefrontal cortex, precuneus, superior temporal sulcus, and the right temporal parietal junction.</p> <p>This ToM network/system is sometimes related to and distinguished from the MNS. In short, mirror neurons are neurons, or areas of neurons, that are – to a certain degree – active during i) observing a specific task, and ii) carrying out this specific task oneself. This has been extended to social contexts and the area of social interaction, as well as empathy. The MNS, with regard to motor-action, feelings, and sensation, typically involves the somatosensory cortices, the anterior parts of the insula, and the anterior cingulate cortex.</p>	<p>(Chaminade et al., 2012; Dumas et al., 2011; Konvalinka & Roepstorff, 2012; Przyrembel et al., 2012; Rilling et al., 2004a; Schilbach et al., 2006)</p>

Table 3: Concepts related to social interaction identified in neuroscience and psychology

Note that, with regard to the issue of human-centredness, most of the above mentioned concepts are generally used within the context of human-human interaction. However, some empirical studies (Gallagher et al. 2002; Kulms and Kopp 2018; Limerick, Coyle, and Moore 2014; McCabe et al. 2001; Rilling et al. 2004; Sanfey et al. 2003; Schilbach et al. 2006) have investigated differences between (alleged) human-human social interaction and the interactions of humans with computers and/or artificial agents. For example, McCabe and colleagues (2001) performed a fMRI study in which participants’ brain activity was recorded while playing a two-person game in the scanner. The subjects believed to be playing either with another human or with a computer. Interestingly, participants tended to cooperate less in the computer condition than in the ‘human’ (who was, in fact, a computer) condition (McCabe et al. 2001). Furthermore, a comparison between the two conditions revealed that, if the subjects *believed* that they were playing with a

human partner (while, in fact, they were not), then activation of specific regions of the prefrontal cortex (which is associated with the ToM-network) occurred (McCabe et al., 2001; Przyrembel et al., 2012).

Using PET, a study by Helen Gallagher and colleagues (H. L. Gallagher et al., 2002) aimed to investigate the ‘intentional stance’ (see also following section on AI and computer science) in terms of the neural activation associated with online mentalising during the two-person game rock-paper-scissors (ibid). Notably, the subjects believed to be playing with either a real person, who they had met prior to the PET-scan, or with a computer. However, all responses of the subject’s partner during the game were – unbeknown to the subject – computer-generated. During both conditions (computer and alleged human) significant activation of certain brain regions was recorded. However, brain activation in the most anterior portion of the paracingulate cortex bilaterally – which is associated with understanding intentions and social interaction (H. Walter et al., 2004) – was significantly greater during the interaction with a ‘human’ (H. L. Gallagher et al., 2002). Activation in the right inferior frontal cortex – which has been associated with motor/action inhibition and control (Forstmann et al., 2008) – was significantly greater in the supposed computer condition (H. L. Gallagher et al., 2002).

Rilling and colleagues (2004a) performed a similar study. However, this study explicitly aimed at immersing the participants in a social interaction that has meaningful consequences to them (ibid). Using a game-theory-based paradigm, specifically, the Ultimatum Game and the Prisoner’s Dilemma, the subject in the scanner played those games with either an (alleged) human partner or with a computer partner outside the scanner. Similar to the results of the two studies mentioned just before, this study showed that activation of ToM-associated brain areas (precisely: the anterior paracingulate cortex, the posterior, the mid superior temporal sulcus, in addition to the hippocampus, and some regions of the hypothalamus) only occurred when the subjects *believed* that they were playing with real human partner – who, in actuality, was a computer partner (Przyrembel et al., 2012; Rilling et al., 2004a).

Some studies also use only virtual characters/agents as interaction partners. Employing an eye-gaze paradigm, Schilbach and colleagues (2006, 2011) investigated the neural activation (via fMRI) during active involvement in social interaction in a virtual reality setting. These characters gazed at either the participant or another bystander character. When gaze-following, in the sense of shared/joint attention – was initiated by the participants, increased neural activity in the ventral striatum (associated with movement and reward) was observed (Przyrembel et al., 2012; Schil-

bach et al., 2006, 2011).

2.1.1.3 AI and Computer Science: Social and Interactive Artificial Agents

Broadly speaking, *AI* (including robotics) and *computer science* (including Human-Computer and Human-Robot Interaction) appear to focus on modelling and investigating socio-cognitive and socially intelligent abilities. On the one hand, this means that the scientific interest lies in modelling such abilities and implementing them into (robotic or virtual/simulated) artificial agents in order to achieve human-like capabilities and/or to further the understanding of these capabilities ('as-is'). On the other hand, this means to investigate if and how humans attribute socio-cognitive and socially intelligent abilities to those agents ('as-if').

Related Concept	Concept in a nutshell	Authors/source publications
<p>Minimal embodiment, minimal approach to embodiment</p>	<p>Embodiment is relational with respect to a certain environment. It is seen as a scale or spectrum, rather than a dichotomy of either being or not being embodied. An entity or agent cannot be embodied to the same degree in all possible environment (e.g. a fish in the water vs on land). A <i>minimal definition of embodiment</i> is that:</p> <p>The system S in a particular environment E is embodied in E if perturbary channels exist between S and E.</p> <p>Specifically, at every time t at which</p> <ul style="list-style-type: none"> • E (or some subsets of its possible states) in relation to S has the capacity to perturb S (or some subset of its possible states), while • S (or some of its possible states) in relation to E has the capacity to perturb E (or a subset of its possible states), <p>S is embodied in E.</p> <p>This is an operational definition that can be applied to a variety of contexts and systems, e.g. from inanimate objects, to robots, to humans.</p>	<p>(Dautenhahn, Bond, et al., 2002; Dautenhahn, Ogden, et al., 2002; Dautenhahn, 2011)</p>
<p>Degrees of embodiment</p>	<p>In relation to a minimal approach to embodiment (see cell above), the Degree of Embodiment (DOM) over the parameters S and E is defined as the function f of the vectors x and y, and t:</p> $DOM_{S,E} = f(x,y,t)$ <p>in which:</p> <ul style="list-style-type: none"> • t is a particular point in time • x is a vector of particular measurements expressing perturbation of S by E • y is a vector containing particular measurements expressing perturbation of E by S <p>Examples variables for x and y are 'perturbary bandwidth' (the range of events S and E can produce that the other may be perturbed by, as well as the force with which these events are produced). and 'structural variability' (the extend in which S and E can structurally</p>	<p>(Dautenhahn, Bond, et al., 2002; Dautenhahn, 2011)</p>

<p>Degrees of embodiment (continued)</p>	<p>vary when perturbed). Simply put, the operationalisation of the variable x (in robotics) is the manipulation of sensors, while y is operationalised as a manipulation of actuators. Examples of x commonly used in robotic are: increasing the number of sensors, using different kind of sensors, increasing exploitation of channels of information provided by the sensors. Plus, in case of y, examples follow the same idea, but with respect to the manipulation of actuators with which S can affect E. Examples are: replacing fixed-response controllers by controllers that allow adaptive behaviour, explicitly exploiting memory and learning (a robots ability to “remember” and “learn” from past encounters and apply them to either current task or predict future responses), increasing the agents/robots sources for learning by complementing individual with social learning and imitation (e.g. using others as ‘social tools’ to adapt to a particular niche).</p>	
<p>Socially situated and embedded agents, interaction-aware agents</p>	<p>An <i>embodied agent</i> – in a minimal approach to embodiment (see above) – is <i>situated</i> if it gains information on its environment through its sensors by interacting with this environment – the situated agent does so on its own, without human intervention. ‘Situatdness’ in this sense means for agent to get sensory-motor feedback from the interactions with a particular environment, which informs real-time sensory data, based on which the agent can act upon in real-time, or acquire its own history (via learning mechanisms). An agent is <i>socially situated</i> if it acquires information about both the social and physical environment by means of interaction. An agent is <i>embedded</i> if the agent and its environment are coupled so tightly, that the are best analysed or modelled as a single agent-environment system (‘structural coupling’, see chapter 3.7.1). An agent is <i>socially embedded</i> if it is part of a social system of agents and their interactions. It is structurally coupled to its physical and social environment and best modelled as a part of a larger system of other agents, rather than as a single agent interacting with them. In order to be socially embedded, an agent has to have awareness (not necessarily in terms of explicit representations) to embody knowledge of the (physical and social) environment. i.e. It has to be (at least partially) aware of social structures and interactions in its environment; it needs to be an <i>interaction-aware agent</i>.</p>	<p>(Brooks, 1991b; Dautenhahn, Ogden, et al., 2002; Dautenhahn, 2011; Edmonds, 1999; R. Pfeifer & Scheier, 1999)</p>
<p>Minimal joint/collective action; agency</p>	<p><i>Joint action</i> is a requirement for ‘successful’ social interaction in the context of social cognition. Requirements for acting jointly is the ability to act (agency). A minimal approach to joint action that includes artificial agents must therefore include minimal definitions of agency. Standard notions of action tend to define intentionality or consciousness as requirements, e.g. to perform goal-directed actions. Understanding intentionality and consciousness (in a strong sense) as human-specific biological constraints, <i>a minimal notion of action</i> does not require the goal of an action to occur in the acting system. I.e. the goal of an action can be generated in another system.. A system acting in a minimal sense needs to perceive the environment</p>	<p>(Strasser, 2015, 2018a; Wachsmuth, 2015)</p>

<p>Minimal joint/collective action; agency (continued)</p>	<p>via sensors, recognise and represent (action-related) goals as goal – but does not have to generate the goal itself – and have actuators to perform actions. An artificial agents fulfilling those minimal requirements is considered to have agency in a sense that allows for joint action. This minimal notion is a functional definition that may be adapted for usage in with respect to different paradigms, research areas, or contexts.</p>	
<p>Minimal joint/collective action, agency; coordination</p>	<p>Socio-cognitive abilities (in the sense of interpreting social cues or information) is an essential condition for coordination in joint actions. Coordination in this sense, does, however, involve ToM, which is generally human-centred. The demanding cognitive resources, e.g. ability to represent a full range of complex mental states or language competencies, of ToM are understood as biological constraints specific to human adults. In a <i>minimal notion of coordination</i>, representations of less complex mental states, e.g. encounters and registrations, are the minimum requirement for an agent to anticipate the behaviours of other agents in a robust enough manner. In this sense, this minimal notion can account for a broader range of agents, like infants, non-human animals, or artificial agents. This minimal notion is a functional definition that may be adapted for usage in with respect to different paradigms, research areas, or contexts.</p>	<p>(Strasser, 2015, 2018a)</p>
<p>Social agents, autonomy, sociality</p>	<p>Social agents are capable of human-like socio-cognitive abilities and posses capabilities for human-like interactions. This requires an interplay of autonomy and sociality. <i>Sociability</i> means to act ‘appropriately’ in the social domain (i.e. with regard to norms, expectations, etc.) It encompasses the ability of an artificial agent to elicit social response from humans by being able to have and/or to follow human and/or human-like social interaction. <i>Autonomy</i> describes the ability of an artificial agent A to reach a certain goal G autonomously from other agents. For A to be autonomous from a human H, it must have the capacity to reach G without assistance or intervention from H. Conversely, If A cannot reach G without the involvement of H, it is (co-)dependent on H, and thus not an autonomous agent.</p>	<p>(C. Breazeal, 2002; Dautenhahn, 2011; Mohammad & Nishida, 2015)</p>

Table 4: Concepts related to social interaction identified in AI (including artificial life, behaviour-based, evolutionary, and social robotics, as well as theory/philosophy of AI).

An example system that is repeatedly referred to in the literature on social artificial agents is a robot called ‘*Kismet*’. The agent *Kismet* (developed by Cynthia Breazeal for her doctoral dissertation under supervision of Rodney Brooks, cf. (C. L. Breazeal, 2000)) is a ‘sociable machine’ that can be considered a socially embedded agent in the sense of being interaction-aware (Dautenhahn, Ogden, et al., 2002). *Kismet* is an embodied agent (specifically, a robot head) capable of emotional expression in the sense of response-specific facial expressions when interacting with a

human. The agent is equipped with visual, auditory, and proprioceptive sensors, and several actuators to control movement of its face and neck. The face is equipped with 15 degrees of freedom (two for each eye, one for each eye lid, two for each eye brow, four for lips, one for the jaw) as well as three for its neck, allowing for a wide range of facial movements and expressions (C. L. Breazeal, 2000; Dautenhahn, Ogden, et al., 2002). Kismet uses these actuators to regulate dyadic human-robot interaction via facial expressions. The produced expressive responses are based on its ability to recognise faces and its use of ‘social amplification’. E.g. if a person moves in very close, Kismet moves back its head (as a human-like retreat-response) (C. L. Breazeal, 2000; Dautenhahn, Ogden, et al., 2002). Kismet can produce emotional expressions of anger, calm, disgust, happiness, sadness, and fear, and also expresses interest (e.g. when approached), boredom (when not being interacted with), and surprise (C. L. Breazeal, 2000; Lindblom & Ziemke, 2003). Due to this large range of emotional expressions, Kismet can perturb the social environment (in the sense of the one human it is interacting with), but can be perturbed by the social environment to a lesser extent. Its design allows for a tight coupling of Kismet and the human interaction partner, but Kismet can, apart from this, not perceive other interactional structures or adapt to other social structures (C. L. Breazeal, 2000; Dautenhahn, Ogden, et al., 2002).

A virtual/simulated agent that can serve as an example of embodiment, social embeddedness, interaction-awareness, and cooperation (in the minimal sense described in the above tables), is the humanoid agent ‘Max’ (Kopp et al., 2005, 2005). Max has been developed in the context of the University of Bielefeld Max Project, and has been researched in a series of experimental studies (for a relatively recent review see (Wachsmuth 2015)). Max is an embodied agent with a human-like appearance that can – by means of sensors like various microphones, trackers, and cameras – perceive its physical and social environment, e.g. it can recognise a human interaction partner as such and process verbal instructions or gestures.

With regard to actions towards the environment, including humans, Max has conversational abilities (natural language, mixed-initiative dialogue, i.e. initiated by the agent or a human) and can produce a series of facial and bodily expressions, including non-verbal behaviours. Max can communicate via a synthetic voice, that can be modulated. Its animated face is based on a model of simulated facial muscle effects including the movement of eyebrows, lip-movement. and lip-synchronous speech (Wachsmuth, 2015). Its body is based on a skeleton model with approximately a third of the degrees of freedom a human skeleton has, and can thus produce a variety of movements and gestures (ibid). Hence, Max can produce a variety of (emotional) expressions. Max

employs those expressions (using its face, body, and voice) in response to external stimuli, like recognising faces or ‘hearing’ specific words. Concerning Max’ modelling, it is controlled by a hybrid cognitive architecture, combining symbolic and dynamic systems paradigms (see chapter 3.3, 3.4, 3.5). The architecture has been adapted and refined over several years (see (Wachsmuth, 2015) for a detailed description of the Max cognitive architecture). What is of interest within the context of this thesis, is Max’ interactional abilities. Notably, Max has been implemented in a real-world setting, in which Max acts as a museum guide capable of holding and following real-time interactions with the visitors of the Heinz Nixdorf MuseumsForum in Germany since 2004 (Kopp et al., 2005; Wachsmuth, 2015). As a museum guide, Max is displayed on a large projection screen from which it can communicate and interact with visitors by engaging them in a conversation, e.g. to greet them, or to provide them with information regarding the museum (Wachsmuth, 2015).

A more recent example for an embodied, socially embedded, and autonomous agent (in the sense introduced in the tables above) is the humanoid robot platform ‘*iCub*’. For example, in a recent study (Tanevska et al., 2020), *iCub* was used in a child-caretaker scenario to investigate human-robot interaction. In such a scenario, the ‘child’ (*iCub*) seeks the attention of the ‘caretaker’ (a human) when being left alone, but no longer seeks caretaker-attention after a while, e.g. after playing with a toy for a period of time (ibid). The study involved two conditions to investigate the fundamentals of interaction; *iCub* would either play on its own, or seek the attention of a caretaker and ask for their company (ibid). The human partner could establish and/or maintain interaction with *iCub* by touching it, showing it their face and smiling, or by showing it toys – which would be recognised by *iCub* (ibid).

Another study employing *iCub* (Marchesi et al. 2019) was concerned with the perception of the robot and its behaviour. For this purpose, a questionnaire was developed in which participants were given a sequence of photos showing *iCub* involved in naturalistic scenarios, and had to choose the likelihood of different explanations for the robot’s behaviours (ibid). These explanations were classified as either ‘mechanistic’ or ‘mentalistic’ (ibid). Interestingly, the ratings were biased towards mechanistic explanations, but also included mentalistic explanations, which – according to the authors (Marchesi et al., 2019) – indicates that humans may adopt an ‘intentional stance’ (see table below) towards artificial agents.

Related Concept	Concept in a nutshell	Authors and source publications
Intentional stance, intentional agents	Developed by the philosopher Daniel Dennett: the intentional stance describes the deposition of humans to take on a perspective in which they ascribe intentions, beliefs, desires and the like to other (human and non-human) entities in order to predict and/or to explain their behaviour. This concept has been incorporated into social cognition research, including the fields of human-computer and human-robot interaction. Humans tend to – at least in some cases – adopt an intentional stance towards artificial agents (robotic or simulated agents).	(Dennett, 1988; H. L. Gallagher et al., 2002; Marchesi et al., 2019; Perez-Osorio & Wykowska, 2019; Strasser, 2018a; Thellman, 2021)
Computers As Social Actors (CASA)	CASA postulates that humans tend to ‘mindlessly’ apply social rules, norms, and expectations (‘social scripts’) from human-human social interactions to human-computer interaction, i.e. they ascribe agency, social qualities, or human-like characteristics to computers and computational agents. Even though they may be aware that the computer is not a ‘conscious’ or ‘living’ being, this is not necessarily reflected in their behaviour and actions. This is due to processes of anthropomorphism and ascription, i.e. attributing human characteristics to non-human entities and treating computational agents <i>as if</i> they were human actors/agents. (See also chapters 2.1, 4.1.1.3)	(Gambino et al., 2020; Limerick et al., 2014; Nass & Moon, 2000; Reeves & Nass, 1996; Schaumburg, 2001)
Extended CASA	The original CASA was formulated around 30 years ago; a time when human-technology interaction was different from today. At that time, social scripts rooted in human-human interactions were the primary social script available to apply to human-computer interactions. The extended CASA argues that today, due to the acceleration of technological advances and their integration into society and daily life, humans may develop <i>specific</i> social scripts for the interaction with computers/computational agents. Such specialised socio-technological scripts may be based on or appear to be similar to human-human social scripts, but are, in fact, distinctly different from human-human social scripts. (See also chapter 4.1.1.3)	(Gambino et al., 2020)

Table 5: Concepts related to social interaction identified in computer science (including the fields of human-robot and human-computer interaction).

3 Social Interaction across Cognitive Science Paradigms

3.1 *Why Investigate Social Interaction on the Level of Cognitive Science Paradigms?*

After having reviewed several conceptions of social interaction in the form of related concepts across different cognitive science disciplines (chapter 2.1.1), the concept of social interaction will be investigated on the level of cognitive science paradigms. Rather than keeping on to review and analyse a multitude of local (i.e. disciplinary) conceptions of social interaction, taking on a perspective on a higher level of abstraction appears promising. This is because approaching the topic of social interaction in cognition at the level of paradigms should reveal the theoretical assumptions, metaphysical claims, and methodological perspectives and means that determine how cognitive phenomena may be modelled or empirically investigated, as well as how research results and insights from different disciplines ought to be obtained, understood, and integrated in *and* across those disciplines.

3.2 *Cognitivism and Computational Approaches to Cognition*

Essentially, the computational paradigm – also referred to as ‘computationalism, ‘representation-ism, or ‘cognitivism’ (Adams & Aizawa, 2010; S. Walter, 2014a) – understands cognition as a form of information-processing. The brain is the central processing unit that handles input-output relations, i.e. the perception of input from the environment and the generation of an output in form of actions. This approach is based on two key principles: 1) the concept of representations, i.e. that mental states are internal representational, and 2) the premises that cognition is a form of information-processing and the computer metaphor of the mind, i.e. the human brain processes information like a computer.

The computational approach holds that:

1. Cognitive processing is a cranial process, i.e. involves the brain *only* (‘intra-cranialism’, cf. (Adams & Aizawa, 2010; S. Walter, 2014b)).
2. Cognition is computation, i.e. a kind of information processing that:

- i. is conceptualised as input-output relations, i.e. perceptions in the form of sensory inputs from ‘outside’ are processed on the inside (brain) to produce a motor output in the form of actions, and
 - ii. is generally associated with functionalism, i.e. cognition is fully determined by its functional role, which represents an autonomous level of analysis (Newen et al., 2018).
3. Information processing is conceptualised as being either:
- i. symbolic, i.e. symbol processing as computation over mental representations (see chapter 3.3), or
 - ii. sub-symbolic, i.e. parallel distributed processing (see chapter chapter 3.4) in artificial neural networks inspired by the human brain (Newen et al., 2018; S. Walter, 2014b)

3.3 *Symbol Processing Approach to Cognition*

Following a top-down and functionalist approach to model the mind, this understanding of cognition focused on the rule-based manipulation – or processing – of symbols. This functionalistic view provided a promising attempt about how to relate (human) biology and psychology: cognitive phenomena are a software running on (biological) hardware (Lindblom, 2015a; S. Walter, 2014a). Central to this understanding of cognition are the terms of ‘symbols’, ‘representation’, and ‘computation’.

In essence, cognitive processes are conceptualised as symbolic information-processing, similar to the way a computer processes information. The information – represented as symbols – comprises algorithmically specifiable symbols that can be manipulated in accordance with a set of specified rules (Beer, 2000; Samuels, 2019).

symbol systems are characterized by sets of rules – typically recursive in form – that specify which combinations of symbols are well-formed or grammatical [...] In short, symbolic representational systems of the sort relevant to classical computation possess a combinatorial syntax and semantics in much the same way as logical systems, and natural languages [...] (Samuels, 2019, p. 104).

Broadly conceived, a symbolic system is a model that processes information in the form of symbols. Symbolic systems are typically expressed as a system (a program), that takes a symbolic description of a problem as an input (Beer, 2000). The purpose of the symbol system is to solve this problem. To achieve that, the system uses the general ‘knowledge’ and/or rules it has on the problem domain in which it operates – which too, is symbolically represented – to process the input

(ibid). That is, the input information is manipulated in a syntactic fashion in order to obtain and output; the solution to the problem (ibid). This manipulation of symbols is based on a set of rules that specify an algorithm (i.e. a specifiable, precise, and step-wise procedure) for ‘mapping’ the inputs to outputs (Samuels, 2019). The algorithm specifies a function on how the symbols are to be computed that is:

formal in that the operations specified by the algorithm are defined with respect to the syntactic, as opposed to semantic, properties of symbols. In this regard, they are akin to familiar grade-school algorithms for multiplication and long division, which are formulated in terms of operations on formally characterized items – Arabic numerals – and not the things they represent – i.e. numbers. One consequence of this is that the task of classically modeling a cognitive process is entwined with the task of specifying the formal properties of the representations involved. Change the symbol system –e.g. from Arabic to Roman numerals – and one must typically change the algorithm as well. (Samuels, 2019, p. 105).

After a the symbol system has ‘run’ the algorithm (i.e. by means of rule-based manipulations of the symbolically represented input), it generates an output. That is, a ‘solution’ to the problem (the input) the system was to solve by means of computation over symbols.

Crucially, the strength of such an approach to cognition, i.e. cognitive processes as a form of information-processing in the sense of computations, is that it makes those processes interpretable. Cognitive processes are modelled by formal procedures (sets of rules, algorithms), but – according to Samuels (2019)“the symbols involved have semantic properties as well [...] it is possible to make sense of – to interpret – the process, not merely as the manipulation of formal tokens, but as mappings from meaningful states to other meaningful states.”(Samuels, 2019, p. 105). Simply put, symbol systems can be interpreted by researchers in a similar fashion as a multiplication of numbers; the multiplication (in the sense of an operation on Arabic numerals, i.e. symbols) and its result can be precisely interpreted due to the associated ‘meaning’ of numbers (in the sense of representations).

Hence, the explanatory focus of a symbol-based model is on the structure and content of the representations and the efficiency of the algorithms used to process/compute them (Beer, 2000), and the explanatory value is that they are interpretable (Samuels, 2019).

As has been noted earlier, central to this understanding of cognition are the terms of ‘representation’ and ‘computation’. Criticism of this paradigms too centres around those concepts.

Common arguments against the computational paradigm within cognitive science revolve around the lack of connection between the external environment and the internal representations with re-

gard to computational models of the human mind (Lindblom, 2015a; S. Walter, 2014a), as well as the role of semantic understanding and meaning (or lack thereof) in intelligence. Dreyfus (1992; 1995)– who was among the first to discuss the concept of ‘situatedness’ – stressed that intelligence has to include a kind of ‘background common sense’, which humans are equipped with by being situated in their physical and social world. Rather than mastering ‘micro worlds’ (i.e. a well-specified and limited problem area, like a specific type of game), an intelligent system has to be situated (ibid). It cannot be separated from the rest of human life, which connotes that cognition must take the body and its specific environment into account (ibid). He argued that the knowledge about the world represented in a computer is provided from the outside. The knowledge of the environment is not localised within the computer program, but provided by the program’s creator, who specified how elements of the external world should be conceptualised, as well as what elements would be represented in which way. The program itself does not have direct access to the environment and the ‘mapping’ of external to internal was done by the programmer.

Searl (1980, 1990) criticised classical AI programs – and, thereby, the idea that the human brain acts like a computer – by pointing out the lack of (semantic) relations between external representations/objects and internal representations. According to Searl, in order to achieve semantic understanding, a cognitive system would require ‘causal powers’ (cf. (Searle, 1980, 1990)) similar to living organisms, i.e. the human brain and the ability to relate internal structures to ‘their’ external environment.

3.3.1 (Social) Action and Interaction within the Symbol-Processing Approach

Positioning the concept of social interaction within this approach appears to be problematic. Within the symbol-processing approach to cognition, the focus of cognition is on the individual and how cognition allows an individual to internally represent the external world – to receive inputs via perceptions and produce outputs in form of actions based on (symbolic) processing of the input. How exactly a cognitive system accesses and ‘understands’ this environment, or how a cognitive system (repeatedly) acts and interacts in the world after processing some input and producing an according output (or how these actions affect – or not affect – the environment³) was simply not of primary relevance within this research programme. It instead focussed on internal processes.

Thus, classical computational approaches to the mind and cognition can be conceived of as ‘dis-

³ This is related to the ‘frame problem’ and associated epistemological debates, see e.g. (Fodor, 1983; McCarthy & Hayes, 1969).

embodied' (Lindblom, 2015a). They lack (direct) access to the environment and have no way of (directly) manipulating or interacting with this environment or other agents in this environment. In short, symbol-processing based computational systems lack a means to link perception and action; they lack a body allowing them to interact and *engage* with the environment (ibid). How agents engage their environment through actions is, however, pivotal for conceptualising social interaction. Regardless of the role representations or symbols may or may not play in human cognition and action, addressing (direct) social interaction must address an agent's body – at least in the sense of a 'basic' means for performing actions in an environment and/or engaging with it. Furthermore, this environment would have to not only include physical elements, such as objects, but also cultural and social aspects, as well as other agents that can be interacted with. A framework for cognition that lends itself to a (fruitful) conceptualisation of social interaction has to consider an agent's actions in and towards the environment.

3.4 Connectionism, Parallel Distributed Processing, and Artificial Neural Networks

An extension of the functionalist approach to cognition was 'connectionism'. Inspired by neural aspects of cognition and the way neurons 'connect', connectionism models cognitive phenomena through a bottom-up approach (Beer, 2000; S. Walter, 2014a).

Resembling the biologically distributed functions of the brain at a general level, connectionists employ artificial neural networks that exhibit activations of neurons and the 'forwarding' of activity from neuron to neuron, thereby either inhibiting or exhibiting it (ibid). A typical connectionist model is a system expressed as a layered computational network that consists of simple, neuron-like elements (Beer, 2000). These elements can be 'trained' (i.e. through different training or learning algorithms) to "transform a numerical input representation into a numerical output representation." (Beer, 2000, p. 96).

While symbol-based models of the mind are based on serial or sequential processing of symbols in a rule-based fashion, connectionist models process information as distributed patterns of (spreading) activity; i.e. 'parallel distributed processing', also referred to as 'sub-symbolic' processing (Newen et al., 2018; H. Walter et al., 2004; S. Walter, 2014a). Furthermore, classical computational models required discrete input and concrete, well-defined tasks, but connectionist neural networks can manage 'fuzzy' input (S. Walter, 2014a). Notably, connectionist neural networks are able to learn, in a broad sense (Beer, 2000). Networks 'learn' by changing the connec-

tions (weights) between neurons by means of different learning or training algorithms (ibid). Such models and are used in the area of pattern recognition; they are able to learn to process and recognise faces and to recognise emotions from facial features or speech (Cottrell, 1991; Valentin et al., 1994), as well as to form the past tense of verbs or simple grammatical rules (Najafi, 2018). Due to their parallel distributed processing, such networks display ‘emergent’ properties (Vernon, 2014; S. Walter, 2014a). The explanatory focus of connectionist models is on the network architecture, the training/learning algorithm, and the intermediate, parallel distributed processing that ‘produces’ its emergent properties (Beer, 2000).

3.4.1 (Social) Action and Interaction in Connectionism

While neural networks are able to handle tasks that the designer/programmer has not directly programmed into them (‘emergent’ properties), most networks still start to learn from the programmer’s specification of a task. Artificial neural networks do create complex patterns to map from input to output by themselves. However, the major criticism directed at symbolic computational models off the mind, still applies to connectionist models. Lindblom (2015a) states that, just like symbol-processing based computational models in the cognitivist paradigm, connectionist models do not solve the ‘symbol grounding problem’ (cf. (Harnad, 1990)), and neglect the role of embodiment, situatedness, and the environment within cognition. Cognition is understood in terms of input-output relations, as something distinct from action or behaviour. Furthermore, considering the relation between biological authenticity and algorithmic implementations, artificial neural networks have been criticised to have limited biological plausibility (Lindblom, 2015a).

Within symbol-processing based and connectionist systems, cognition is understood as a kind of information processing within a central system, detached from a body, an environment or the world (ibid). Thus, both symbol-processing based and (early) connectionist approaches to cognition do not appear to be fruitful starting points for positioning and discussing social interaction within cognitive science.

3.5 A Note on Social Interaction in Computational Paradigms

It should be noted that the line of argumentation against computational paradigm described above only holds – strictly speaking – if such systems are, in fact, ‘disembodied’. That is, if those system do lack a physical or virtual body which is embedded into a (physical or virtual) environment. Dreyfus (1992) discussed the possibility of creating an artificial embodied agent using

computers. He suggested that it would in fact be possible to create an artificial embodied agent if this agent possessed a set of sensor-motor mechanisms similar to a living organism. Pfeifer and Scheier (1999) too point out that neural networks may primarily be viewed as ‘detached’ information processors. But if artificial neural networks become embedded, i.e. integrated in a physical robot in a specific environment, then they may give rise to *intelligent autonomous agents* (H. L. Dreyfus, 1992; Lindblom, 2015a). Particularly the integration of connectionist systems and symbol processing/manipulation based systems – ‘*hybrid*’ systems, i.e. systems including elements of both classical computational and connectionist models – appears promising (see next chapter). Implementing – hybrid or dynamical (see chapter 3.5) – systems into autonomous embodied (virtual/simulated) agents or robots may potentially allow for investigation of social interaction. If their (physical or simulated/software) environment was to include other (human or artificial) autonomous agents who are interacted with, they may be capable of social action and interaction. Hypothetically, if an autonomous system is capable of producing intelligent behaviours and actions in an environment in which it is situated along with other autonomous agents, it could be capable of participating in social interaction. For example, the work of Di Paolo and colleagues (Di Paolo et al., 2008, 2013) demonstrated social interaction, social contingencies, and coordination (see also chapter 3.11.3) in simulated agents that have been modelled based on enactivist and evolutionary robotics approaches.

3.5.1 Cognitive Architectures and the Implementation of Computational Models in Artificial Agents

As pointed out above, cognitivist computational models are – on a conceptual level – ‘disembodied’. However, they may still be applied to study social interaction or cognition via implementation of such systems in (virtual or physical) artificial agents, controlled by e.g. a cognitive architecture. Cognitive architectures are an approach of computational cognitive modelling meant to represent a model of human cognition on a general level. Cognitive architectures are an “attempt to fulfil the goal of integrating many of the results of experimental psychology [...] into a unified theoretical framework with broad coverage” (Chipman, 2016, p. 3) in terms of “veridical theories of the phenomena of human cognition” (Chipman, 2016, p. 3).

Broadly, cognitive architectures can be classified into three different categories: biologically-inspired architectures (e.g. by neural brain activity, like in connectionist modelling), philosophically-inspired architectures (inspired e.g. by philosophical theories on explicit and implicit knowledge), and ‘classical’, i.e. symbolic, problem-solving AI architectures (Adam et al., 2016). Such architectures can be symbolic, emergent⁴, or hybrid (Vernon, 2014; Ye et al., 2018).

There are many examples⁵ concerning the application of such architectures. Notably, Cognitive architectures have been implemented in ITS (e.g. the symbolic architecture *ACT-R*, see (Corbett et al., 2000; S. Ritter et al., 2007), as well as virtual intelligent agents (e.g. the hybrid architecture *CLARION*, see (Lynch et al., 2011)) and developmental robotic agents (e.g. *Darwin*, see (Krichmar & Edelman, 2005; Reeke & Krichmar, 2005)).

Unfortunately, describing types and corresponding examples of cognitive architectures in greater detail would go beyond the scope of this master’s thesis. Except for cases in which a specific cognitive architecture is mentioned explicitly (e.g. *ACT-R* in chapter 4.1.2.1, subsumption architecture in chapter 3.12) this topic will not be addressed further.

Another way of modelling (embodied) cognitive processes is to employ dynamical systems (see next chapter).

4 In his book ‘Artificial Cognitive Systems’ Vernon (2014) uses the term ‘emergent’ to refer to the non-functionalist paradigms subsuming connectionist, dynamic, and enactivist modelling of cognition in artificial systems.

5 For a general overview of symbolic, emergent, and hybrid architectures see (Vernon, 2014). For a more recent review of symbolic, emergent, and hybrid cognitive architectures see (Kotseruba & Tsotsos, 2020; Ye et al., 2018).

3.6 Dynamical Systems Perspective

Within cognitive science, dynamic systems theory is a broad meta-theoretical framework that has been ‘imported’ from physics and theoretical mathematics. Attempts to account for cognition by means of dynamical systems theory arose due to scepticism or rejection of representational approaches to cognition (S. Walter, 2014a). Dynamicists argue that this framework is better suited for situating cognitive or psychological phenomena, i.e. it may account for them more comprehensively and/or accurately than information-processing-based accounts of cognition (ibid). According to Fusella (Fusella, 2013), the (human) mind is

an abstraction for the neurological underpinnings in the brain and these are not machines they are biological organs made up of cells and organic molecules and they are part of the natural world and could arguably be better accounted for by a meta-theory that has been successful in capturing the diverse natural phenomena (Fusella, 2013, para. 4).

Such dynamical modelling served as the basis for the dynamical systems metaphors in the philosophy of mind and cognition. According to Tim van Gelder (van Gelder, 1995, 1998), dynamical systems theory is a combination of the metaphysical thesis that cognitive systems are dynamical systems, and the epistemic thesis that cognitive systems can be described using the terminology and methods from dynamic systems theory:

The dynamical hypothesis has two major components: the nature hypothesis (cognitive agents are dynamical systems) and the knowledge hypothesis (cognitive agents can be understood dynamically). [...] The dynamical hypothesis (DH) is the unifying essence of dynamical approaches to cognition. It is encapsulated in the simple slogan, cognitive agents are dynamical systems (van Gelder, 1998, p. 615).

Furthermore, dynamical systems emphasise the relevance of time, i.e. that systems change over time, as well as a close interrelationship of cognition and action (Beer, 2000; van Gelder, 1998; S. Walter, 2014a; L. M. Ward, 2002). The brain, the body, and the environment are not only interactive components of cognition, but comprise a complex coupled system. Attempting to describe one of these components in isolation, without considering the simultaneous change within the other components of the system, is not possible or meaningful. Cognition is no longer ‘in the brain’ but encompasses a dynamic system (Schöner, 2013; S. Walter, 2014a). The complex interactions between brain, body, and environment give rise to cognition; it is the product of the complex interaction between the system’s components (ibid).

In the above sense, a dynamical system is rather broad and abstract construct. Formally, a dynamical system is a:

mathematical object that unambiguously describes how the state of some system evolves over time. More formally, a dynamical system is a triple $\langle T, S, \phi_t \rangle$ consisting of an ordered time set T , a state space S , and an evolution operator $\phi_t : S \rightarrow S$ that transforms an initial state $x_0 \in S$ at time $t_0 \in T$ to another state $x_t \in S$ at time $t \in T$. (Beer, 2000, p. 92)

How exactly the elements of this triple are operationalised (i.e. implemented in a concrete dynamical systems model) can vary:

- The time T
 - can be discrete or continuous.
- The state space S
 - can be numerical, symbolic, continuous, discrete, or hybrid, and
 - may be either finite or infinite-dimensional (depending on the number of variables required to describe the state of the system)
- ϕ_t may be
 - given explicitly or be defined implicitly,
 - either deterministic or stochastic, and
 - may receive inputs (non-autonomous) or no inputs (autonomous system) (Beer, 2000).

In short, a dynamical systems model is typically expressed as a set of equations (either differential or difference) that describe how the system's state changes over time (ibid).

Different to symbolic or connectionist models, the explanatory interest or focus of dynamical systems lies on the space of possible trajectories and its structure, as well as the system's internal and external forces that shape the particular trajectory as it develops over time (ibid). A crucial difference between symbolic, connectionist, and dynamic models is that inputs to a dynamical model serve as a source of perturbation of the system's intrinsic dynamics; i.e. the internal state of a dynamical system is not uniquely specified by some external input with reference to some external state (ibid).

Formal examples of dynamical systems models are cellular automata, finite state machines, and Turing machines (ibid). A more concrete example concerning what kind of problems or phenomena may be investigated with this approach, is the A-not-B error. In short, the A-not B error (as formulated in 1954 by Piaget, see (Bond, 2012; Piaget, 1954)) is a specific task used to investigate object permanence in infants (age 7-12 months) within the area of developmental psychology.

The infant is placed in front of two opaque containers (container A and B) and trained to reliably look for a hidden object in container A after seeing that another person hid it in there. Then, the object is instead hidden in container B and the infant has to wait for a given period of time (ibid). Although they watch the object being placed in B, most infants will still reach for container A (hence the name A-not-B error). Piaget (1954) interpreted this as a lack of object permanence in the development of infants in this age group. However, other studies contested this interpretation and showed that the error is context-sensitive in rather specific ways (Beer, 2000). Infants are far less likely to make the A-not-B error when they are interested in the object (e.g. a special toy or sweets), if they are not made to wait for a bit after watching the object being hidden in container B, as well as when the two opaque containers do not look identical, if they have to only look at the correct container but do not reach for it, or when the infants changes its posture between switching from the hidden-in-A to the hidden-in-B condition (ibid). This circumstances, i.e. the specific patterns of context-sensitivity, can be accounted for and modelled as a dynamical system. Using the dynamical theory of motor programming (Thelen et al., 2001), a dynamical systems approach can be used to model the motor actions – in the sense of embodied (see chapter 3.7) cognition – involved in the A-not-B task (Beer, 2000; Thelen et al., 2001).

To name another research area, the dynamical systems approach can be used for the dynamic modelling of artificial ‘nervous systems’ in autonomous agents within evolutionary robotics (for a detailed review of concrete studies on the matter, see (Beer, 2000)). This makes dynamical systems suitable for situated accounts of cognition (see chapter 3.6), because the agent’s body (and its ‘nervous system’), the agent’s environment, and its actions towards or in this environment can be conceptualised and modelled as coupled dynamical systems (ibid).

3.6.1 (Social) Action and Interaction in Dynamicism

Opponents of the dynamicist approach criticise that dynamical systems can only describe basic cognitive processes in dynamic real-time interaction (S. Walter, 2014a) But it is exactly this aspect that is of interest with regard to social interaction and its role in or relation to cognition. As an *abstract meta-theory*, dynamical systems may be considered as a (initial) theoretical step towards a basis for describing social interaction and its role in cognition, as this framework pays tribute to the body as well as its interaction with the environment. From a conceptual point of view, dynamic systems theory provides the *theoretical language and terminology* for understanding cognition as a self-organised system. It informs conceptions of cognition in situated approaches to cognition (S. Walter, 2014a). Computational and/or brain-centred approaches to cog-

dition generally locate such problems solely within the individual organism. Situated approaches, on the other hand, employ conceptions of cognition in terms of agent-environment systems, including basic processes and more complex tensions or conflicts between agent and environment, which are understood in relational terms (Schöner, 2013). In this sense, dynamical systems metaphors and/or models are used in the context of situated cognition (cf. (Lamb & Chemero, 2018; Richardson & Chemero, 2014), e.g. to adequately capture the concepts of coupling and autonomy within enactive conceptions of embodied (inter)actions within (social) cognition.

3.7 *Situated Cognition and 4E-Approaches to Cognition*

The following chapter will provide an overview of more recent approaches to cognition, namely situated cognition, and discuss the concept of social interaction within these approaches.

In this paradigm, cognition is understood to be ‘situated’ in the sense that cognitive processes are assumed to not be ‘bounded’ by the brain as a central processing unit (as in classical approaches to cognition). Rather, cognitive processes are dependent on the body and the interaction with the natural, technological, and social environment (S. Walter, 2014a).

In a nutshell, situated cognition denotes that cognitive processes are not limited to the brain/neural processes and heavily rely on the body and its specific physical constitution (embodied cognition) and the environment a cognitive agent (or system) is embedded in (embedded cognition). While the classical/computational and connectionist approach has an internalist conception of cognition (‘intracranialism’, cf. (Adams & Aizawa, 2010; S. Walter, 2014b)), all situated approaches have an anti-internalist perspective on cognitive processing. Situated cognition proposes that the non-neural body and the physical environment are essential to explaining individual cognition (Newen et al., 2018). Thus, cognition extends beyond the brain and/or body into the technological and social environment (extended and/or distributed cognition, respectively) and arises through the interactions with these environments (enactive cognition) (ibid).

Although it may seem relatively clear at first glance what is meant by ‘situated’ or ‘4E’ cognition, i.e. embodied, embedded, extended and enactive cognition, a closer examination of cognitive science literature reveals that this is not the case. Some of these terms seem to be used synonymously, while others act as placeholders for a whole set of approaches that only sometimes seem to be compatible. While cognitivism and computational approaches to cognition provided a clear

framework for the analysis of cognition that includes unified concepts, theories, and models, situated cognition as a research programme is still rather ill-defined. Situated cognition is a loose-knit family of various approaches within cognitive science, which are, however, *not* based on a common or shared terminology or clear-cut conceptualisation of cognition (S. Walter, 2014b; R. Wilson & Clark, 2009). While they do share the common idea that cognition is not limited to the brain and neural processes (or even the body) in terms of a centralised unit of operation, and that cognition is closely coupled to action and the environment in terms of brain-body-world interactions, multiple – and not necessary compatible – variants of embodied, embedded, extended, and enactive cognition exist in parallel. Moreover, even if this may initially seem to be the case, 4E cognition is only partially a complete departure from the core concepts of the preceding classical paradigms, i.e. representational and/or computational conceptions (S. Walter, 2014b). Many theories associated with situated cognition do lean towards conceptualisations of cognition in dynamic terms, i.e. anti-computational and/or non/anti-representationalism⁶, but this is not the case in all of them.

Based on the following generalisations (see list just below), the terms ‘situated cognition’ and ‘4E cognition’ will be used synonymously as an umbrella term for the approaches of embodied, embedded, extended, and enactive cognition (‘4E’ cognition) in the context of this master’s thesis, with ‘situated cognition’ furthermore being a broader category that comprises not only the 4E-approaches, but also distributed cognition.

Although not all researchers commit to these assumptions, or do so to a varying degree (which will be outlined in detail in the following chapters), according to Walter (2014b) and Newen et al. (2018), the assumptions guiding the basic understanding of situated cognition can be generalised to:

1. Cognitive processing is *not* limited to the brain, thus involving extra-cranial, bodily processes in terms of a *brain-body unit*.

6 Note that the terms ‘non-representational’ and ‘anti-representational’, and ‘non-computational’ and ‘anti-computational’, respectively, have been encountered during the literature review for the purposes of this master’s thesis. Although they seem to imply some kind of distinction, they also seem to sometimes be used synonymously. The interpretation of those terms within the context of this master’s thesis is the following. A theoretical concept of or perspective on cognition can be based on anti-representational and/or anti-computational stance; that means that representation (and computations thereof) as an explanatory means for cognition are rejected. Non-representationalism and/non-computationalism implies that a certain theory, model, or process of cognition does not involve or exclusively rely on representations and/or computation, without claiming that cognition in general (all cognitive processes) do not involve or rely on representations and/or computation. In this sense, a theory or concept within the context of this thesis can be described as non-representational/non-computational without necessarily also implying an anti-representational/anti-computational stance towards cognition.

2. Cognitive processing involves the natural, technological, and/or social environment, i.e. it involves extra-bodily aspects in terms of a *brain-body-environment unit*.
3. In such a ‘cognitive unit’, cognitive processing can be either a:
 - i. partially *dependent*/co-dependent (in the sense of a diachronous non-constitutional relationship, cf. (S. Walter, 2014b)) process, or be
 - ii. partially *constituted*/co-constituted by extra-cranial and/or extra-bodily process (see also coupling-constitution fallacy, chapter 3.9.2).

Furthermore, conceptions in which:

1. Cognition is *co-dependent upon extra-cranial/bodily processes* are associated with *weak embodiment* (see chapter 3.7.1).
2. Cognition is *co-constituted by extra-cranial/bodily processes* capture the idea that cognition is *strongly embodied* (chapter 3.7.1).
3. Cognition is *co-dependent upon extra-bodily processes* (i.e. the natural, social, and/or technological environment) are generally referred to as *embedded cognition* (chapter 3.8).
4. Cognition is *co-constituted by extra-bodily processes* (i.e. the natural, social, and/or technological environment) are associated with *extended cognition* (chapter 3.9).
5. Cognition is *either co-dependent upon or co-constituted by extra-bodily processes* that comprise unit(s) of interacting agents who utilise natural and/or technological resource are generally considered to be instances of *distributed cognition* (chapter 3.10).
6. Cognition is ‘localised’ in neither the brain, nor the body, nor in the natural, social, and/or technological environment, i.e. cognition involves action and interaction, thus being a *relational* process of *sense-making between autonomous systems*, are generally considered to be approaches of *enactive cognition* (chapter chapter 3.11), in which:
 - i. the question of the ‘where’ of cognition is replaced with a relational understanding of cognition with a constitutional, rather than a dependence-claim (cf.(S. Walter, 2014b)). Specifically:
 - a cognitive process that is *co-dependent* on the ability or disposition to act is *weakly enacted*, and
 - a cognitive process that is *co-constituted* by the ability or disposition to act is considered to be *strongly (or radically) enacted* (cf. (Newen et al., 2018)).

Whereby:

All situated cognition approaches take on an *anti-interalist* position, but

- embodied, embedded, and extended cognition approaches can be based on *either* a computational/representational or an anti-computational/anti-representational (sometimes also referred to as ‘dynamic’) understanding of cognition, and *only*
- enactive cognition is purely an *anti-internalist, anti-computational, and non/anti-representational* perspective, with
- the idea of distributed cognition – at least the ‘classical’ variant based on the work of Hutchins (Hutchins, 1995a) – being the *computational/representational* ‘counterpart’ of non- and/or anti-representational enactive approaches (Newen et al., 2018; S. Walter, 2014b).

As will be outlined in the following chapters, the concept of social interaction and its relevance for (social) cognition has found its way into embodied, embedded, as well as extended, and distributed accounts of cognition, which treat the social environment and actions towards and with other cognitive agents (i.e. social interactions) as a key factor of explanations of (individual) social cognition and the brain-body-world interactions it involves. Some proponents of enactivist accounts of cognition even go as far as claiming that social interaction is more than a mere context of social cognition, stating that social interaction can complement, replace, and possibly even constitute social cognition (e.g. (De Jaegher et al., 2010a; Froese et al., 2015a)). Before discussing how the concept of social interaction can be understood within situated cognition, it is necessary to more closely examine the approaches generally associated with situated/4E cognition. Therefore, I will first clarify what is meant by the individual paradigms or approaches, specifically embodied, embedded, extended, distributed, and enactive cognition, and then examine social aspects and conceptualisations of social interaction within these paradigms, respectively.

3.8 Embodied Cognition

As a concept, ‘embodiment’ has been used in cognitive science and (nouvelle) AI since the 1980ies and is considered to be one of the key concept that distinguishes situated approaches from ‘disembodied’ computational theories of classical cognitive science (Ziemke, 2003). Embodied cognition postulates that intelligence or cognition requires a body, rather than existing in the form of abstract representations which are processed by the brain in the sense of a ‘central unit of operations’ (ibid). It highlights the role of the body in cognition to produce actions, as well as agent-environment interaction.

Several authors have tried to clarify what ‘embodied cognition’ means and have discussed the concept of embodiment on different levels of abstraction. Some of this contribution – that seem to be of interest within the context of this thesis – will be reviewed in the following chapters.

3.8.1 Variants of Embodied Cognition

On a general level, Clark (1999) describes two forms of embodiment: 1) simple embodiment and 2) radical embodiment. These approaches mainly differ in their leaning towards anti-representational perspectives on cognition. *Simple embodiment* maintains the fundamental ideas of the computationalist/functionalist approach. Here, embodiment acts as a kind of constraint for ‘inner’ processing in which the body influences the representational capacities of the cognitive system or organism. This notion appears to correspond to being ‘weakly’ embodied as outlined in the generalisation regarding situated cognition in the previous chapter. In contrast, *radical embodiment* views the body not only as an influencing, but rather as a *constituent* part of the cognitive system that goes beyond the brain (Clark, 1999). According to this perspective, cognitive processing involves an interplay between the brain, the body, and the world (Clark, 1999; Lindblom, 2015a, p. 20; Shapiro, 2014). This notion appears to correspond to being ‘strongly’ embodied as outlined in the generalisation regarding situated cognition in the previous chapter.

In a similar vein, Kiverstein (2012) points out that current variants of embodiment mainly differ with respect to the question whether such theories still retain computational and information-processing-based ideas about cognition. He differentiates three forms of embodiment: 1) body-functionalism, 2) body-conservatism, and 3) body-enactivism (Kiverstein, 2012). While the latter type, i.e. *body-enactivism*, entirely rejects the concepts of representations and computation – claiming that such an approach cannot provide an agent with common-sense knowledge about its environment – the first two types retain computational ideas (ibid). Embodiment as *body-functionalism* holds that the body merely implements the computational processes that underpin cognition; the body links external information/input to the internal “biological machinery” (Kiverstein, 2012, p. 740). *Body-conservatism* more strongly admits the relevance of the body for cognition and views the body as a means to contribute to information-processing by ‘supplying’ inputs to the brain, as well as executing motor-instructions ‘sent out’ by the brain (ibid). It is not clear whether other cognitive systems could ‘supply’ such inputs or if ‘outputted’ actions could be directed towards others (i.e. other agents) in the environment, as the focus of these approaches still seems to be on the individual and its internal processes in terms of neural/brain-processes. Notice

that not all embodied approaches emphasise the role of the environment in cognitive processes, or at least not in form of cognitive extension (see chapter 3.9).

The third form, body-enactivism, is concerned with how the body enacts or produces meaning in a certain environment in a situated manner, i.e. in a way that is meaningful to the agent given its environment (ibid). Such processes of ‘meaning-making’ could include other agents, i.e. by means of social interaction (see chapter 3.11, 3.11.3)

In a similar fashion, Gallagher (2011a) differentiates between minimal embodiment (see also ch X), embodied functionalism (similar to Kiverstein’s body-functionalism), biological embodiment (similar to the notion of organismic embodiment, see below), embodied semantics (concerned with linguistic aspects), and radical or enactive embodiment (alike to Kiverstein’s body-enactivism, see also chapter 3.11.1.3).

In contrast, Wilson (2002) does not focus on the ways in which different forms of embodiment depict the body’s contributes to cognition, but rather on the basic claims an embodied understanding of cognition rests upon:

1. Cognition is *situated*; it involves perception and action that takes place in a real-world environment.
2. Cognition is *time-pressured*, it has to be understood in the context of real-time interactions with the environment.
3. Cognition *exploits the environment*. It is used to ‘off-load’ cognition, i.e. to ‘hold’ or manipulate information, and
4. the *environment is part of* the cognitive system. Thus, the brain ‘on its own’ is not a meaningful unit of analysis and scientists have to consider the environment and its role in cognitive activities.
5. *Cognition is for action*; its function is to guide action and cognition has to be understood in relation to the production of situation- or context-appropriate actions and behaviours.
6. Cognition is *rooted in interactions* with the environment and even ‘decoupled’ offline cognition is body-based (ibid).

If embodied cognition is situated, serves action, and includes the environment – and thereby extends the ‘cognitive unit’ and, thus, the of analysis beyond the individual – cognitive activities may include real-time actions towards others in the same environment. As such, embodied cognition may be a promising framework for social interaction. However, several authors point out that

it is not at all clear what it means for a cognitive agent, entity, or system to be ‘embodied’ (see e.g. (Alessandroni, 2018; S. Gallagher, 2011a, p. 20; Kiverstein, 2012; Ziemke, 2003)), or what kind of ‘body’ is required for cognition to be embodied, or even whether it requires a specific type of body at all (Ziemke, 2003). Furthermore, as pointed out in chapter 2.1, many conceptions focus on human cognition; they are not concerned whether or under which conditions an artificial agent, the interaction between multiple such agents, or between artificial and human agents, can be embodied. A fruitful notion of embodiment needs to be restrictive in order to allow for sufficient differentiation between those cognitive systems that are embodied, and the ones that are not. A too broad notion of embodiment could include all possible systems situated in any complex environment (cf. (Riegler, 2002)). Different forms or variants of embodiment imply different possibilities and constraints with regard to social interaction. What form such a thing as ‘embodied’ social interaction could take on, or what kind of bodies (and their respective environments) could be capable of taking part in these social interactions, is determined by what it means to be ‘embodied’. Depending on which forms or notions of embodiment are considered, conceptualisations of social interaction could either be restricted to human-human social interaction, or include interactions with artificial or robotic agents, other non-human entities, or even artefacts. Furthermore, it is not clear in which way the body contributes to cognition, or whether this contribution does still involve some degree of information processing based on computational ideas and representations (cf. Kiverstein 2012).

Building on the work of Wilson (2002), Ziemke (2003) focusses on the question of what type of body is capable of embodied cognition. He identifies six notions of embodiment:

1. Embodiment as structural coupling/structural embodiment,
2. historical embodiment,
3. physical embodiment,
4. organismoid embodiment,
5. organismic embodiment, and
6. social embodiment (Ziemke, 2003).

With the exception of organismic embodiment, all of the above notions could include both human and artificial agents, some of them also other non-human entities, as well as (social) interactions between them. The broadest notion, *embodiment as structural coupling*, describes systems as embodied if they are *structurally coupled* (cf. (Maturana & Varela, 1980, 1987)) to their environ-

ment. This means that a system is embodied in an environment, if bidirectional perturbatory channels exist between the system and its environment. This could encompass human as well as artificial cognitive systems. Albeit providing a relatively clear definition of embodiment, it is also too broad to effectively distinguish between cognitive and non-cognitive embodied systems, both physical or non-physical (Ziemke, 2003). As Mohammad and Nishida (Mohammad & Nishida, 2015) point out: “Even a rock on a river’s floor is structurally embodied.” (Mohammad & Nishida, 2015, p. 209).

Another notion of embodiment is the one of *historical embodiment*, which highlights aspects of development and adaptation. An embodied cognitive system is affected by the environment it develops in. Its current state is the result of a specific history of interactions with its environment and the adaptation of these interactions to this environment. According to Riegler (2002), a system is embodied if it gained competence in its environment. Essentially, this notion of embodiment applies to all living systems, but does not necessarily exclude non-biological systems. *Physical embodiment* requires embodied systems to have a physical body as intelligence or cognition cannot “merely exist in the form of an abstract algorithm” and thus requires a “physical instantiation” (R. Pfeifer & Scheier, 1999, p. 649). In this sense, all living systems are embodied. Artificial systems require a similar body, e.g. as a physical robot or a simulated agent body (cf. classification of autonomous agents in (R. Pfeifer & Scheier, 1999)). According to Ziemke (Ziemke, 2003), this notion of embodiment is more restrictive than structural coupling. It still does, however, possibly include inanimate, ‘non-cognitive’ objects, such as chairs or tables. A more restrictive understanding of physical embodiment is the view that embodied systems require a body equipped with some kind of sensors and motors that physically ‘ground’ them in their environment (Brooks, 1991b, 1991a; Ziemke, 2003). Both historical and physical embodiment can be considered special cases of structural coupling, but don’t necessary include or exclude each other. For example, a living systems (e.g. a child) may be physically and historically embodied, but a physical system (e.g. an automated vacuum cleaner) may not be historically embodied (Riegler, 2002; Ziemke, 2003).

Another notion of physical embodiment, *organismoid embodiment*, holds that cognition is limited to organism-like bodies, i.e. physical instantiations that, to some degree, resemble sensorimotor abilities/capacities of living bodies. This notion of embodiment encompasses all living and artificial organism-like systems, including human and human-like cognitive systems. However, Ziemke notes that “it might very well be true that having hands, arms, legs, etc. is essential to hu-

man (-like) embodied cognition, but this notion in itself does not tell us *why* [emphasis added] this might be so” (Ziemke, 2003, p. 1308).

Organismic embodiment depicts cognitive systems as physical and organismoid-like, but is restricted to organisms in the sense of *living* bodies, which are autonomous and autopoietic (cf. (Maturana & Varela, 1980, 1987)), thereby excluding mechanic or otherwise artificial cognitive systems/bodies.

Finally, the notion of *social embodiment* centres on the role of embodiment in social cognition, including embodied actions in social environments. Some authors (e.g. (Alessandroni, 2018; Lindblom, 2015a; Lindblom & Ziemke, 2008)) view this form of embodiment as a distinct set of embodied cognition theories that is “orthogonal” (Ziemke, 2003, p. 1309) to other notions of embodiment. But Ziemke (2003) also points out that this notion appears compatible with the notions of organismoid embodiment and body-based offline cognition. It may also be compared to embodiment as structural coupling if the environment is a social environment (Mohammad & Nishida, 2015).

Many authors (e.g. (Barsalou et al., 2003; Lindblom, 2015a; Lindblom & Ziemke, 2008; Niedenthal et al., 2014) focus on human cognition and human-human interactions. However, as long as the underlying notion of embodiment is e.g. organismoid embodiment or embodiment as structural coupling, embodiment is not limited to living systems and has been applied in AI and (social) robotics research (see e.g. (Dautenhahn, Ogden, et al., 2002; Dautenhahn, 2011; Mohammad & Nishida, 2015)).

The next chapter will focus on different contributions to social embodiment with regard to the role of the body in social interaction and cognition.

3.8.2 (Social) Action and Interaction in Embodied Cognition: Social Embodiment

Broadly speaking, embodied cognition is concerned with the way the body and its interactions with the environment shape cognition. Embodied *social* cognition moves its investigation of embodied cognition beyond interactions of an embodied organism/system with its physical environment, to the interactions between agents and their social environment, i.e. how cognition is *socially* embodied.

Within socio-cognitive research, one influential contributions explicitly using the term “social embodiment” is the work of Barsalou and colleagues (2003).

Barsalou et al. (2003) focus on human cognition and human social interactions, rather than different notions of embodiment or what kind of body is required for what kind of cognition. They

define social embodiment as “states of the body, such as postures, arm movements, and facial expressions, arise during social interaction and play central roles in social information processing” (Barsalou et al., 2003, p. 43) and state that “embodied theories represent knowledge as partial simulations of sensory, motor, and introspective states” (Barsalou et al., 2003, p. 44). It should be noted that for Barsalou and colleagues (2003) introspective states include events that are perceived inside the body or mind and lack counterparts in the external world, e.g. affects, emotions, or beliefs. Depending on the situation, embodiment can range from motor simulations to partial or full executions, and such embodiments constitute the core of social information processing (ibid).

3.8.2.1 Social Embodiment Effects

In their work on embodied theories of knowledge, Barsalou et al. (2003) propose four ‘social embodiment effects’. They address cognitive, bodily, and affective states in their description of these social embodiment effects:

1. Perceived social stimuli don’t only produce cognitive states but also bodily states, e.g. when adopting a more upright posture when receiving a compliment.
2. Observing bodily states in others causes mimicry of bodily states in the observer, e.g. when automatically mimicking someone’s facial expressions or body posture during conversation.
3. Bodily states in the self also produce affective states in the self, e.g. experiencing more pride in an upright position than in a slumped one.
4. Performance in social interaction is modulated by the compatibility of bodily and cognitive states, e.g. nodding when agreeing with someone in relation to a positive preference towards someone or something.

(cf. (Barsalou et al., 2003; Lindblom & Ziemke, 2008; Ziemke, 2003))

These social embodiment effects suggest a strong relationship between cognitive and bodily, as well as affective, states in social situations, especially when perceiving others in direct (social) interaction. Research on embodied social perception, emotions, and attitudes from the field of social psychology too indicates a relationship between bodily movements and posture, e.g. nodding one’s head in agreement or, conversely, shaking one’s head in disagreement in relation to positive or negative predispositions towards objects, respectively (Lindblom & Ziemke, 2008; Niedenthal et al., 2005). Furthermore, it has been suggested that the mirror neuron system (often in relation

to simulation theories of the mind) provides the neurological basis for the above mentioned social embodiment effects (ibid).

3.8.2.2 Metaphors for Social Interaction and Cognition

Information-processing-based approaches, including the work of Barsalou et al. (2003) – generally view (social) cognition as a process that primarily takes place in the skull or brain. The body functions as a kind of “input and output device” (Lindblom & Ziemke, 2008, p. 50). This view implies that agents relate to each other through the use of explicit internal representations, similar to the way agents relate to other parts of their external environment. In this view, the body is a “physical interface” (Lindblom & Ziemke, 2008, p. 50) between the external world and ‘internal programs’. Bodily aspects, such as eye gaze, body posture or gestures that occur during communication or other interactions are considered to be visual outcomes of ‘inner’ mental processes that are – through processes of perception – ‘transmitted’ from one cognitive agent/system to another. (ibid)

An alternative contribution regarding the role of the body in social situations is the work on embodied social cognition by Lindblom and Ziemke (2003) and Lindblom (Lindblom, 2015a, 2020) Aiming to clarify the role of the body in social interaction from an embodied cognition perspective, Lindblom and Ziemke relate the concept of social interaction to intersubjectivity as “the cradle of social interaction” (Lindblom & Ziemke, 2008, p. 51). Claiming that “information is not a predefined and discrete entity which can be sent, through signals, from one agent across time and space to another agent in the form of internal mental representations” (Lindblom & Ziemke, 2008, p. 51), they suggest a shift in perspective from the *information-transmission metaphor* to the *dance metaphor*. Rather than understanding internal (social) information as being externally transmitted back and forth between agents, the *dance metaphor* focuses on the emergence of information – or meaning – in dyadic agent interactions. Embodied actions⁷ are no longer conceived of as linear or discrete processes, but as dynamically emerging in a particular context of interactive social situations.

Furthermore, Lindblom (2015a) connects the dance metaphor to the basic ideas of *distributed cognition* (for this paradigm, see chapter 3.10) on a *methodological* level: the unit of analysis is extended beyond the individual organism/agent and includes material artefacts and other organisms/agents, as the cognitive system is distributed across the (social) environment. The benefit of

⁷ Lindblom (2015a) emphasises the use of the term ‘action’ instead of ‘behaviour’ to indicate that embodied actions are intentional movements, i.e. “socially embodied actions are a set of movements that have agency for the actual agent” (Lindblom 2015a, p. 284).

such an approach is that cognition becomes directly observable. Aspects of (social) cognition that have been considered to be ‘inside’ (i.e. the brain) can now be observed and studied through socially embodied actions such as gestures facial expressions, speech, and the like, e.g. visible embodied actions in unfolding dialogue (ibid). This understanding of social interaction as a generalisation to intersubjectivity, thus, focusses on the experience of being embodied in the material/physical *and* the social environment. This framework for embodied social cognition does, however, not necessary reject mental concepts, such as intentions or beliefs. But it questions the central role those concepts play in information-processing approaches, i.e. their role as an underlying requirement for cognitive processes (information-transmission metaphor). The framework of embodied social cognition rather suggests that cognitive processes emerge from and are grounded in embodied (inter)actions (Lindblom, 2015a; Lindblom & Ziemke, 2008).

3.8.2.3 Four Fundamental Functions of the Body in Social Interaction and Cognition

Socially embodied actions do not merely express some kind of internal cognitive processes, but are fundamental aspects of experiencing cognition. Positioning her work within radical embodiment, Lindblom (2015a) draws from theoretical and empirical research from the fields of social psychology, social neuroscience, embodied linguistics, and cognitive anthropology to generalise the role of the *human* body and its sensorimotor processes in social contexts⁸ to four fundamental functions:

- The body functions as social resonance mechanism
 - The body functions as a means and end in social interaction
 - Embodied actions and experiences function as a ‘helping hand’ in shaping, expressing and sharing thoughts
 - The body functions as a representational device
- (Lindblom, 2015a, p. 149)

The body functions as a social resonance mechanism. The function of the body as social resonance mechanism states that cognitive states of interacting agents are reflected in the self and in-between them on a basic level, i.e., they ‘resonate’ in and through the body and its actions. Social embodiment suggest that perceptions in form of embodied social stimuli are not ‘encoded’ into cognitive states or higher knowledge structures, which, in turn, are again ‘decoded’ into executive motor-functions (ibid). Rather, bodily states in themselves are also cognitive or affective states, as

8 Embodied actions are always in relation to others. Here, “social context” – as used by Lindblom (2015) – is meant to indicate the relational aspects of social interaction and cognition, i.e. a relational conception of the (developing) social mind, which is strongly influenced by the work of Mead and Vygotsky (cf. (Lindblom, 2015a; Mead, 1934; Vygotsky, 1978)

also described in the social embodiment effects by Barsalou et al. (2003) and related work from the field of social psychology. There is a bi-directional relation between cognitive, affective, and bodily states, which occurs automatically in direct/online (i.e. when both interacting agents are present) and indirect/offline interaction (i.e. in absence of an interaction partner). Phenomenologically, embodied actions and their experience within the self and in-between interacting agents inform intersubjectivity and understanding others in social contexts. Furthermore, Lindblom (2015a) states that the mirror neuron system serves as the neurological underpinning for the social resonance mechanism.

The body functions as a means and end in social interaction. The function of the body as a means and end is related to the previous function (social resonance mechanism). The ‘resonance mechanism’ states that the body and its sensorimotor processes serve a resonance function, while the function of ‘means and end’ focusses on *how* this is accomplished by addressing the relationship of first-hand (e.g. performing an action) and third-hand (e.g. observing an action) experiences in social interaction (Lindblom, 2015a).

Lindblom (2015a) points out that, while the mirror neuron system may not sufficiently account for the complexity of social interaction, it does link ‘action’ and ‘action perception’, allowing interacting agents to perceive and understand the meaning of observed actions. Thus, the mirror neuron system provides the neurological and embodied basis for social experience and social interaction. Perception and action are close interrelated and “not only bounded by the brain” (Lindblom, 2015a, p. 150). Recognising an action – and understanding the ‘what’ and ‘why’, i.e. the intention of an action – activates the perceptual as well as the motor system and there appears to be a close relation between first-hand and third-hand experiences, which may involve different degrees of embodied simulations (S. Gallagher, 2005; Lindblom, 2015a). Lindblom (2015a) further states that the activation of the mirror neuron system and other parts of the brain form a continuum between the basic-level understanding of actions to more advanced actions, including gestures, speech, and language.

Embodied actions and experiences function as a helping hand in shaping, expressing and sharing thoughts. Gestures and speech complement each other and are crucial to social interaction. With the function of a ‘helping hand in shaping, expressing, and sharing thoughts’, Lindblom (2015a) highlights the role of gestures and their experience – in the sense of embodied actions – in social interaction and cognition. She states that gestures operate both ‘inwardly’ (i.e. in-

tra-subjective) and ‘outwardly’ (i.e. inter-subjectively) during social interaction, e.g. during unfolding dialogue and conversations. Furthermore, gestures may not only serve as embodied externalisations to express an agent’s thoughts and ideas, they may also shape them: “gesturing functions as a vehicle of thought [...] enables people to embody and communicate their thoughts in action” (Lindblom, 2015a, p. 150).

The body functions as a representational device. In relation to the functioning of gestures as embodied actions, Lindblom (2015a) further addresses the role of gesturing as a representational function. She postulates that gesturing functions as a symbol that carries meaning: “the hallmarks of human (spoken) language is its representational capacity, which is, so to speak, the ability to correspond to something else. [...] certain kinds of gesture, portraying representational aspects, are the least provocative and most obvious examples of the body as an (external) representational device” (Lindblom, 2015a, p. 152) She provides the relatively simple example of a twisting hand movement as a symbol for the actual action of opening a jar. As a communicative act, this gesture ‘signifies’ the action of opening a jar and thus, has representational characteristics that involve intent and (cognitive) meaning.

Cognitive processes of social interactions are, at least to some degree, grounded in such ‘embodied representations’, i.e. embodied actions that carry representational content. Furthermore, the perception of another persons actions in online cognition may be enough to gain an understanding of the other person and their intentions, with embodied simulation being an ‘extra step’. The mirror neuron system may be the neurological underpinning of this ability and the “linkage between “action” and “action-perception” might propose a kind of “action representations” that are directly enacted in social interaction” (Lindblom, 2015a, p. 153).

In summary, human social interaction is a relational process, as embodied actions are always in relation to another agent. The four functions of the body emphasize the crucial relevance of embodiment for social interaction and mean to integrate ‘inner’ processes, such as the mirror neuron system or embodied simulations, and ‘outer’ aspects, like posture, gesture, or facial expressions. The ‘social mind’ is assumed to be relational, radically embodied, and situated in a socio-cultural environment. Meaning and intention emerge dynamically as a product of socially embodied (inter)actions between agents and their environments.

On a methodological level, a socially embodied understanding of social interaction and cognition as proposed by Lindblom and colleagues has various implications:

- The unit of analysis changes: from the individual agent and its internal (brain-)processes to visible – and thus observable – embodied actions and the (co-regulated) interactions between agents and/or their environment.
- Embodied actions are *always* in relation to others. If the social dimension of embodied cognition within agent-agent and/or agent-environment interactions is of interest, “it is useless to distinguishing between individual and/or social actions, given that they are two sides of the same coin. [...]” as humans “always are in dialogue—either with our selves or with other individuals” (Lindblom, 2015a, p. 284).
- Meaning and intentions are emergent phenomena that arise during social interactions in an (observable) ‘out there’. Thus, emergent meaning becomes ‘accessible’ to third-person observers of social interactions and can be studied by employing adequate means and methods of analysis (e.g. natural inquiry in Lindblom’s case, cf. (ibid)).
- The ‘social mind’ (in the sense of (Mead, 1934; Vygotsky, 1978)) develops; embodied actions and their respective meaning⁹ are not pre-given; social interaction and cognition is a process, rather than a discrete state. In line with historical embodiment (Ziemke, 2003) every agent has his/her own history, which has to be considered. There is a temporal dimension to the investigation of social interaction and how it unfolds dynamically: “embodied actions are always unfolding, and every action is considered on behalf of the prior actions” (Lindblom, 2015a, p. 284)
- The study of embodied actions is not limited to the interaction between agents, but also includes their environment. The socio-cultural context and the environment are integral parts of social interaction. One should “not examine embodied actions apart from their context of use because the actual situation at hand adds meaning and intentions to the embodied actions” (Lindblom, 2015a, p. 284).
- However, there is a difference between interactions with the physical environment and the social environment, i.e. interactions with other agents are more complex and diverse than interactions with the physical environment.

These methodological implications are, essentially, in line with Wilson’s (2002) claims, but explicitly focus on the social dimension of embodied cognition, with a particular interest in social in-

9 Here, ‘meaning’ could be interpreted as a ‘mild’ form of ‘body-enactivism’ (in the sense of Kiverstein, 2014). In Lindblom’s framework ‘meaning’ refers to 1) the person performing the embodied action, i.e. the embodied action has meaning to the agent. This implies a theoretical “move from the experienced body to the enacted body” (Lindblom, 2015a, p.284). 2) meaning in dynamic communicative acts, i.e. in unfolding interaction and communication with others, e.g. via gestures as ‘representational devices’.

teraction and the role of the body in social cognition. It should be noted that the above notion of social embodiment and the functions of the body in social interaction solely focus on human-human interactions. Whether some aspects can also be applied to the interaction with or between artificial agents/systems, specifically, ITS, will be discussed in part III.

3.9 Embedded Cognition

Generally, proponents of situated cognition are not only concerned with the role of the body in cognition but also with the (natural *and* social) environment the cognitive system is *embedded* in.

From a situated point of view, interactive strategies that manipulate the environment are central to cognition. This holds for human cognition as well as for artificial cognitive systems or agents. A prime example within cognitive science are Brooks's (1991a, 1991b, 2014) reactive 'Creatures' or 'Mrobots' which do not have a central processing unit to produce internal representations of external inputs and use "the world as its own model" (Brooks, 1991b, p. 140). Perception and action are closely related and internal, symbol-based, representations become (at least partially) dispensable. This may be best expressed in the idea that 'perception is for action', or even that 'perception *is* action' (cf. (Noë, 2004)). In this sense, embedded cognition is closely related to the idea of embodiment, as the way in which a cognitive agent can strategically manipulate its environment is not only dependent on the environment, but also constrained by the specific configuration of the body and its sensorimotor capacities.

3.9.1 Embedded vs Embodied vs Extended

Similarly to other 4E-approaches, embedded cognition does not have a clear-cut or unified definition and the demarcation between paradigms – especially in the case of embedded cognition – is not always well-defined. Confusingly, the term 'embeddedness' and 'embedded cognition' are sometimes used synonymously with the terms 'situatedness' and 'situated cognition', respectively (see (Lawrence, 2010; Lyre & Walter, 2013)). As already hinted at above, embedded cognition often seems to be discussed in conjunction with embodied cognition and is closely related to the concept of embodiment. In fact, some authors seem to use the concept of embodiment in a way that latently implies the (social) embeddedness of cognition as well (e.g. (Lindblom & Ziemke, 2008)). While embodied cognition focuses on the role of the body and its specific (physiological)

properties in cognition, embedded cognition is concerned with how cognitive systems can – through their body – manipulate their environments in a way that allows them to use this environment as a cognitive resource. The idea of embeddedness has also been related to – and confused with – cognitive extension and extended cognition (Palermos, 2014; Rupert, 2009; Varga, 2016), as well as distributed cognition, which can be seen as an alternative or bridging perspective on cognition (Hutchins, 2001a; S. Walter, 2014b). While some authors seem to use these terms synonymously or at least in close relation to each other (e.g. embedded and embodied cognition) to conjointly explain a cognitive phenomenon, other theorists state that either one or the other approach may be better suited to explain a certain phenomena, as they are not all compatible (e.g. in the case of extended and embedded cognition, cf. (Rupert, 2009)). Aiming to distinguish embodied cognition from embedded, extended, and distributed cognition, Wilson and Foglia (2017) state that embodied cognitive science appeals to the idea that cognition deeply *depends on* an agent's body and its involvement in both sensing and acting. Embedded cognition focussed on the specific (*co-*)*dependent* strategies organisms use to ‘offload’ their cognition and the ways in which cognitive processes are actively distributed over the physical, social, and cultural environment (ibid). Extended cognition is concerned with the ways cognitive systems themselves extend beyond the individual body and how features of the physical, social, and cultural environment can (temporarily) become a *constitutive part* of an agent's cognitive system (ibid). A similar perspective is taken for the purposes of this thesis. Albeit being closely related to the other 4E-paradigms, especially with regard to the role of action and interaction, embedded cognition will be treated as a distinct paradigm within the context of this thesis.

3.9.2 (Social) Action and Interaction in Embedded Cognition

According to S. Walter (S. Walter, 2014a) three basic ideas are relevant with respect to embedded cognition as a distinct paradigm:

- 1) rather than being mediated by internal representations, perception and action are more closely related than assumed in the classical ‘sandwich model’ of cognition. Cognitive processes can depend on the environment in such a way that an active interaction with the environment renders internal representations (in part) obsolete.
- 2) Based on the idea that internal representations do not mediate perception and action, (some) sensorimotor theories of perception do not only conceive of perception and action as being closely related, but abandon this distinction entirely, i.e. perception is a kind of action (ibid).

3) Essential to this perspective of cognition is the role of the environment, i.e. environmental dependency of cognition and, furthermore, that this environment is not passively perceived, but can be actively exploited.¹⁰

These ideas are closely related to the body (i.e. embodied cognition, see), as the body and its configurations allows interaction with the environment. Note that theorists (e.g. (Clark & Chalmers, 2010a)) who consider the relation between cognition and the environment to be more than a relation of dependency and state that the environment is considered to not only be an external resource than can be actively exploited by a cognitive systems (embedded cognition), but an ‘extended’ constitutive part of the cognitive system, are better classified as theorists of extended cognition (cf. (Varga, 2016; S. Walter, 2014a).

As for the concept of *social* action and interaction within embedded cognition, the idea of embeddedness seems to already incorporate social aspects of the environment.

Considered to be ‘embedded’ are cognitive processes which – to a substantial degree – dependent on the environment. This may include the physical, technological (e.g. devices or tools), and social environment (Newen et al., 2018; S. Walter, 2014a). In the present context, social embeddedness would refer to cognitive processes which depend on the social environment, i.e. cognitive activities that exploit resources in an environment that includes multiple (human or artificial) agents who are capable of performing action towards and interactions with each other. According to Edmonds (1999) “social embedding is a special case of embedding in general - the ‘social’ bit comes from the fact we are dealing with collections of parts that are worthy of being called *agents* [original emphasis]” (Edmonds, 1999, p. 327). In the context of this thesis, the concept of embedded cognition is assumed to include the case of social embeddedness, even when not explicitly referring to it with the ‘social’ prefix.

In the following, two concepts often related to embedded approaches to cognition, i.e. strategies of environmental exploitation, will be discussed: 1) scaffolding and 2) affordances. Furthermore, as both concepts extend the cognitive unit – and thus, the unit of analysis for cognition – beyond

¹⁰ These ideas are closely related to the body (i.e. embodied cognition), as the body and its configurations allows interaction with the environment. Furthermore, note that theorists who consider the relation between cognition and the environment to be *more than a relation of dependency* and state that the environment is considered to not only be an external resource than can be actively exploited by a cognitive systems (embedded cognition), but an ‘extended’ *constitutive* part of the cognitive system, are better classified as theorists of extended cognition (see e.g. (Rupert, 2009; Varga, 2016)). In a nutshell, embedded cognition holds that (human) cognition routinely exploit structures in the environment to reduce cognitive load, whereas extended cognition would further this understanding by claiming that such external structures or resources in the environment can also become a constitutive part of an agent’s extended cognitive system.

individual and/or internal cognitive processes into the (social) environment, the next chapters also briefly describe selected conceptions of social scaffolding and social affordances with regard to the topic of social interaction.

3.9.2.1 Scaffolding

In the context of cognitive science and, particularly, the environmental dependency of cognition, ‘scaffolding’ (in the sense of (Clark, 1997a)) generally refers to a specific kind of (actively) using resources in the environment, i.e. a metaphorical ‘scaffold’ that is used to reduce the workload of information processing when solving cognitive problems.

An illustrative example – that is sometimes also used in the context of extended cognition – for the exploitation of the environment during cognitive activities is the game ‘Tetris’. When playing the computer game Tetris, it is easier for the player to actually rotate the pieces to quickly figure out where and how they fit, rather than mentally rotating them to plan how to fit them into the existing pieces. That is, during cognitive activities, the environment is actively exploited to carry cognitive load and solve the problem of rotating shapes to fit them together (Clark, 1997a; S. Walter, 2014a). Similarly, when playing the game ‘Scrabble’, a certain number of letters has to be rearranged to form a word. It is more efficient to solve this problem by physically arranging and rearranging the letters to form a meaningful word, instead of imagining all the possible combinations of letters (S. Walter, 2014a).

As mentioned just above, scaffolding – in terms of *exploiting an external structure* – has been associated with both embedded and extended cognition. The demarcation between theoretical approaches concerning the concept of scaffolding are not clear-cut, since the 4E-movement within cognitive science too is a loose-knit family of more or less compatible or complementary approaches that do *not* share common core concepts or definitions. In the context of this master’s thesis, scaffolding will be primarily discussed in relation to embedded cognition as one of main theoretical approaches subsumed under the umbrella term of ‘situated cognition’.

The concept of scaffolding is based on the idea that the environment of a cognitive system is of a specific nature; a different type of environment therefore leads to a different possibility of scaffolding. In this respect, scaffolding would still be compatible with classical computational approaches, since a different type of input leads to different cognitive processing (S. Walter, 2014a). A *situated* perspective on embedded cognition implies that the *environmental dependency* must be of a different kind. Cognitive systems do not only passively receive and process information

form the environment in terms of input-output relations. They actively use resources in the (physician, social, and technological) environment, and even intentionally structure the environment in a certain way that allows scaffolding (ibid). Such a kind of environmental dependency and manipulation is based on (dynamic) interactions with the environment. Cognitive systems can actively exploit their environment to reduce cognitive load by ‘offloading’ it into the environment (ibid). Thus actions towards and interactions with the environment are crucial to successfully exploiting the environment during cognitive activities via scaffolding.

3.9.2.2 Social Scaffolding

If embedded cognition is closely linked to embodiment, and does not only pertain to the physical, but also the social environment, then (social) cognition may too involve scaffolding rooted in the interactions with (human or artificial) other agents: *social scaffolding*.

Child-caretaker interactions can serve as an example for such a scaffolding. When a child learns gestures through interactions with caretakers, the social environment can function as a social scaffold for developing and using pointing in terms of an intentional action (Lindblom & Ziemke, 2008). When discussing embodied actions and the role of the social environment for developing them, Lindblom and Ziemke refer to Vygotsky’s (1978) example of gesture development in children:

Initially, it is only a simple and incomplete grasping movement directed towards a desired object, and is only constituted by the child’s bodily movements, and nothing more. When the caretaker assists the child, the meaning of the situation itself changes. [...] The individual gesture ‘in itself’ becomes a gesture ‘for-others’. The caretaker interprets the child’s reaching movement as a kind of pointing gesture, resulting in a socially meaningful communicative act, whereas the child at the moment is not aware of its communication ability. After a while, however, the child becomes aware of the communicative function of the performed movements, and then begins using referential gestures towards other people. (Lindblom & Ziemke, 2005, p. 89)

In such a scenario, the caregiver provides scaffolding to the child so it can develop the ‘for others’ gesture, and the social interactions between child and caregiver scaffold the child into the socio-cultural environment (Lindblom & Ziemke, 2005; Varga, 2016; R. a Wilson, 2005). This example illustrates that (human) cognition can be dependent not only on the physical and technological environment, but also on cultural and social scaffolds, as social interactions and social practices can scaffold cognition (Hutto & Myin, 2012; D. Ward et al., 2017).

Whether (social) cognition involving interactions between artificial agents and/or between hu-

mans and artificial agents involves a form of social scaffolding in terms of (observable) agent-environment interactions will be discussed in chapter 5.3.

3.9.2.3 Affordances

In close relation to the idea that ‘perception is for action’, Gibson’s (1986a) *non-representational stance* in his ecological theory of perception and, specifically, the concept of affordance, depicts another perspective on the environmental dependency of cognition. Essentially, ‘*affordances*’ are opportunities for action provided or offered by the environment (Borghi, 2018; Gibson, 1986b). Gibson's theory of affordances (1986b) was a counter-movement to the computation theories of perception. According to Gibson's theory, perception does not serve to provide representations independent of action, because perception is already directly related to action (ibid). Crucially, the environment offers possibilities of action that agents perceive *directly*. Affordances are acted upon through 'direct perception', i.e. without any kind of mediation through internal representations (Gibson, 1986b, 1986a). Thus, affordances are not some kind of internal property, but always and only arise through the interaction of an embodied agent with its environment; by perceiving and understanding it directly and immediately (ibid). The concept of affordance therefore shows a clear reference to perception and action, as well as to embodiment and environment.

3.9.2.4 Social Affordances

Generally, social affordances are a subcategory of affordances that provide opportunities for social interaction (Borghi, 2018; Rietveld, 2012; Rietveld et al., 2013; Rietveld & Kiverstein, 2014). When presented with a social affordance, an agent is drawn to a certain possibility of social actions and/or interactions (ibid). In relation to the example provided in the section on social scaffolding above, infant-caretaker interactions may also be viewed as social situations that afford, i.e. provide possibilities for, specific social interactions. Social embodied actions, such as gestures or facial expressions, as well as social situations or practices can carry a social affordance (Borghi, 2018; S. Gallagher, 2008; Rietveld et al., 2013). For example, facial expressions of emotions, like a sad expression, can serve as a kind of social affordance that prepares others to react in a certain way, e.g. to approach and comfort the agent with the sad expression (Matsumoto et al., 2008; Rietveld & Kiverstein, 2014). Similarly, an extended hand may socially afford a greeting handshake, or being in the lift with others affords to moves away to an appropriate distance of one another (Rietveld, 2012; Rietveld et al., 2013).

Notably, social affordances can be distinguished from object affordances, e.g. a cup that affords grasping (Rietveld 2012, 2013). However, object and social affordance should not be seen as separate, because “object affordances make up an important part of the context of social affordances, and vice versa [...] object affordances is not independent of the social.” (Rietveld, 2012, p. 208)

Furthermore, social affordances may also be understood in terms of how the presence of others influences object affordances, i.e. what kind of responsive actions are evoked (see e.g. (Matsumoto et al., 2008; Rietveld, 2012)). Strictly separating the physical and social context may be misleading as the environment supports interaction through both physical and social affordances, and social input from other agents may help to fully benefit from the affordances offered by the environment (H. Dreyfus, 2002b, 2002a; Rietveld, 2012). Environmental affordances may provide a multitude of possible actions and interactions; the affordance of an object and the relationship between objects is characterised by the physical, as well as the social context they are embedded in. Therefore, affordances might evoke different actions depending on the social context (ibid), since some actions caused by object affordances might not be considered appropriate or meaningful in the presence of others. Some social situations may even afford us to alter or actions. For example, one’s favourite song might afford dancing and jumping around when being in a dance club, but not in the office or when using public transport. Or – returning to the cup-example – a fresh cup of tea may afford immediate grasping of the cup and slurping of the tea, but probably not when sitting next to the Queen, which would likely afford a rather specific way of grasping, stirring, and drinking the tea that is considered ‘appropriate’ in this social context.

Whether (social) embedded cognition that also involves exploiting the technological environment, e.g. situations involving interactions between humans and artificial agents, can be conceptualised by employing the framework of social affordances, such as gestures or facial expressions generated by artificial agents, will be discussed in chapter 5.3.

3.10 Extended Cognition

According to the position of extended cognition (Clark, 1997a; Clark & Chalmers, 1998a, 2010a), cognition is not limited to the brain or the body, but – in some cases – extends into the environment to include external artefacts, such as technological devices or tools.

Two concepts appear to be central to determine if and when cognitive systems are (at least temporarily) extended, i.e. include physical artefacts/devices: 1) (Causal) coupling between a cognitive agent and the artefact/device in question, and 2) the parity principle, also called functional parity (or alternative principles, see discussion of parity in chapter 3.10.3 below).

If a cognitive agent is linked with an external artefact in a two-way interaction, then this interaction creates a kind of *causal coupling* which can be considered as an (extended) cognitive system (ibid). That is, all relevant components in the system have a *causal role* and jointly organise cognition. If the external components were removed, the system's competence would decrease; similar to removing parts of the brain (ibid). As Clark and Chalmers state: “as we confront some task, a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is (so we claim) part of the cognitive process”(Clark & Chalmers, 2010a, p. 29). Furthermore, the active role of an external resource in and for the cognitive system is of relevance for Clark and Chalmers. Their understanding of cognitive extension is a form of ‘*active externalism*’: “the relevant external features are active, playing a crucial role in the here-and-now”(Clark & Chalmers, 2010a, p. 29). The external elements directly impact the organism and its behaviour and are “in the loop, not dangling at the other end of a long causal chain”(Clark & Chalmers, 2010a, p. 29).

Returning to the Tetris-example: When a person plays the computer game Tetris, he/she could either imagine how the pieces fit together (i.e. mental rotation), or he/she could actually rotate the pieces in the game to see how they fit. Are both scenarios cognitive processes? If the agent does not, or only partially, rely on mental rotation, but (also) rotates the pieces directly in the game, is the device the game is played on *part of* his/her cognitive system during the task of rotating the shapes? If someone is sitting next to the player and yells “rotate it to the left!” – is this person also part of the cognitive system? As the following chapter will discuss, criteria for the ‘inclusion’ of an environmental resource into the ‘cognitive unit’, i.e. whether it is causally coupled, are an ongoing debate.

3.10.1 Criteria for Cognitive Extension

Clark’s and Chalmers (1998a, 2010a) paradigmatic example for an extended cognitive system is the coupling of Otto and his notebook: Otto has memory issues and always carries his notebook with him, in which he takes notes of all kinds of things he deems important or interesting. Otto

routinely relies on his notebook to look up information. Inga does not have memory issues and routinely relies on her memory. Both of them want to go to the museum, but, initially, are not sure where it is located. Inga retrieves the information about the museum's location from her memory (as an internal and 'biological' information storage). Otto, on the other hand, does not know the location and cannot retrieve it from his memory. But he does know that this information must be in his notebook and looks it up – he uses his notebook as an external information storage, similar to Inga's memory. Both Inga and Otto have correctly 'remembered' the museum's location, but – during the task of information retrieval – Otto's cognitive system consists of the causal 'Otto-notebook-coupling', i.e. the notebook is part of Otto's cognitive system¹¹. If someone were to take away his notebook, Otto could not, or only poorly, perform the task of retrieving the museum's location (ibid).

However, not just any external object can be causally coupled to an individual and become a constitutive part of a cognitive system. Concerning criteria for inclusion of external, non-biological artefacts into an individual's cognitive system, Clark (2010b) has specified a 'rough-and-ready' set of criteria:

1. That the resource be reliably available and typically invoked. (Otto always carries the notebook and won't answer that he "doesn't know" until after he has consulted it).
2. That any information thus retrieved be more or less automatically endorsed. It should not usually be subject to critical scrutiny (unlike the opinions of other people, for example). It should be deemed about as trustworthy as something retrieved clearly from biological memory.
3. That information contained in the resource should be easily accessible as and when required. (Clark, 2010b, p. 46)

Furthermore, Clark (2010b) states that certain devices or artefacts may be causally coupled to an individual, while others don't: "[...] A book in my home library would not count. [...] Mobile access to Google would not (it would fail condition (2)). Otto's notebook would." (Clark, 2010b, p. 46).

The principle behind the question regarding when or what kind of devices can become an (extended) part of an individual's cognitive system is the *parity principle*, also referred to as principle of *functional parity*. Rooted in functionalist thinking, this principle states that there has to be a kind of "functional isomorphism between internal and external cognitive resources" (Krueger,

¹¹ It should be noted that Clark and Chalmers differentiate between extended *cognition* and extended *mind*. The example described here would be an instance of extended cognition (i.e. involving cognitive processing). Extended *mind* refers to mental states, such as beliefs, desires, or emotions. E.g. Inga *believes* that the museum is located on street X, i.e. she *believes* that she remembers the location correctly. Otto does not know the location, but *believes* that the location is in his notebook. Here, the causal coupling between Otto and his notebook are an example of an extended mind. (cf. (Clark and Chalmers 1998, 2010))

2014a, p. 537) concerning the relevant aspects for the given task. I.e. with respect to the aspects relevant to the given task, internal and external resources have to function in a similar way, fulfil the same purpose, and yield the same results as their internal ‘counterpart’. Coming back to the above example of Otto and the museum: the way Otto uses his notebook is functionally isomorph to the way Inga uses her biological/internal memory. In both cases, it is some kind of information storage or memory, both of them use it to retrieve information, and both processes yield the same result; both know where the museum is located.

3.10.2 Extended Cognition as a Constitution Thesis

It should be highlighted that extended cognition does not only suppose causal relations between coupled human and non-human entities. The extended cognition thesis is actually a *constitution thesis*. In a nutshell, cognitive extension in terms of a constitution thesis suggests that external resources or artefacts play a constitutive role during (some) cognitive tasks and/or that they (temporarily) constitute a part of the cognitive system.

Whether inferring constitutive relations from causal relations is (metaphysically) legit or empirically fruitful is still a matter of debate (see e.g. (Adams & Aizawa, 2010; Clark, 2010a, 2010a; Menary, 2010; Ross & Ladyman, 2010)). Common criticism includes the issue of ‘over-extension’ of cognition or ‘cognitive bloat’, i.e. considering any external element that causally affects cognitive processes to be part of the cognitive system or confusing the coupling with an external device or process that is merely helpful to a certain cognitive process with one that is of a ‘proper’ cognitive kind to lacking a clear principle of demarcation (Adams & Aizawa, 2010). A pivotal part of this debate is the so-called *coupling-constitution fallacy*¹².

As proponents of an internalist view of cognition (i.e. cognition as cranial, brain-bound processing), Adams and Aizawa (2010) point out that causal relations do not automatically entail constitutive relationships. Furthermore, they criticise the lack of a distinction between cognitive and non-cognitive (i.e. lacking a ‘mark of the cognitive’, c.f. (Rupert, 2009)) and inferring cognitive constitution of non-cognitive objects due to them being causally coupled to a cognitive agent.

12 In discussions in the literature, some authors also refer to this issue as ‘causal-constitution fallacy’ (vs the original term ‘coupling-constitution fallacy’ as used by (Adams & Aizawa, 2010; Clark, 2010a) debated here). According to Rottschaefer (2017) – who basis his discussion on the work of Palermos (2014, 2016) – this is to indicate that they understand the *coupling* characteristic of continuously, recurrently, and mutually interacting systems as a kind of *constitutive* relation, rather than a *linear causal* relation between two distinct entities. The causal-constitution fallacy thus (also) focuses on the fallacy of confusing causal relations with coupling relations or to distinguish them in such a way (cf. (Rottschaefer, 2017; Gallagher, 2018; Ross & Ladyman, 2010)), whereas the original term ‘coupling-constitution fallacy’ indicates the fallacy of confusing coupling (as a linear causal relationship) with constitution. Furthermore, it appears that ‘coupling-constitution fallacy’ is used with regard to debates in and/or about classical/computational conceptions of cognitive extension, whereas the term ‘causal-constitution fallacy’ seems to be leaned on more in the work of authors who’s work follows an anti-internalist, dynamic, and/or enactive conceptions of (extended) cognition.

They point out that this kind of confusion is the most common mistake extended cognition theorists make: “If you are coupled to a rock in the sense of always having it readily available, use it a lot, trust it implicitly, and so forth, Clark infers that the rock constitutes a part of your memory store. Yet coupling relations are distinct from constitutive relations [...] the fact that an object or process X is coupled to a cognitive agent does not entail that X is a part of the cognitive agent’s cognitive apparatus”(Adams & Aizawa, 2010, p. 68). Concerning more concretely the use of tools or devices during cognitive activities, they (ibid) also point out that if a mathematician used a pencil to calculate the addition ‘ $2+2 = 4$ ’, then Clark and Chalmers would consider the pencil to be cognitive, as it is part of the cognitive system; but the pencil on its own does not know ‘ $2+2$ ’ (ibid).

Proponents of the extended approach (e.g. Clark 2010b), however, point out that the hypothesis of extended cognition (HEC) claims only that cognition is extended in *some* cases, usually in relation to (and for the time span of) *certain* tasks; the extended resource or artefact *temporarily* becomes a part of the cognitive system. Secondly, the relation between cognitive agent and cognitive resource appears to be misinterpreted. Extended theorists do not claim that a non-cognitive artefact becomes “magically cognitive in itself” (Krueger, 2014a, p. 534) when coupled with a cognitive agent. Rather, both the agent and the artefact become part of one extended cognitive system.

Furthermore, Ross and Ladyman (2010) point out that the issue of causal and constitutive relations is a matter of metaphysical considerations, which should not define how to model cognition. Modellers should draw boundaries of models of cognition (e.g. extended beyond or limited to the brain or body) in a way that maximises efficiency of capturing a phenomenon.

Nonetheless, the difference between causality, coupling, and constitution when investigating cognitive couplings should not be disregarded entirely. From a methodological perspective, one clear advantage of extended cognition – as it is also the case in other situated approaches – is that processes that have originally been considered to be ‘private’ or ‘internal’ become visible, i.e. observable:

seeing cognition as extended one is not merely making a terminological decision; it makes a significant difference to the methodology of scientific investigation. In effect, explanatory methods that might once have been thought appropriate only for the analysis of “inner” processes are now being adapted for the study of the outer, and there is promise that our understanding of cognition will become richer for it. (Clark & Chalmers, 2010a, p. 30).

When observing such phenomena, an adequate methodology has to include criteria for distinguishing parts of the cognitive system from other external artefacts that are *not* part of the cognitive system, as well consider ‘coupling-constitution-issues’ – and account for them. Clark’s

(2010a, 2010b) criteria for causal coupling and the parity principle, as well as the work of other theorists that is based on functional parity (also called ‘*first-wave*’ extended theories) may suffice. But alternatives have been proposed – often in relation to a (implicit) *rejection of functionalism*.

3.10.3 Alternatives to extension as functional parity

While early work within extended cognition relies on the principle of functional parity (i.e. ‘*first-wave*’ externalist theorists, such as the original version of Clark and Chalmers in 1998), other views of cognitive extension are not based on the parity principle (sometimes doubted ‘*second-wave*’ or ‘integrationist’ approaches), effectively sidestepping the coupling-constitution fallacy and related concerns.

Two examples of alternative principles are 1) *integration* of external devices based on the *complementarity principle* (e.g. (Krueger, 2014a; Menary, 2010)), and 2) *incorporation* of external devices or objects via *phenomenological transparency* (e.g. (Colombetti, 2017)).

Cognitive *integration* (cf. (Menary, 2010)) has a stronger focus on body-based, i.e. embodied, aspects of extended cognition and views external entities as tools or vehicles that become integrated into one cognitive system. Based on a *manipulation thesis*, rather than causal coupling, Menary (2010) points out that bodily manipulations of external resources allow

the integration of the bodily “internal” and “external” aspects of cognition into a whole. This is to think of a cognitive process as *hybrid* [emphasis added], straddling both brain and bodily manipulation of environmental vehicles. [...] There is, of course, a continuous looping causal interaction between neural processes and bodily manipulations of external vehicles” (Menary, 2010a, p. 228).

Through bodily manipulation, cognitive neural processes and external, environmental resources become integrated into one cognitive system (ibid). But the focus is not on the causal coupling of these resources, but on bodily manipulations and how they become integrated with neural processes to form a ‘hybrid’ cognitive system. Conceptually, what determines parts of the cognitive system in integrationist approaches is

not functional parity but rather *complementarity* [original emphasis]— that is, the way internal and external resources, each exhibiting different formats and functions, harmoniously integrate and in so doing collectively generate new cognitive capacities [...] it is precisely because Otto’s notebook is *not* [original emphasis] subject to bio-effects like recency, chunking, and decay that it can be such a reliable memory aid and shore up his deficiencies.” (Krueger, 2014a, p. 538).

Yet another alternative to parity can be found in the work of Colombetti (2017, 2018a). Rather than conceptualising extension from the functionalist perspective present in first-wave approaches to cognition, Colombetti discusses cognitive extension from an *enactive* point of view. Specifically, Colombetti (ibid) discusses cognitive extension as *incorporation*. Based on the work of Thompson and Stapleton (2009), Colombetti argues that an external object or artefact is incorporated into a cognitive system if it becomes “*phenomenological transparent*” (Colombetti, 2018a, p. 583). Phenomenological transparency means that the artefact is no longer experienced as such, but rather, an organism’s world is experienced with or through it. A paradigmatic example of such incorporation is a blind person’s cane.

Additionally, ‘*third-wave*’ approaches to extended cognition have been proposed. Third-wave externalists (e.g.(Cash, 2013)) have a broader understanding of cognitive extension that goes beyond the typical examples of notebooks or technological devices. They typically draw on enactive (or closely related) approaches to cognition and are particularly concerned with the issue of social interaction. However, they generally argue that cognition does not only involve social interaction, but also complex cognitive-cultural collectives, social practices, and institutions that allow for socially extended cognition (Cash, 2013; S. Gallagher, 2013; Thompson & Stapleton, 2009). One such an approaches is the ‘socially and culturally distributed cognition’ by Cash (2013). In such accounts of social extensions, individual cognition takes place within a mutually co-constructed larger context of social and technological systems, and cultural practices and institutions, e.g. the legal system (Cash, 2013; De Jaegher, 2013; S. Gallagher, 2013). This approach does not only take into account direct social interaction, but addresses cognitive extension as interaction with cognitive-cultural practices and institutions as a complex two-way interaction between individual and practices.

Such third-wave approaches of cognitive extension, will, however, not be discussed further within the context of this master’s thesis. Concerning social aspects of cognitive extension, including social interaction, third-wave approaches initially appear to be of special interest. However, as the focus of this thesis lies on social interaction in terms of direct and dyadic interaction, the following section on socially extended cognition will instead first discuss socially extended cognition in general, and then focus on ‘emergent extended cognition’ (Varga, 2016), which seem to be particularly suited for capturing dynamically unfolding dyadic social interactions.

3.10.4 (Social) Action and Interaction in Extended Cognition: Socially Extended Cognition

Concerning socially extended cognition and/or socially extended mind, Clark and Chalmers state:

“What about socially extended cognition? Could my mental states be partly constituted by the states of other thinkers? We see no reason why not, in principle. [...] What is central is a high degree of trust, reliance, and accessibility. [...] the waiter at my favorite restaurant might act as a repository of my beliefs about my favorite meals (this might even be construed as a case of extended desire). In other cases, one’s beliefs might be embodied in one’s secretary, one’s accountant, or one’s collaborator.” (Clark & Chalmers, 2010a, p. 38).

Thus, some cases of (dyadic) social interaction may qualify as *socially extended cognition*. To provide an example, one can return to the case of Otto and his notebook: Otto has Alzheimer, but in this specific imaginary case, he is also visually impaired and increasingly has problems with reading his notes. His long-term partner, whom he trusts and has been with for several years, has neither memory issues nor problems with his/her eyesight. Otto routinely lives with his partner and readily relies on him/her during their every-day life, for example when running errands, following a recipe to cook dinner – or when going to the museum. Hence, Otto’s partner may very well fulfil Clark’s (2010b) criteria. Otto trusts his partner, who is available and reliable, automatically endorsed, and accessible. In terms of parity, Otto’s partner may fulfil the same purpose and function as Otto’s own memory or visual perception, was it not impaired. Otto’s partner could remember where the museum is, as well as look it up in Otto’s notes to tell him the location of the museum – would this be a form of ‘*socio-cognitive*’ coupling with regard to functional parity?

Note that, however, if the relation between Otto, his notebook, and Otto’s partner was instead viewed in terms of complementarity, rather than parity, the same scenario may be viewed as a case of ‘*socio-cognitive*’ *integration*.

In yet another imaginary example scenario, Otto’s partner may be hearing impaired, and Otto has become blind. Otto and his partner routinely rely on each other – e.g. when having conversation with others, or introducing someone, or when going grocery shopping – to function as the other’s hearing or eyesight to accomplish cognitive task, respectively. Assuming that another person could be experienced similarly phenomenologically transparent as e.g. a hearing aid, a cane, or even a guide dog, one might even go as far as considering it a potential case of social extension as a form of ‘*socio-cognitive*’ *incorporation*. Fuchs and De Jaegher (2009) do in fact describe a social kind of incorporation between individuals, i.e. *mutual incorporation*. Mutual incorporation is a conception of social extension via incorporation that explicitly addresses social instances of interaction and cognition. Fuchs and De Jaegher (2009) define it as a “reciprocal interaction of two

agents in which each lived body reaches out to embody the other” (Fuchs & de Jaegher, 2009, p. 474). To clarify this kind of incorporation they provide the example of an experienced and skilled tennis player. The player does not only incorporate the ball and its trajectory, but also the opponents movements and position in space – as does the other player when returning the ball – thereby fluently switching from incorporating the body of the other to one’s own embodiment and back again (cf. (Fuchs & de Jaegher, 2009)). This conception of mutual incorporation does, however, draw on the *enactive* conception of coupling (see chapter 3.11) to qualify such an extension. They furthermore refer to the phenomenological conception of incorporation in a similar sense as Colombetti (2018a) does when referring to Merleau-Ponty (1963, 2013) and the paradigmatic example of a blind individual and his/her cane. Fuchs and the De Jaegher (2009) classify the cane-example as *uni-directional* incorporation. *Mutual* incorporation, however, occurs “Through the mutual coupling of their lived bodies—mediated through eye contact, facial expressions, voice, touch, gesture and other kinds of intentional action—they enter into a dyadic bodily state [...] and, in a certain way, incorporate the perceived body of the other”(Fuchs & de Jaegher, 2009, p. 472).

Although Clark and Chalmers (1998a, 2010a) explicitly mention the possibility of a *distinct* theory of socially extended cognition that extends not only in the physical but also the social environment, common examples of extension “involve individual props like notebooks or cellular phones, somewhat neglecting that human problem-solving often relies on social interaction” (Varga, 2016, p. 2471). Furthermore, all examples provided above involve rather simple cases of socio-cognitive extension in the sense that they revolve around a rather clear cognitive task or activity that is to be achieved, and both parties possess prior knowledge about each other’s knowledge and/or their cognitive capacities, as well as rather specific knowledge about the devices involved. But in many ‘everyday’ cases, social interaction is spontaneous and unfolds dynamically. In many instances, social interaction during cognitive activities occurs in situation in which one does not know about the other’s cognitive capacities, knowledge, intentions, motivations, goals, etc. Can such instances too be viewed as cases of socially extended cognition?

3.10.4.1 Emergent Extended Cognition

The cognitive system ‘Otto-and-the-notebook’ allows Otto to use cognitive capacities he would not have without the notebook, i.e. he could not know the location of the museum, was he not coupled with his notebook, with which he (temporarily) constitutes an extended cognitive system

(a ‘body-environment-unit’). Implicit to this example is a weak notion of *emergence*: The (emergent) properties of the ‘Otto-notebook-system’ are linear and “simply aggregate decomposition” (Varga, 2016, p. 2484). Essentially, Otto’s action of going to the museum is a direct and predictable result from Otto’s belief that the location is in the notebook, thus looking up the location in the notebook (ibid). Even if Otto relied on the memory of his long-term partner (as in the example provided above), rather than his notebook, the ‘Otto-and-his-partner-system’ in this scenario would still be a rather simple ‘aggregate decomposition’ that could be predicted from a ‘pre-emergent’ state (cf. (Varga, 2016)). The belief concerning the location of the museum is still reducible to the belief already existent in the system. According to Varga (2016), such instances may still be explained by the Hypothesis of Extended Cognition (HEC) as proposed by Clark and Chalmers (1998a, 2010a). But it does not suffice to capture many ‘standard’ or ‘real-world’ instances of dyadic social interaction, which are more complex. Socially extended cognition may involve dyadic social interactions in which cognition is not about achieving a pre-defined task by manipulating external resources (be it a device or another agent), but rather emerges as an *unintended* result of social interaction (Varga, 2016). Such dyadic social interactions are a distinctly different extended phenomenon than the one described by Clark and Chalmers (1998a, 2010a). There is a need to “take into account the asymmetry in the dyadic system, and thus the gradual difference between a prime cognizer and a ‘mere’ cognition supporter. [...] we need to bear in mind that our dyadic system arises spontaneously, without involving common goals and knowledge about domains of expertise” (Varga, 2016, p. 2479). Dyadic cognitive systems may involve non-linear elements of *irreducibility* and *uncontrollability* (ibid). Thus, some cases of socially extended cognition may be better captured by applying a stronger notion of emergence than the original HEC: ‘*emergent extended cognition*’ (Varga, 2016).

Focussing on emotion regulation, Varga argues that the dyadic interaction between infant and caretaker is a case of socially extended cognition since “cognitive emotion regulation in the infant is achieved within the framework of the interaction, by the inclusion of extra-somatic environmental resources provided by the caretaker” (Varga, 2016, p. 2472). Applying Clark’s (2010b) criteria of availability and reliability, automatic endorsement, and accessibility to the ‘infant-caretaker-coupling’ is problematic (see (Varga, 2016) for a detailed discussion). Generally, the possibility for (social) interaction a caretaker provides for the infant can only be compared to the information provided by Otto’s notebook to a limited extent. In the infant-caretaker-scenario “there is only the interaction itself, which in certain synchronic constellations gives rise to emotional regulation as a *previously unintended* [emphasis added] by-product of interaction” (Varga, 2016,

p. 2479). Furthermore, emotion regulation in the ‘infant-caretaker-system’ involves cognitive processes that are beyond the control of either part of this dyadic system and cannot be reduced to distinct parts of the system (ibid). It is a case of socially extended cognition, in which some cognitive properties are irreducible emergent properties of the *coupled* ‘infant-caretaker-system’, thereby rendering it an example of *emergent* extended cognition.

Varga (2016) proposes two notions of emergence, resulting in two *Hypotheses of Emergent Extended Cognition (HEEC)* to address such cases of socially extended cognition. These hypotheses should be viewed as complementary to the hypothesis of extended cognition postulated by Clark and Chalmers (i.e. as in (1998a, 2010a)). Both hypotheses of emergent extended cognition are meant to capture cognitive processes in dyadic social systems in terms of *complex* systems (Varga, 2016).

‘Emergence₁’ and the HEEC₁ are intended to capture *irreducible* systemic properties, i.e. when the systemic properties of a (cognitive) system are not reducible to properties of the system’s parts. Irreducibly emergent properties of a system arise autonomously as an *uncontrollable* effect of (social) interaction, “which cannot be completely explained by recourse to the intentions of the parts” (Varga, 2016, p. 2458). Furthermore, emergence₁ involves ‘*diachronic novelty*’. That is, novel emergent properties. Some of these emergent properties may only be trivially novel in the sense that they are novel in comparison to the system’s parts, but others may be *non-trivially novel*, meaning that they *diachronically* appear for the first time in the system’s history (ibid).

HEEC₁ depicts cognitive emotion regulation in the infant-caretaker-system as emergent₁; it is uncontrollable, irreducible, and diachronically novel. The novel system property cannot be reduced to the system’s parts – the ‘parts’ being the infant and the caretaker. As the infant and caretaker form two-way interactions, the system’s emergent property is more than “the simple linear ‘sum’ of the infant’s and the caretaker’s states that provides the emotion regulation ability” (Varga, 2016, p. 2488) at lower levels of complexity. The emotion regulation ability of the infant-caretaker-system is irreducible and emergent.

‘Emergence₂’ and the HEEC₂ additionally capture top-down influences in extended cognitive systems. An extended emergent₁ property is also an emergent₂ property, if the higher-level systemic properties may also have a *top-down effect* on lower-level properties in the system. ‘The whole’, i.e. the emergent systemic property, directly influences the behaviour of the system’s parts (ibid).

HEEC₂ focusses on the *unfolding diachronic process* in dyadic interaction. It depicts emotion regulation in the infant-caretaker system as it did in HEEC₁, with the addition that the dyad of infant

and caretaker as the system's 'whole' has top-down effects on the properties and processes of the systems parts., i.e. the individual emotion regulation (I.e.). Considering the infant-caretaker-system as a socially extended cognitive system in the sense of HEEC₂, allows to capture the infants emotion regulation in complex unfolding dyadic interaction. The ongoing process that regulates the infant's *real-time* emotions occurs simultaneously with the unfolding diachronic process that develops its emotion regulation ability: "the dyad exhibits top-down effects on the infant and establishes an ongoing change, which, together with the whole development of the infant contributing to this process, eventually leads to the development of individual emotion regulation abilities" (Varga, 2016, p. 2489).

Thus, HEEC provides a framework for analysing systemic processes in dynamically unfolding interaction, which become observable because the dyad (infant and caretaker) are considered to be a socially extended cognitive system. By considering the role of emergence and its properties, HEEC accounts for the complexity of spontaneous and dynamically unfolding social interaction in dyadic cognitive systems. Similar to other 4E-approaches, Vargas' proposition of HEEC has important methodological implications for the analysis of cognition, especially with regard to social interaction. As in social embodiment, the unit of analysis is no longer limited to the individual and its brain: "Instead of analysing causation ultimately in terms of microphysical causes, the idea is to apply a more coarse-grained perspective that also includes the interaction between systemic and lower-level processes [...] organism and environment at times create unified cognitive systems that should be regarded as proper units of analysis" (Varga, 2016, p. 2487).

3.11 Distributed Cognition

Distributed cognition is closely associated with the work on 'cognition in the wild' by the anthropologist Edwin Hutchins (1995a). In a nutshell, distributed cognition is based on the idea that cognitive processes are distributed over complex networks of interactions between a) different actors and b) technological devices or resources, as well as c) both actors and devices/resources in the environment (ibid).

Although distributed cognition does not stand for one of the 'Es' in '4E cognition', due to the focus on interaction with both other human agents and technological devices, as well as Hutchins' consideration of cognitive science in conjunction with anthropological and sociological perspect-

ives, distributed cognition is prone to be discussed in the context of a thesis that means to discuss the concept of social interaction within cognitive science, and especially social interaction with ITS. Thus, within this thesis, distributed cognition will be counted as one of the situated approaches subsumed under the umbrella term ‘situated cognition’.

According to Hutchins (2001a)

Distributed cognition is a framework for thinking about cognition which seeks to understand how the cognitive properties of aggregates emerge from the interactions of component parts. It can be applied to cognitive systems at many levels of complexity, from areas of an individual brain to communities of interacting persons. [...] The fundamental question here is how the cognitive processes we normally associate with an individual mind can be implemented in a group of individuals? (Hutchins, 2001a, para. 1).

The ‘component parts’ Hutchins writes about are multiple human actors and the material resources (e.g. artefacts or technological tools/devices) they use during a specific cognitive activity or task. In this sense, distributed cognition is sometimes construed as a special kind of cognition that occurs when individuals interact with one another and with artefacts. However, Hutchins (2001a, 2006), stresses that this understanding of distributed cognition is only partially correct, because distributed cognition does *not* describe a specific kind of cognition or cognitive processes. It is a *specific kind of thinking about cognition* that allows to investigate the interactions between the individual(s) and its environment. At its core, it is concerned with understanding the relationships of ‘the social’ and ‘the material’ to cognitive processes that take place inside an individual human actor, as it does not seem possible to examine ‘the cognitive’ by looking at “what is inside our heads alone” (Hutchins, 2001a, p. 2071). Thus, within cognitive science the term ‘distributed cognition’ does not mean the study of any particular cognitive phenomena or kinds of cognition. Rather, it is a theoretical approach to or perspective on cognition in general.

A final note on how the terminology is used in this master’s thesis: the special case of cognitive processes that occur exclusively through the collective effort of and interactions between social actors *without* the aid of additional technological devices (e.g. phenomena such as ‘collective memory’ or ‘group cognition’) may be called ‘socially distributed cognition’ (e.g. in (Dillenbourg & Self, 1992; Perry, 1999, 2003a)). Since distributed cognition is explicitly meant to include ‘the social’ in terms of interactions with others, the case of distributed cognition that exclusively involves interaction between social agents will not explicitly be referred to as ‘socially distributed cognition’ within the context of this master’s thesis. Distributed cognition ‘automatically’ includes instances of socially distributed cognition.

3.11.1 (Social) Action and Interaction in Distributed Cognition

Hutchins' distributed cognition is concerned with cognitive processes that not only involve the interaction between multiple actors and/or material resources, but, in fact, are only possible *because* of these interactions. Simply put: "distribution means interaction" (Hutchins, 2006, p. 376), and social interactions are a *sine qua non* of distributed cognition.

Similar to other approaches within cognitive science (especially classical or computational approaches), distributed cognition means to understand how cognitive systems are organised. It takes cognitive processes to be those processes that are involved in decision making, memory, reasoning, learning, and the like (Hutchins, 1995a, 2001a; S. Walter, 2014a). Furthermore, it depicts cognitive processes in terms of transformation of representations based on a computational conception of cognition. However, Hutchins' understanding of computation is more along the lines of Wilson's (1994) 'wide computation' (cf. (S. Walter, 2014a)). The computational processes of interest consist not only of the transformation of internal or mental representations, but also of a kind of 'hybrid' representations, which are distributed over different materials and actors (*ibid*). Thus, the computational processes are implemented by the interactions of a collective composed of human actors and technological resources.

What *distinguishes* distributed cognition from classical computational approaches to cognition is that it does *not* view cognition as a purely internal process: "The distributed cognition perspective aspires to rebuild cognitive science from the outside in, beginning with the social and material organization of cognitive activity" (Hutchins, 2001a, p. 2071). While classical cognitive science paradigms view cognitive events as symbol manipulation inside individual agents, distributed cognition encompasses a broader understanding of cognitive events that are not limited by an individual's skull or skin (*ibid*). The boundaries of the unit of analysis for cognition is extended beyond the individual. The socially interacting individuals and materials in a distributed system are observable cognitive events on the 'outside'.

In line with Hutchins' social-anthropological background and his 'outside-in' perspective, distributed cognitive phenomena are most often investigated with ethnographic methods, but might also be investigated with simulation models (Hutchins, 2001a). The framework of distributed cognition is particularly suited for the (ethnographic) study of three kinds of distribution:

1. Interaction between social others (e.g. the distribution of cognitive labour among persons mediated by human interaction),

2. interactions with the material environment (e.g. the interaction of a person with a cognitive artefact), and
3. interactions of the present and past (i.e. temporal aspects of interaction, as social environment develop over time and past interactions impact present interactions). All of these kinds of interaction take place simultaneously in real-world activities that can be observed (Hutchins, 2006)

The most prominent examples of applied studies employing the distributed cognition framework are Hutchins' analysis of ship navigation (i.e. analysis of the complex interactions between multiple individuals and environmental resources/devices when navigating a ship, cf. (Hutchins, 1995a), and his analysis of the interactions in a cockpit (Hutchins, 1995b). But distributed cognition has also been applied to workplace studies (Perry, 1999) the investigation of scientific knowledge in research teams (Giere, 2002), as well as to studies in the area of human-computer interaction (cf. (Perry, 2003a)).

3.11.1.1 *Distributed versus Extended versus Embedded*

As mentioned in the beginning of Part I, distributed cognition is not one of the 'E-approaches' to cognition, but may very well be considered as a situated approach to cognition. When taking into account that distributed cognition is concerned with agents and their interactions with one another, as well as the social and technological/material environment, one might be inclined to relate distributed cognition to embedded and/or extended cognition as a kind of 'in-between' complementary paradigm. In fact, it does share some assumptions about cognition with both extended and embedded approaches to cognition, but also differs from them in central points.

While some cognitive processes may be explained by multiple approaches (albeit differing in their focus) in terms of taking on a certain theoretical perspective, other phenomena might be best explained by either the one or the other approach. Especially (socially) distributed cognition and socially extended cognition appear to be related. For example, some authors suggest that a socially extended understanding of cognition has a more 'distributed' than 'externalist' perspective, or that some socially extended phenomena may better be viewed as distributed cognition (Rottschaefers, 2017; Slors, 2019; Stephan, 2018). Some authors state that that extension involving physical or material components is best viewed as extended cognition, while extensions to other cognitive agents is distributed cognition (cf. (Rottschaefers, 2017)). However, as Hutchins' (1995a, 2001a, 2006) points out, distributed cognition is not necessarily concerned with a specific

kind of cognition, but is an alternative way of approaching and investigating about cognition.

The following section will briefly outline this via the previously established ‘Otto-plus-notebook’ and ‘Otto-plus-partner’ (see chapter 3.9) example.

It should be noted that the present chapter does not intent to entirely demarcate or ‘de-tangle’ the different approaches in a clear-cut fashion, but rather to highlight their relations, similarities and differences, as well as the problems that arise when attempting to compare them in such a way, particularly when social interaction is involved.

Distributed cognition is concerned with cognitive processes that not only involve the interaction between multiple actors and/or material resources, but, in fact, are only possible because of these interactions. As a theoretical perspective on cognition, it ask “how the cognitive processes we normally associate with an individual mind can be implemented in a group of individuals?”(E. Hutchins, 2001, p. 2068). Distributed cognition is typically applied to situations that involve the interaction between (multiple) actors and/or the (technological) resources in their environment. Furthermore, reviewing the most prominent examples of distributed cognition seems to suggests that distributed cognition is particularly suited to explain cognition involving multiple individuals and/or resources in situations in which an agreed upon task or common goal is shared within the collective of interacting agents.

Embedded cognition is concerned with the ways an agent routinely exploits structures in the physical and/or social environment (Edmonds, 1999; Rottschafer, 2017; S. Walter, 2014a). Here, the focus lies on the environment and on how it is used (through the body). Operating in-between perception and action, the concepts of scaffolding (Clark & Chalmers, 1998a, 2010a; Lindblom & Ziemke, 2005; Rottschafer, 2017) and affordances (Baggs, 2021; Gibson, 1986b; Rietveld, 2012; Rietveld & Kiverstein, 2014) can be considered as example strategies of environmental dependency in the sense of exploitation and/or exploitation. Scaffolding refers to a specific kind of environmental exploitation in which structures in the environment are used as a ‘scaffold’ to reduce cognitive load (Clark, 1997a; S. Walter, 2014a). Affordances are opportunities for action provided by the environment (Gibson, 1986b). Both are sometimes described by means of an (individual) agent exploiting its environment via interaction with objects, technological devices, or other agents. In comparison, *distributed cognition* is mainly concerned with the interaction between multiple agents and/or material devices, which all happen to co-inhabit the same (socio-cultural)

environment (Hutchins, 1995a, 2006). The focus in *distributed cognition* lies on these interactions and how they affect the individual agents, not on the (individual) agent and how it affects or uses the environment.

Embedded cognition is strongly related to the concept of *embodiment*, while distributed cognition is not. *Distributed cognition* (in its original form) is not concerned with embodiment or the role of the body in cognitive processes. Despite the emphasis on interactions between agents and their social environment, the framework of distributed cognition “offers little on the embodied nature of human cognition, and is currently peculiarly ‘disembodied’.” (Lindblom, 2015a, p. 5). Even if the basic idea of distributed cognition might initially appear to be compatible with interactional and social account of embodied cognition, distributed cognition focusses on interaction, not the body and its role in ‘executing’ such interactions.

The example of Otto-and-his-notebook and Otto’s memory may be viewed as – qua the components’ interaction – either distributed over Otto and his notebook, or as embedded (some authors even refer to this as an example of being embodied, embedded, and extended, cf. (Lindblom, 2015b, 2015a; Rottschaefer, 2017)). One might consider the use of the notebook as Otto ‘offloading’ the information of the museum onto the environment. The notebook might afford Otto to look into it to retrieve the information, or flipping through, and writing into his notebook might act as a kind of memory scaffold. Otto uses his body to interact with the notebook, which he exploits as a resource of his environment. In this sense, Otto’s cognitive activity is *embodied* and *embedded* (and/or extended).

The relation of and distinction between *distributed* and *extended* perspectives on cognitive processes, appears to be more complex.

Theorist of distributed and extended cognition understand cognition as (partially) ‘spread’ across external resources/non-human artefacts, i.e. objects within the agent’s environment, as well as other cognitive agents (in the case of socially extended cognition described in chapter 3.9.4). *Extended views* tend to conceptualise cognition as located within the individual that extends into the environment, e.g. artefacts or technological devices that become integrated into the cognitive system. It is the individual that extends its cognition into the environment, e.g. a specific tool, to temporarily form an extended cognitive system. In views of *distributed cognition*, there does not seem to be such a ‘centre’ or a ‘locus’ of cognition, as the interaction between the cognitive components are of interest (Scholz, 2013; S. Walter, 2014a) Put differently, *extended cognition* (as

well as *embedded cognition*) is an ‘inside-out’ perspective, and *distributed cognition* is an ‘outside-in’ (Hutchins, 2000) perspective on cognition.

In the Otto-plus-notebook example, it appears to be Otto who is the locus of cognition. The notebook extends his cognition to form one extended cognitive system; it becomes a part of Otto’s *extended memory system*. Even in the more complex ‘social’ case of the Otto-and-his-parter example, one might still argue that – during the cognitive task of retrieving the information regarding the location of the museum – Otto’s cognition is extended by the memory of his partner. From a distributed point of view, however, the Otto-and-his-parter system are a cognitive collective. The properties of the *distributed system* differ from the properties of the systems components, i.e. Otto, his notebook, and/or his parter. The distributed system – as a kind of ‘multi-person-agent’ – would possess the information due to their interactions, as the memory and information about the museum is distributed over the system.

However, such distinctions appear to become less clear, as the complexity of the (social) situations increases, especially when technological devices such as computers, smart devices, or interactive or ubiquitous technologies are involved, particularly concerning the example of extended memory (Hamilton & Yao, 2018; S. Walter, 2014a; R. Wilson, 2005). What if Otto and his parter were not to use his notebook, but a smartphone with a smart/voice assistant like Siri, Alexa, or Cortana, which they trust, readily invoke, and regularly and reliably rely on to retrieve the correct information? What if Otto asked his parter about the museum, who then consults a smart assistant, which provides the desired location of the museum to both of them? Or: Otto asks his parter, who then gives him the information, but tells him to also consult Siri (the smart assistant) to be sure, which he then does, upon which Siri confirms the museum’s location: Would such instances be best viewed as a case of extended or distributed cognition? This seems to depend on the perspective of the chosen approach. If the memory in terms of the information about the location of the museum is attributed to the ‘Otto-plus-parter-plus-smartphone assistant’ system with Otto in its centre, then one might consider this to be a case of (socially) extended cognition. A focus on the Otto-parter-smartphone assistant interactions as distributed process of information retrieval might be a case of distributed cognition, as the task is attributed to the collective distributed system of Otto, his partner, and the smart virtual assistant.

S. Walter (2014a) suggests to resolve the tensions between distributed and extended cognition by aligning distributed with extended cognitive processes. Ascribing the cognitive process to one

actor as the centre or locus of cognition would be the main point in such an attempt. Other agents and/or resources would be viewed as ‘extended’ (ibid). Such a perspective acknowledges that cognitive processes are primarily provided by individuals, but also that they are part of an organised social collective, which is (sometimes) supported by technology. At the same time, the distributed system as a collective is not in itself the bearer of the cognitive performance ((S. Walter, 2014a), cf. social manifestation thesis (R. Wilson, 2005)). However, in some situations determining the centre or locus of cognition might not be as straightforward or meaningful, for example in situations where multiple agents (including the use of technologies) of equal ‘rank’, i.e. all having a comparable level of expertise, knowledge, etc., participate in a decision-making process, Such situations may occur when a research group discusses the interpretation of their research results, or when members of a political party vote on an issue.

Particularly when cognitive activities or tasks involve (social) interactions between human agents and/or non-human entities, distinguishing the different foci of the distinct approaches and the perspective they employ appears to be a complex issue, which is still a matter of ongoing debates..

How the case of social interaction between a human learner and an ITS could be viewed from an embedded, extended, or distributed point of view will be discussed in the chapters 5.3-5.5.

3.12 Enactivism

In comparison to the other ‘E-approaches’, enactivism has an even stronger focus on body-world interactions than the previously described situated approaches to cognition, and is strongly characterised by its non/anti-representationalist stance towards cognition (Kyselo, 2013a; S. Walter, 2014a, 2014b; D. Ward et al., 2017)

From a historical perspective, enactivism is not rooted within one of the traditional cognitive science core disciplines; it draws on the phenomenology of perception from Maurice Merleau-Ponty, the philosophical biology and phenomenology of life from Hans Jonas, and the theory of autopoietic systems of Francisco Varela and Humberto Maturana (S. Walter, 2014a; D. Ward et al., 2017). In a strict sense, enactivism can be traced back to the work of Varela, Thompson, and Rosch (1991) ‘The Embodied Mind: Cognitive Science and Human Experience’ (TEM), which was “An ambitious synthesis of ideas from phenomenology, cognitive science, evolutionary biology, Buddhist philosophy and psychology, it attempted to articulate a new research programme:

an enactive cognitive science, that would bridge the gap between the empirical study of the mind and the disciplined reflection on our lived experience that characterises phenomenological and Buddhist practices.” (D. Ward et al., 2017, p. 365).

Work on TEM began in the mid 1980ies as a parallel movement to cognitivism and representational approaches to cognition. Unlike cognitivism and computational approaches, which based their model of cognition on the computer metaphor, enactivism has the living organism at the heart of its model of cognition (S. Walter, 2014a; D. Ward et al., 2017; D. Ward & Stapleton, 2012).

While computational approaches to cognition understand cognition in terms of information processing and the transformation and/or manipulation of internal representations, TEM instead emphasised the explanatory importance of interaction between an organism (in terms of a ‘lived’ body) and its environment (ibid). In terms of ancestor theories or models of cognition and its antecedents, TEM works in the same tradition as several related theories, models, or empirical work that developed at roughly the same time – all of which involve aspects of co-dependent organism-environment relations to some degree. Among others, this includes dynamical systems theory, Gibson’s ecological psychology and his theory of affordances and direct perception, as well as the work of several phenomenologists, like Merleau-Ponty (ibid). Another related area was the work on situated robotics and Rodney Brook’s (1991b) reactive, behaviour-based mobile robot ‘Creatures’— which did not have a central processing unit but were instead composed of several special-purpose subsystems – a ‘subsumption architecture’ (cf. (Brooks, 1991a, 1991b) – which guided simple sensorimotor behaviours to use the world as ‘its own model’ (Brooks, 1991b). In fact, in TEM Varela et al. (Varela et al., 1991) even claim that Brook’s Creatures are an illustrative example of enactive cognition (D. Ward et al., 2017). All of these developments and trends within cognitive science gained momentum roughly around the time TEM was developed. In this sense, cognitive science gradually moved towards an ‘active’ understanding of cognition that was embodied and embedded, and TEM meant to further these trends (D. Ward et al., 2017; D. Ward & Stapleton, 2012).

Enactivism emphasises that cognitive structures emerge from processes of self-organisation which result from *interactions* between an organism and its environment. In their initial definition of enactivism Varela and colleagues (1991) state that: “the enactive approach consists of two points: (1) perception consists in perceptually guided action and (2) cognitive structures emerge from the recurrent sensorimotor patterns that allow action to be perceptually guided.”(Varela et al., 1991, p. 173).

A paradigmatic simple, yet clear example later provided by Varela (1997) is the case of the single cell bacterium as an autopoietic system:

The bacterial cell is the simplest of living systems because it possesses the capacity to produce, through a network of chemical processes, all the chemical components which lead to the constitution of a distinct, bounded unit. [...] An autopoietic system—the minimal living organization—is one that continuously produces the components that specify it, while at the same time realizing it (the system) as a concrete unity in space and time, which makes the network of production of components possible (1997, p. 75).

As a biological organism, the bacterium maintains itself by moving towards and absorbing nutrients (sucrose), and moving away from other noxious substances in its environment (De Jesus, 2016; Varela, 1997; D. Ward et al., 2017). The bacterium is separated from its environment by a semi-permeable boundary that is created and maintained by a network of processes (moving towards and absorbing sucrose, avoiding other substances) that repeatedly cross the organism-environment boundary while the bacterium self-regulates itself with respect to its condition (De Jesus, 2018; D. Ward et al., 2017). The interactions between the bacterium and its environment are a necessary requirement for its self-preservation. Because sucrose is not only an existing structure or condition in its environment, but a relevant nutrient for the bacterium in order to maintain itself, it becomes a ‘meaningful’ condition to the bacterium (ibid). The existence of sucrose is an environmental property; its status as valuable nutrient for the bacterium, however, exists only in relation to the bacterium. Both the bacterium and the ‘meaningful’ structures in its environment emerge from a set of self-organised and dynamic processes (ibid). Thus, these *meaningful* structures are ‘enacted’ as part of the dynamical process that both creates and sustains the bacterium and the environment it is responding to (De Jesus, 2018; D. Ward et al., 2017). This kind of co-production of the bacterium and its environment through dynamic interaction is what TEM conceives of as ‘emergent cognitive structures’ and ‘perceptually guided action’.

This example illustrates some of the *core terms and concepts* of the enactivist approaches to cognition: autonomy, autopoietic systems (a term which, notably, is not mentioned in TEM), emergence, structural coupling (or ‘sensorimotor coupling’), adaptivity/adaptive systems, experience, sense-making, and embodiment.

Autopoietic systems are autonomous systems that are:

organized (defined as unity) as a network of processes of production (synthesis and destruction) of components such that these components:
(i) continuously regenerate and realize the network that produces them, and
(ii) constitute the system as a distinguishable unity in the domain in which they exist.
(Varela, 1997, p. 75).

In relation to *autonomy*, autopoietic systems exhibit self-sustaining dynamics that create both a distinction between an organism and its environment, as well as a distinct domain of interactions relevant for the organisms survival (D. Ward et al., 2017). The bacterium from the above example is an *autopoietic system* because it is an operationally closed biological unity that *emerges* through the interactions with its environment. It is separated from its environment by a semi-permeable boundary that is created and maintained by repeatedly moving towards and absorbing sucrose (and avoiding other non-nutrient substances) in order to self-regulate itself with respect to its condition (De Jesus, 2016; D. Ward et al., 2017). The bacterium thereby “generates and specifies its own organization through its operation as a system of production of its own components” (

(Maturana & Varela, 1980) as cited in (D. Ward et al., 2017, p. 368)). This autonomy does not entail complete independence from the environment since the bacterium’s self-maintaining operations require resources in the environment, i.e. sucrose. In this sense, the bacterium as an autopoietic system is *structurally coupled* to its environment through its recurring self-preserving sensorimotor interactions: the environmental interaction of swimming towards and absorbing sucrose yields a change within the bacterium that allows it to maintain itself (Maturana & Varela, 1980; Varela, 1997; D. Ward et al., 2017). Thus, the bacterium develops a specific set of interactions with the environment that allow the bacterium to sustain itself, and it employs these interactions *adaptively* as it regulates itself in accordance to its condition. Crucially, the ‘transformation’ from sucrose as a neutral property of the environment to a *meaningful* structure (i.e. a nutrient) for the bacterium exists only in the relation to the bacterium and its adaptive sucrose-seeking behaviour (De Jesus, 2016; Varela, 1997, p. 19; D. Ward et al., 2017). Through these interactional dynamics a ‘*point of view*’ – and, thus, subjective *experience* – of the bacterium *emerges*, in which sucrose becomes a valuable nutrition, while other noxious substances become something non-nutritious that has to be avoided (De Jesus, 2018; Varela, 1997; D. Ward et al., 2017). In the enactive approach, this process is called *sense-making* (De Jesus, 2018; Kyselo, 2013a; S. Walter, 2014a). Through sense-making, the meaningful structures are brought forth – they are ‘*enacted*’ – as part of the dynamic interactional process between the bacterium and its environment (ibid). Sense-making and the interactional dynamics between an organism and its environment depend on and are constituted by the particular way in which the organism is *embodied* (ibid). An organism’s embodiment determines how structures or properties in the environment can support its self-preservation because its body allows for particular sensorimotor interactions with its environment (ibid). The bacterium seeks out sucrose in a specific way of ‘swimming’ towards sucrose-rich

areas because this is how it can sense, reach, and absorb sucrose due to its physical composition; the way in which it is embodied.

3.12.1 Three Variants of Enactivism

Since its publication, different interpretations of ‘The Embodied Mind’ (Varela et al., 1991) and its core conceptual ideas have been presented as ‘enactivist’, resulting in different distinct and partially overlapping variants of enactivism, that focus on different aspects of cognition or ‘mindedness’ (D. Ward et al., 2017). In the following chapters, these variants will briefly be described, namely 1) autopoietic enactivism, 2) sensori-motor enactivism, and 3) radical enactivism (Hutto & Myin, 2012; D. Ward et al., 2017).

3.12.1.1 Autopoietic Enactivism

Whereas TEM views cognitive systems more generally as autonomous, i.e. as autonomous systems, autopoietic variants (e.g. (E. Di Paolo, 2005; Thompson, 2010; Varela, 1997; A. Weber & Varela, 2002)) of enactivism view cognitive systems as autopoietic. Autopoietic forms of enactivism (AE) strongly rely on the principle continuity of life and mind (*‘life/mind continuity thesis’*) as they attempt to ground cognition in the biodynamics of *living* systems (Kyselo, 2013a; D. Ward et al., 2017). In AE, *living systems are cognitive systems* and cognition is a kind of sense-making that is enacted through the interactions of adaptive autonomous systems – specifically, autopoietic systems – and their environment (De Jesus, 2016; Kyselo, 2013a; D. Ward et al., 2017). The life/mind continuity thesis states that the structure of higher cognitive systems can be described as autonomously organised according to the principles of autopoiesis, whereby the autonomy ranges from the simplest cell organisms to more complex organisms. Due to the continuity of life and mind, higher cognitive principles, e.g. in humans, follow the same principles as those that fall under the simplest of life forms, such as the single cell bacterium (ibid). They are an enrichment of life itself like it is already found in simple single-cell organisms: “organisational structures and principles distinctive of mind are simply enriched versions of the structures and principles grounding life itself.” (D. Ward et al., 2017, p. 370)

In their analysis of the different contemporary forms of enactivism, Ward et al. (D. Ward et al., 2017) point out three main aspects when comparing AE to the original TEM: 1) teleological directedness, 2) adaptivity and 3) how the life/mind continuity thesis is conceived of.

Returning to the bacterium example, the simple sensorimotor interactions between the bacterium

and its environment (e.g. moving towards sucrose, moving away from noxious substances) are understood in terms of ‘*teleological directedness*’ to the environment (De Jesus, 2016; D. Ward et al., 2017). This sense-making process (i.e. the enactment of meaningful cognitive structures) is teleologically directed in the sense that sucrose has ‘significance’ or ‘value’ to the bacterium because – in relation to the bacterium – it becomes something ‘good’ (a nutrient) for the bacterium with regard to the dynamics of its embodiment (De Jesus, 2018; Varela, 1997; D. Ward et al., 2017).

As already hinted at in the bacterium example above, the conception of biodynamics in TEM emphasised the dynamics of *autonomy*, while AE further speaks of autonomous systems as *autopoietic systems* and proposes that cognition requires *adaptivity* (De Jesus, 2016; E. Di Paolo, 2005; D. Ward et al., 2017). That is, a capacity to modify the relationship to the environment in a way that aids an organism in self-sustaining itself as an autopoietic unity. In the bacterium example, the bacterium senses and seeks out sucrose in its environment in accordance to its condition; this capacity and the interactional dynamics through which it reaches and absorbs sucrose makes it an adaptive autopoietic unity (ibid).

The way AE theories conceive of the life/mind thesis differs from the original understanding in TEM and the premises that cognition is grounded in biodynamics. Whereas TEM implies that the dynamics of autonomous organisation entail teleological relationships between the organism and its environment, AE theorists (e.g. (A. Weber & Varela, 2002) argue that the *continuity of life and mind is foremost phenomenologically* evident. With regard to ‘lived’ experiences and how they ground cognition in living systems, AE tends to supplement TEM’s phenomenological accounts of perception after Merleau-Ponty ((2012), originally published in 1945) with the phenomenology of life after Hans Jonas (1966) and the ‘Jonasian inference’ that ‘life can only be known by life’ (see (Jonas, 1966) for a more detailed discussion on this issue).

3.12.1.2 Sensorimotor Enactivism

While AE focusses on cognition in living systems more generally, sensorimotor enactivism (SME) restricts its analysis of cognition to the *sensorimotor dynamics relating perception to action* (D. Ward et al., 2017). Thus, SME theorists focus on the structure, content, and character of *perceptual experience as relationships between sensations and embodied activities* (ibid). Various sub-strands of SME exist, and while they differ in their scope, e.g. a focus on perceptual experiences versus analysing perceptual experiences as a means to understanding consciousness more generally, the different forms understand perception as ‘*perceptually guided action*’ (Varela et al.,

1991) similar to the original TEM. Furthermore, SME is not particularly concerned with the Jonasian interpretation of the life/mind continuity thesis and related (phenomenological) claims. Ward et al. (D. Ward et al., 2017) even state that SME “largely ignores or downplays the other theoretical principles associated with enactivism, such as the co-production of organism and environment, emphasizes on biodynamics and a commitment to life/mind continuity” (D. Ward et al., 2017, p. 371).

Instead, SME account for the content and character of perceptual experiences through ‘*sensorimotor contingencies*’ or ‘*sensorimotor regularities*’ (Noë, 2004, 2005; O’Regan & Noë, 2001). That is, “patterns of dependence obtaining between perception and exploratory activity” (D. Ward et al., 2017, p. 371) in the sense of systematic interrelations between (potential) perceptions, sensations, actions, and movements (ibid). Perceptual experiences, i.e. seeing, in a cognitive system, means learning to master such sensorimotor contingencies and understand how (potential) perceptions and movements are related. In this sense, perception *is* a kind of action: a “particular way of exploring the environment” (O’Regan & Noë, 2001, p. 939) A prominent sensorimotor view on perception is the sensorimotor theory by Noë (2004), who claims that perception is ‘all the way in’ since it does not break down into locally represented components, but depends in its entirety on the capacity to exercise *sensorimotor skills* to access the environment¹³.

An example illustrating this, that is sometimes referred to in the literature, is the perceptual experience of a tomato (see e.g. (D. Ward et al., 2017; D. Ward & Stapleton, 2012)). In a nutshell, the example concerned the question of how an agent can perceive, or see, a whole tomato when being presented only with its facing side? Humans, as agents, may only see one side of a tomato, yet they understand the tomato in its ‘entirety’. For SME, this is possible because an agent implicitly understands the ways in which potential exploratory movements would bring the other sides of the tomato into view, i.e. the agent has the capacity to understand how details of the environment can be accessed through exploratory activities (Noë, 2004; O’Regan & Noë, 2001; D. Ward et al., 2017; D. Ward & Stapleton, 2012). An agent perceives the whole tomato because the perception of its complete shape results from the understanding of how an agent is related to various perceptual aspects of the tomato’s shape, which can be accessed through exploratory movements (ibid). Through this ‘skilful mastery’ (ibid) of the relational patterns of sensorimotor contingen-

13 Note that in this sense SME also endorses Brook’s (1991) anti-representational account applied in his ‘Creatures’ – which, through their specialised sub-systems (a bottom-up ‘subsumption architecture’, see e.g. (Brooks, 1989, 2014) – that coupled sensory information to actions, posses capacities for accessing detailed information in the environment to use it ‘as its own best model’, rather than information processing in terms of constructing and manipulating detailed internal models (cf. (D. Ward et al., 2017)). In this regard, SME endorses cognitive systems in a ‘non-biological’ sense.

cies, and thus understanding how potential perceptions and movements are related, a human perceptually experiences ‘the whole’ tomato. What an agent can ‘do’, in terms of movements, thus affects what can be perceived. In this sense, SME also emphasises that cognition (at least in terms of perceptual experience) depends on the specific bodily configuration of a cognitive system and its active movement (Kyselo, 2013a; O’Regan & Noë, 2001; D. Ward & Stapleton, 2012).

3.12.1.3 *Radical Enactivism*

Radical Enactivism, also called “Radical Enactive Cognition” or “REC” in short, (Hutto & Myin, 2012) is not a competing variant of enactivism, but an attempt to improve and ‘unify’ enactive approaches to cognition. REC does include the variants outlined above, but is also complemented by work in dynamical systems theory, embodied robotics, artificial-life research, or Gibson’s psychology (Hutto & Myin, 2012; Kyselo, 2013a; D. Ward et al., 2017). Nonetheless, Hutto and Myin (2012) state that “REC is the only truly tenable kind of enactivism. The only good enactivism is a properly radical enactivism.” (Hutto & Myin, 2012, p. 5).

Similar to the previously described variants, REC rejects cognitivism and representational approaches to cognition, and emphasises interactional dynamics in its analysis of cognition. However, REC does not share the commitment to a strong life/mind continuity thesis of AE. Furthermore, while agreeing with AE’s contention that biodynamics can underpin cognition, REC heavily qualifies the way in which AE grounds intentional relations to the environment in biodynamics through ‘sense-making’ that enacts or ‘brings forth’ some kind of meaning or ‘significance’ in the organism-environment relationship. Particularly, it *objects to the attribution of content* in this regard, e.g. characterising the bacterium in the example above as swimming towards or ‘valuing’ sucrose. That is because REC abandons the idea that “all mentality involves or implies content” (Hutto & Myin, 2012, p. 1). For REC, organismic activity in terms of *engaging* the environment in specific ways (i.e. through sensorimotor interactions)

suffices for the most basic kinds of cognition. Such activity does not depend upon individuals retrieving informational content from the world—content that is then processed and manipulated—in order to attribute properties to the world. In short, not all mentality requires individuals to construct representations of their worlds.¹⁴ (Hutto & Myin, 2012, pp. 3–4).

¹⁴ Importantly, Hutto and Myin (2012) differentiate between (basic) *cognition* and *mentality*, with basic cognition being a narrower conception of mentality. They differentiate the two terms in the following way: “‘basic cognition’ means something narrower than ‘basic mentality.’ It denotes mental activity that exhibits intentional directedness, but it doesn’t necessarily imply phenomenality. ‘Basic mentality’ denotes mentality that may exhibit both intentional directedness and phenomenality.” (Hutto & Myin, 2012, p. x) and further state that such basic cognition and mentality can be ‘without content’: “It may be believed that wherever there is intentionally directed cognition there must be content [...] More modestly, it may be believed that wherever there is at least perceptual experience there must be content [...] We advance the fortunes of the counter-thesis: that there can be intentionally directed cognition and, even, perceptual experience without content.” (Hutto & Myin, 2012, p. x).

In comparison to the other forms of enactivism, REC does not only take an anti-representational stance towards cognition, but also explicitly puts forward the concept of '*contentless*' cognition. REC thus draws on the same – or similar – theoretical principles and assumptions as the other variants of enactivism. It does, however, also go beyond them in specifying 'basic' cognition that is grounded in sensorimotor interactions between an organism and the environment it engages with, in contrast to the socio-culturally scaffolded capacities that are characteristic for mature human cognition (D. Ward et al., 2017; D. Ward & Stapleton, 2012). According to Ward et al. (2017), it thus analyses:

cognition in terms of an interplay between the biological and sensorimotor dynamics emphasised by the above forms of enactivism [AE and SME] (suitably purged of representationalist undertones) and social dynamics that bootstrap basic minds into the realm of contentful thought and experience. (Ward et al., 2017, p. 373).

Rather than advancing (or defending) the ideas of other forms of enactivism, REC's understanding of cognition essentially builds on the *strong* version of 1) the embodiment thesis and 2) the developmental-explanatory thesis (Hutto & Myin, 2012).

For REC, *embodiment* is defined as the contextually embedded *sensorimotor interactions* – involving not only the brain but also the non-neural/extra-cranial body – between an organism and the environment. Furthermore, these dynamic interactions *constitute* cognition in the form of activity that involves individuals that *engage* with aspects of their environment over time (ibid). The *strong version of the embodiment thesis* dictates that basic cognition has to be understood in terms of the concrete patterns of environmental activity of situated organisms, and that these patterns of organism-environment interactions constitute (basic) cognition. Mentality *consists of* the wide-ranging “extensive ways in which organisms interact with their environments” (Hutto & Myin, 2012, p. 7). This also holds for phenomenal experience, which is understood as a specifiable kind of activity (e.g. neural activity) that too has to be understood through the ways in which individuals engage with certain aspects of their environment via sensorimotor interactions (ibid).

The *developmental-explanatory thesis* is concerned with temporal dimensions of interactions and an organism's *history of previous interactions*. According to Hutto and Myin it “holds that mentality-constituting interactions are grounded in, shaped by, and explained by nothing more, or other, than the history of an organism's previous interactions” (Hutto & Myin, 2012, p. 8). This entails that it is not a set of internal representations, but *repeated engagements* with an environ-

ment that shape and explain current interaction tendencies, skills, know-how, and embodied tendencies of an organism (ibid).

3.12.2 Enactivism and other E-approaches

The relation of enactivism to the other E-approaches to cognition is disputed, especially with regard to embodied and, foremost, extended cognition. Many theorists claim that enactive cognition incorporates embedded and embodied accounts of cognition to some degree or, rather, in a very specific sense of being embodied (Chemero, 2013; S. Gallagher, 2017; Hutto & Myin, 2012). The – still ongoing – debate revolving around the concept of cognitive extension, however, appears to be less straightforward. For example, some authors argue that enactivist accounts based on a deep continuity of life and mind cannot be aligned with the principle of cognitive extension (see e.g. (Villalobos & Silverman, 2018; S. Walter, 2014a; Wheeler, 2011)), while others argue for enactivism as subsuming extended cognition (e.g. (D. Ward & Stapleton, 2012)). The spectrum of the different positions ranges from theorists stating that *some* of those approaches are compatible (D. Ward & Stapleton, 2012), to authors understanding enactivism as subsuming, or even being a condition for, *all* the other E-approaches (e.g. (Hutto & Myin, 2012)), to others stating that, due to its theoretical basis and assumption, enactivism is in *no* way compatible with the other situated approaches (S. Walter, 2014a).

3.12.2.1 Enactivism as Embedded, Embodied, and (sometimes) Extended

An enactive understanding of cognition holds that perception and cognition depend on the cognising agent's/organism's interactions with the environment. Some enactivist accounts (e.g. (S. Gallagher, 2017; D. Ward et al., 2017; D. Ward & Stapleton, 2012)) propose an understanding of enactive (*human*) cognition that is at the same time embedded, embodied, (potentially) extended, and, furthermore, affective (sometimes doubted '4EA' approach or 'ecumenical enactivism', see e.g. (D. Ward et al., 2017; D. Ward & Stapleton, 2012)). Those authors understand enactivism as a kind of unification or continuation of the other E-approaches. Enactive cognition is even viewed as a condition for this: "if cognition is enactive then it is also embodied, embedded, affective, and potentially extended" (D. Ward & Stapleton, 2012, p. 89).

In a similar vein, Gallagher (S. Gallagher, 2017), who's understanding of enactivism appears to be more strongly influenced by phenomenological and (neo-)pragmatist philosophy (see pragmatic turn chapter 2), formulates seven key assumptions or central principals, that too imply an understanding of enactivism as embedded, embodied, extended, and affective:

1. Cognition is not only a brain-event or ‘bound’ to the brain, but emerges from interactive brain-body-environment processes.
2. ‘The world’ in terms of meaning and intentionality, is not something that is ‘pre-given’, but is enacted by cognition, action, and social interactions.
3. An understanding of cognition or cognitive processes has to take into account the role of the ‘*worldly contexts*’ in which they acquire ‘meaning’; they do not involve a kind of representational mapping or internal model of the world.
4. Enactivism is strongly linked to dynamical systems theory and thus emphasises the relevance of dynamic brain-body-environment couplings.
5. Cognitive systems are *socially situated*, intersubjective, and *extended*, whereby the conception of extension is an enactive one, and *not* equivalent to the functionalist conceptions of cognitive extension from (first wave) extended cognition theorists (e.g. (Clark, 2010b; Clark & Chalmers, 1998a, 2010a) .
6. Higher-order cognitive functions (e.g. imagining, reflective thinking) are closely coupled with *situated* and *embodied actions*.
7. These complex cognitive functions are grounded not solely in *sensorimotor coordination*, but also *affective and autonomic aspects of the body*.
(S. Gallagher, 2017).

Whereas the other situated approaches reviewed in the present thesis share an *anti-internalist* (i.e. the view that cognition/cognitive processes take place entirely inside an organism's boundaries of the body or the head/brain) with enactivism, not all are based on a non- or anti-representational understanding of cognition. The variants of embedded, embodied and extend cognition presented in this thesis seem to include sub-variants that differ in their theoretical position regarding cognition. I.e. they either retain computational and/or information-processing assumptions or view cognition in dynamical systems and/or anti-representational terms, whereas enactivism clearly and solely takes on a non- or anti-representationalist position. Necessarily, if enactivism is to incorporate an understanding of cognition as embedded, embodied, and (sometimes) extended, it can only address them within an understanding that too is based on a dynamic conception.

This issue is also explicitly addressed by authors who argue for an understanding of enactivism as embedded, embodied, and extended, like Ward and Stapleton (D. Ward & Stapleton, 2012). Not-

ably, Gallagher's¹⁵ (2017) notion of enactivism is explicitly a reaction to weak notions of embodiment that view the body not as a constitutive and active part in cogitation, but as a kind of 'vehicle' or shell that provides 'body-formatted representations' to the brain (S. Gallagher, 2017, 2019)). In a similar fashion to REC (S. Gallagher, 2017, 2019), Gallagher points out that enactivism is based on a strong notion of embodiment. Thus, an enactivist understanding is not compatible with embodiment in general, but only with those variants that are rooted in a conception that adheres to a *strong notion of embodiment* (Hutto & Myin, 2012), e.g. what Kiverstein (2012) calls body-enactivism, or Chemero (2013) refers to as radical embodiment, and Gallagher (2011a) calls radical or enactive embodiment.

The different notions of embodiment are in part related to the bigger *dispute regarding the role of cognitive extension* in enactivism. Both extended cognition and enactivist theorists explicitly address the role of the body in and for cognition. For enactivists, cognition as a phenomenon is enacted via (sensorimotor) interactions in terms of dynamic interactions and agent-environment-couplings. However, despite acknowledging the role of the body, embodiment plays a different role in (first wave) externalist theories of cognition. At least in cognitive extension based on the principle of functional parity (e.g. (Clark & Chalmers, 1998a) the body is a means to extend cognition to include external devices in the sense of a non-neural 'vehicle' for cognitive processes. The cognitive system is extended, but rather than achieving this through dynamic interactions, cognition 'starts' in the brain and then includes the body and (specific aspects of) the environment (Clark & Chalmers, 2010a, 2010a; S. Gallagher, 2017).

On a conceptual level, the dispute around cognitive extension of (first wave) externalists in contrast to enactivists may pertain to the use of the concepts of 'coupling' with the environment and that external aspects are part of the cognitive system in a sense of 'constitution'¹⁶. Similar to the notion of embodiment, in enactivism, these concepts have to be rooted in dynamical systems thinking, rather than in representationalist, computationalist, or functionalist conceptions.

15 Regarding the relation of enactivism to other situated approaches, particularly embodiment, I want to point out that Gallagher (2017) seems to use 'embodied cognition' (apparently in the sense of radical embodied, similar to e.g. (Chemero, 2013; Hutto & Myin, 2012) as the central concept or a kind of umbrella concept/term for situated approaches to cognition. In the context of this thesis, 'situated cognition' is used in such a way, i.e. embodied, embedded, extended, and enactive (as well as distributed) cognition are treated as distinct approaches within situated cognition as a paradigm in cognitive science.

16 This is in part related to a metaphysical debate revolving around the concepts of 'coupling', 'constitution', and 'causality' that has already been mentioned in chapter 3.9.2. For example, Gallagher (S. Gallagher, 2017, 2018b) points out that extension in an enactive understanding of cognition requires a particular conception of coupling as well as constitution that differs distinctly from functionalist accounts. Coupling has to be understood as dynamic, not linear, and constitution cannot be (only) understood as compositional in terms of synchronous 'part-of-the-whole' relationships. For a detailed discussion of these concepts in a dynamic sense and in relation to the fallacious 'coupling-constitution'/'causal-constitution' issue regarding enactive and extended approaches see e.g. (S. Gallagher, 2018b)

At least first wave extended cognition, i.e. approaches based on the parity principle, view cognitive extension as linear, *causal couplings*. In an extended cognitive system, an external artefact that is causally coupled with an agent, can play a constitutive role in a cognitive process, i.e. external resources may (temporarily) *constitute* a part of the extended cognitive system (ibid). These external resources actively impact cognitive processes as they influence the cognitive task “here-and-now” (Clark & Chalmers, 2010a, p. 29). Environmental resources may not be only passively involved at the end of a long causal chain, since Clark and Chalmers state that they are “in the loop” (Clark & Chalmers, 2010a, p. 29) via ongoing causal couplings. However, this conception is based on a *functionalist* understanding of cognitive extension. In contrast, enactivism generally understands the interactional dynamics between organism and environment that *constitute* cognition as ‘loopy’ in a *dynamic* sense (S. Gallagher, 2018b; S. Gallagher & Zahavi, 2008; Hutto & Myin, 2012) and also speak of *structural or sensorimotor coupling* with the environment in a *dynamic* sense (ibid). Rooted in dynamic systems thinking, interaction dynamics between agent and environment are thus ‘loopy’ and recurrent in the sense of a “simultaneous *reciprocal* [emphasis added] causation” (Hutto & Myin, 2012, p. 6), and not linear. Their dynamics cannot be isolated and viewed as causally linked ‘inner’ responses and ‘outer’ contributions from the environment, like it is the case in (first wave) externalist accounts of cognition (Hutto & Myin, 2012, p. 6).

Gallagher (S. Gallagher, 2017) argues for an enactivist and situated account of extension that is situated, dynamic, and relational, and draws on the enactive conception of (social) affordances:

The situation includes not just our notebooks, computers, and other cognitive technologies, and not just the social and cultural practices and institutions that help us solve a variety of cognitive problems, it also includes *us* [original emphasis]. We are in the world in a way that is not reducible to occupying an objective position in the geography of surrounding space [...] We, as minded beings, are definitively ‘out there’, *dynamically* [emphasis added] coupled to artifacts, tools, technologies, social practices, and institutions that *extend* [emphasis added] our cognitive processes. Enactivist and extended mind conceptions are, or at least should be, of one mind in this regard. (S. Gallagher, 2017, pp. 59–60).

Alternatively, Ward and Stapleton (2012) too propose that cognition can sometimes indeed be extended in a dynamic sense, but ground such cognitive extensions in the enactive concept of *transparency* (similar to e.g. (Colombetti, 2018a)).

Hutto and Myin (2012), on the other hand, point out that cognition (in the sense of REC) is better understood as *extensive* and not as extended:

cognition is essentially *extensive* [emphasis added] and not merely, as Clark and Chalmers (1998) famously argued, extended [...] The difference between these claims is that the Extended Mind Hypothesis (EMH) doesn’t rule out the assumption that biologically basic cognition is, by default, brain-bound [...] only in exceptional cases—for example, when non-bodily add-

ons are required in order to make the achievement of certain cognitive tasks possible—do minds extend. By contrast, those who endorse REC and thus the strongest version of the Embodiment Thesis assume that minds are already, *in their basic nature* [emphasis added], extensive and wide-ranging. (Hutto & Myin, 2012, p. 7).

3.12.2.2 Enactivism as an Independent Theoretical Approach

Yet another perspectives on the relation between enactivism and the other situated approaches is put forward by Sven Walter (2014a). According to Walter (*ibid*) enactivism – as a research programme – is a distinct theoretical framework that has to be understood independently from other situated approaches. Enactivism is based on a conception of cognition as interactional dynamics involving an organism (including its brain *and* non-neural body), its sensory-motor interactions, and the environment (*ibid*). This conception of cognition is ultimately at odds with other situated approaches. As mentioned earlier, enactivism is strictly a non-representational approach and thereby at odds with all other strands or variants within situated cognition that retain computational or cognitivist aspects within their models of cognition. Walter (2014a) further claims that enactivism is not compatible with *any* of the other situated approaches; neither the ones based on computational, nor the ones rooted in dynamical conceptions of cognition (*ibid*). This is because, so Walter argues, the other situated approaches are essentially concerned with the question of ‘where’ cognition is (not) located. In enactivism, however, cognition is a relational phenomena that emerges from dynamical interactions that enact processes of sense-making, and thus cannot be pinned down to a specific place or ‘where’ of cognition. If the other situated approaches are framed as a dispute over the ‘where’ of cognition, then enactivism can serve neither as a condition for, nor as a ‘merger’, consolidation, or as as kind of enhancement of the other situated approaches (*ibid*). In this line of thinking, enactivism has to be understood as a distinct and independent theoretical position that *bridges* the sub-personal and personal dimensions of cognitive processes, rather than being concerned with the ‘where’ of cognition (*ibid*).

Walter (2014a) also points out that enactivism still has to clarify its position and theoretical value within cognitive science as a novel research programme. Due to its holistic understanding of cognition and a lack of unified theoretical tenets, enactivism struggles to provide fruitful empirical means or methods that further explanations of their claims. It also struggles to further the analysis of concrete cognitive processes that were investigated within cognitivism, such as short or long-term memory, action-coordination, planning, an the like (*ibid*). Gallagher (2017) too notes that the ‘holistic’ enactive account of cognition presents a challenge for science:

Enactivists, by focusing on not just the brain, not just the environment, not just behavior, but on the rich dynamics of brain–body–environment, offer a holistic conception of cognition. To put it succinctly, however, it is difficult to operationalize holism. [...] On the one hand, enactivism makes empirical claims, for example, about the work of sensorimotor contingencies, and in this respect it resembles a research program that can suggest new experiments and new ways of interpreting data. On the other hand, its emphasis on holism presents problems for empirical investigations [...] enactivism may be better thought as a philosophy of nature than a scientific research agenda. (S. Gallagher, 2017, pp. 21–22)

Hutto and Myin (2012) clearly see enactivism as a promising, radical, and even a revolutionary approach within cognitive science, but acknowledge that its legitimacy is still a matter of debate. Concretely, they point out that:

the legitimacy of these revolutionary approaches remains hotly contested. It is certainly true that there hasn't yet been a definitive articulation of the central and unifying assumptions of such approaches. [...] there is some reason to doubt that it will be possible to group together all of the offerings that nominally travel under the banner of Enactive and Embodied cognition by identifying their commitment to a set of well-defined theoretical tenets. (Hutto & Myin, 2012, pp. 3–4).

They furthermore state that neither enactive nor embodied cognition (in their 'radical' sense in terms of REC) can act as a label for a well-defined theory of cognition, but rather denote broad frameworks for understanding the basic nature of cognition (ibid).

In the context of this thesis, I am more in line with Walter than with the other perspectives towards enactivism and its relation to other approaches within situated cognition research. I treat all situated approaches to cognition, including enactivism, as distinct and rather broad frameworks that aid the explanation and analysis of cognition, particularly, with respect to social interaction. Some of these approaches may be (historically or conceptually) related to one another, but not all of them are mutually inclusive or contingent upon each other. Given the vast amount of different variants and theoretical interpretations of all situated approaches, I therefore do *not* treat enactivism – nor one of its variants – as a set of premises or conditions that entail that if cognition is enactive, it is also embedded, embodied, and extended. Instead, I discuss these situated approaches as *distinct* theoretical frameworks. Conversely, a phenomenon may qualify as (some variant of) embodied, embedded, or extended, but not as enactive cognition. The same holds for the 'social variants' of the different approaches outlined within this master's thesis.

3.12.3 (Social) Action and Interaction in Enactivism

An aspect that all variants of enactivism described above have in common is their clear focus on action and interaction in their analysis of (basic) cognition. Essentially, for enactivists, perception

is a form of action, and the dynamic interaction between an organism and its environment *is* cognition in the sense that those interactions are what grounds and *constitutes* cognition (Hutto & Myin, 2012; Noë, 2004; Varela et al., 1991). Within this understanding of cognition, some enactivists (e.g. (De Jaegher et al., 2010a; De Jaegher & Di Paolo, 2007; Froese et al., 2015)) do not only emphasise the relevance of interaction, but specifically also the role of social interaction. Concerning the role of interactional dynamics in cognition, some enactive theorists (ibid) even make similar assumption with regard to social cognition, i.e. that social interaction in the sense of dynamic interaction (or ‘coupling’) between individual cognisers, their environment, and, dominantly, each other, *constitutes* social cognition or at least plays a crucial enabling role (De Jaegher et al., 2010a). In this sense, perception is not only for action – in social contexts, perception is also for interaction with others.

How social interaction, or more generally the relations between individual and ‘sociality’), is addressed and conceptualised differs between the variants of enactivism.

Within a ‘unified’ 4E or 4EA understanding of enactivism (e.g. (S. Gallagher, 2017; D. Ward & Stapleton, 2012)) the relevance of the social context and environment, including the interaction with others, is often addressed explicitly. For example, Ward and Stapleton (2012) note that *human* cognition socially situated and socio-culturally scaffolded. Similarly, Gallagher (2017) points out that for enactivists, cognition is understood in a larger context of brain-body-environment interactions, and that this includes not only the physical, but also the cultural and social environment:

on the enactivist view, the explanatory unit of perception (and action, and cognition) is not the brain, or even two (or more) brains in the case of social cognition, but dynamic relations between organism and environment, or between organisms, which include brains, but also include their own structural embodied features that enable specific perception–action loops involving social and physical environments, which in turn effect statistical regularities that shape the structure and function of the nervous system. The question is, what do brains do as part of a dynamical attunement of organism to environment in the complex mix of transactions that involve moving, gesturing, and interacting with the expressive bodies of others, with their eyes and faces and hands and voices; bodies that are gendered and raced, and dressed to attract, or to work or play; bodies that incorporate artifacts, tools, and technologies, that are situated in various physical environments, and are defined by diverse social roles and institutional practices? [...] How an agent responds and what an agent perceives will depend to a great degree on the overall dynamical state of the brain, but also on environmental factors, embodied-affective and intersubjective factors, the person(s) with whom she is interacting, her worldly and intentional circumstances, the bodily skills and habits she has formed, her physical condition, as well as her history of personal experiences, and what the other person may expect in terms of normative standards stemming from communal and institutional practices. (S. Gallagher, 2017, p. 126).

In this understanding of enactivism, social interaction involves the brain, body, and physical engagements as distinct kinds of sensorimotor couplings in form of embodied dynamical interactions, such as facial expressions, gestures, and movements. Through social interactions, complex sensorimotor couplings are generated by perceiving and responding to the social embodied actions (e.g. facial expressions) of others in “rich pragmatic and social contexts” (S. Gallagher, 2017, p. 12). In this sense, social cognition is as a kind of *attunement process* to the environment that allows humans to perceive and understand others that heavily draws on the concept of *intersubjectivity* (ibid). This account of enactivism incorporates aspects of the above variants of enactivism, but also Gibson’s theory of affordances or, more generally, his ecological psychology, as well as phenomenology, and pragmatist philosophy. According to Gallagher (2017) such a pragmatic understanding of the environment as engaging not only physical or material artefacts, but also social and cultural aspects, and his general emphasis of intersubjectivity is consistent with the enactivism associated with Varela, and Thompson and their original TEM.

To provide another example, De Jesus (2016) addresses social interaction – based on AE and the life/mind continuity thesis – within the broader context of relationships between individual and sociality with regard to the paradigmatic example of the sucrose-seeking bacterium:

a nonnegotiable but often overlooked implication of adopting a life-mind continuity thesis is that, sociality needs to be taken as a fundamental ontological condition of organismic life itself. That is, a commitment to life-mind continuity requires one to acknowledge that sociality is not only to be found in the human domain but inherent in life itself, beginning already with single cell organisms. There cannot be sociality without “individual” organisms but, unlike what the single cell example strongly suggests, without sociality there cannot be any individual organisms. This general, but important insight, is by no means new and can already be found in pragmatism and more recently in biosemiotics and certain strands of anthropology. (De Jesus, 2016, p. 877)

McGann and de Jaegher (2009) too point out that the continuity of life and mind involves sociality, but address it from another perspective. With reference to ‘the social’ they seem to understand this continuity rather generally as continuity from life to cognition, from self (as a perspective) to other, and from individual to society (ibid). With regard to an enactive understanding of perceptual experience in SME’s sensorimotor contingencies (Noë, 2005; O’Regan & Noë, 2001), McGann and De Jaegher (2009) formulate an approach to social perception via the concept of *self-other contingencies*, which they also refer to in broader terms, e.g. as *social contingencies* or *intersubjective contingencies*. They state that contingencies of action are not only sensorimotor, but also social (ibid). They conceive of perception as a skill in the sense of mastering the set of contingencies or regularities within a specific domain (after (O’Regan & Noë, 2001)). They thus conceptualise social perception as a kind of skill within the social domain. This way, social per-

ception – conceived of as a social skill – can be examined in terms of self-other contingencies that structure social interaction (McGann & De Jaegher, 2009). According to the authors, this approach is also inline with TEM or the enactive approach by e.g. (Di Paolo, 2005; Thompson, 2010; Varela, 1997) in the sense that they approach social perception in the context of social engagement, crucially involving the concepts of autonomy, sense-making, embodiment, emergence, and experience. It also seems to be aligned with the ecumenical, 4EA, and intersubjective frameworks proposed by or Ward and Stapleton (2012) and Gallagher (2017), respectively.

While several enactive theories are potentially of interest within the context of this thesis, an enactive approach that explicitly analyses social interaction and its role in cognition is the framework of participatory sense-making (De Jaegher & Di Paolo, 2007). According to de Jaegher and colleagues (De Jaegher, 2018; De Jaegher et al., 2010a; De Jaegher & Di Paolo, 2007), social interaction can serve as an explanatory tool for social cognition research. Interestingly, the proponents of this approach do not only explicitly address the relevance of social interaction, but also make their motivation and the – in their opinion – necessity to scientifically investigate social interaction and its role in cognition explicit. In their opinion (see below), addressing and properly conceptualising social interaction is necessary to further social cognition research by bridging the apparent gap between social science and cognitive science approaches to social understanding:

Whilst anthropologists and other social scientists – traditionally the investigators of social interaction processes – are not often interested in relating their findings to questions about individual cognition, psychologists and cognitive scientists seem generally not aware of, or take for granted, the importance of the interaction process for social cognition, and focus instead on individual capacities. On one side, the focus has been too exclusively on the interaction process, whereas on the other, the individual has been over-exaggerated. [...] Recently proposed embodied accounts of social cognition [...] go beyond traditional cognitivist explanations and emphasise the role of the body in our understanding of another's intentions. However, a drawback of many of these approaches is that the emphasis is still too exclusively on the *individual* [original emphasis] body. In the enthusiasm for embodiment in the social realm the fact is sometimes overlooked that social understanding is crucially an *interactional* [original emphasis] process. To social scientists, this may seem a trivial insight, but in cognitive science the importance of the interaction process is only beginning to trickle through the still very individualistic net. (De Jaegher & Di Paolo, 2008, p. 34).

Essentially, they are proposing the concept of participatory sense-making as a means to overcome the limiting focus on the individual and its capacities within social cognition research (methodological individualism, see chapter 2.1). That is, a non-individualistic account of social cognition that integrates individual cognition and the interaction process (ibid). The framework of participatory

atory-sense-making is particularly suited to analyse direct face-to-face interactions, and especially dyadic interactions, and will therefore be reviewed in detail in the following chapter.

3.12.3.1 Participatory Sense-Making

Participatory sense-making (De Jaegher & Di Paolo, 2007, 2008) is an enactive account to social cognition. It appears to be in accordance with 4EA interpretations of enactivism by e.g. Gallagher (2017) or Ward and Stapleton (2012) and mainly draws on the enactive concepts of autonomy, sense-making, emergence, and experience as they have been outlined in chapter 3.11. These core concepts are used to formulate an enactive account of social cognition: Such a theory would be concerned with defining the social in terms of the embodiment of interaction, in terms of shifting and emerging levels of autonomous identity, and in terms of joint sense-making and its experience.” (De Jaegher & Di Paolo, 2007, p. 489). Rather than focussing on the enactment of meaningful structures through an *individual* agent and its environment, participatory sense-making is concerned with the sense-making processes that emerge in the interaction of (at least) two agents. Specifically, participatory sense-making is defined as:

the coordination of intentional activity in interaction, whereby individual sense-making processes are affected and new domains of social sense-making can be generated that were not available to each individual on her own. (De Jaegher & Di Paolo, 2007, p. 497).

In order to illustrate such processes of participatory sense-making, the authors provide the example of playing the pantomimes game charade:

players have to ‘act out’ a phrase without speaking, while the other members of the same team try to guess what the phrase is. Imagine that the phrase is the title of a film with two words. [...] The player starts miming the second word. With his fingers he draws a large square in the air in front of him. Someone suggests “box”. He gestures “no”. Another guesses “picture”, again no. He changes the gesture, drawing the big square again and then moving his hands together to the centre of the square and pulling his closed hands towards him while opening his arms outwards. “Cupboard.” “Wardrobe.” He gestures “no” but adds a waving ‘almost there’ movement of the hand. Then he makes the same square, puts his hands in the middle, but now pushes his hands forwards and outwards with an opening movement and then leans forwards with his upper body and moves his head and gaze from left to right. The right answer comes: “Window!” Then he points to his back with his thumb over his shoulder. “Rear Window!” [...] all interactors have to adjust their sense-making in a way that converges towards the ‘right’ gesture and the ‘right’ interpretation. The new meaning of the gesture is jointly constructed during interaction. (De Jaegher & Di Paolo, 2007, pp. 500–501).

Crucially, participatory sense-making of this kind involves embodied and expressive actions in the context of social interaction¹⁷. In order to put interaction at the centre of their enactivist perspective, they explicitly draw on the concept of *coordination* between coupled systems – as it is understood in dynamical systems theory – to define social interaction. Coordination is defined as the non-accidental correlation between the behaviours of at least two systems that are either i) sustained in coupling, or ii) have been coupled in the past, or iii) have been coupled to another, common, system (ibid). In this context, a correlation is “a coherence in the behaviour of two or more systems over and above what is expected, given what those systems are capable of doing.” (De Jaegher & Di Paolo, 2007, p. 490). Coordination of this sort often rely on rather simple coupling mechanisms and are a common phenomenon in both living/biological and physical systems (ibid). Coordinated coupling can take place under a wide range of conditions and may occur between any two systems with broadly similar properties, e.g. synchronisation of pendulum clocks, synchronised behaviour of fireflies, coordination of walking speed between child and caretaker when walking next to each other, and similar instances (De Jaegher & Di Paolo, 2007, 2008). Coordination may take on different forms of patterned behaviour, such as synchronisation, imitation, mirroring, anticipating, and the like (for a more elaborate discussion on the different kinds of coordination see (De Jaegher & Di Paolo, 2007)). Something that is present primarily in biological systems in social contexts, but rather uncommon in couplings of physical systems, is the circumstance that patterns of coordination can directly influence the interactors and their disposition to maintain or modify their encounters (ibid). Biological systems may – via embodied actions such as gestures or utterances – either facilitate or end the continuation of the coordinated coupling. Autonomous systems engaged in coupled coordination may influence the coordinated pattern of interaction that is generated, but this patterned interaction, including the history of previous coordination, can in turn influence the individuals (ibid). Instances in which this kind of bidirectional influence between interactors (in the sense the autonomous systems sustained in coordinated coupling) and their emergent interaction (i.e. the dynamic pattern of coordination) occurs, are instances of *social interaction* (ibid). This conception of coordination depicts social in-

17 Note that the definition of social interaction employed by De Jaegher and colleagues (De Jaegher et al., 2010a; De Jaegher & Di Paolo, 2007, 2008) does only include social encounters that involve *direct* interaction, thereby excluding social situations that lack elements of interaction or “mutuality” (De Jaegher et al., 2010a, p. 443) in the sense of direct co-regulated coupling, e.g. when remotely observing a social scene in which others are merely present, or the belief that someone is present. However, the authors (2007, 2010) point out that even these aspects of human social cognition, i.e. third-person or observational situations, involve social interaction in either a direct or in a derivative way because “third-person observational social capacities develop on the back of interpersonal social capacities [...] an observer of a social phenomenon, even though he *can* [original emphasis] indeed be a passive bystander as it were, is *qua* [original emphasis] sense-maker always in some way, even if minimally, engaged with the other whose behaviour he is observing” (De Jaegher & Di Paolo, 2007, p. 503). In such instances, social interaction may at least acts as an “enabler” (De Jaegher et al., 2010a, p. 443).

teraction as a process at the relational level of collective dynamics that unfolds over time, including loss and recovery of coordination (2007, 2008). Furthermore, De Jaegher and Di Paolo (2007) claim that social interaction in itself is a proper level of analysis, since analysing how coordinated couplings occur, how they are sustained, change, and, eventually, break down during social encounters reveals temporal aspects of social interaction and their consequences for participatory sense-making. That is, how meaning is jointly generated and transformed in social interactions (ibid).

Formally, participatory sense-making is a kind of coupling between two autonomous agents, which they also refer to as interactors:

Social interaction is the regulated coupling between at least two autonomous agents, where the regulation is aimed at aspects of the coupling itself so that it constitutes an emergent autonomous organization in the domain of relational dynamics, without destroying in the process the autonomy of the agents involved (though the latter's scope can be augmented or reduced). (De Jaegher & Di Paolo, 2007, p. 493).

This formal definition acts as a distinction criteria for identifying an interaction as ‘properly’ social, hence (potentially) involving participatory sense-making. For example, the unintentional transfer of body heat when waiting in a group of people at a crowded bus stop is not a social interaction because despite being coupled, the agents do not actively regulate their coupling with respect to the interaction (ibid). Bumping into someone at the same crowded bus stop when trying to walk past the bus stop also does not qualify as social interaction, unless one or both parties start to regulate their coupling in an unfolding interaction (ibid). This would be the case if they tried to move past each other, but repeatedly fail to do so due to unintentionally mirroring each others movements, i.e. moving to the same side at the same time, until one of them finally gestures to the other to move past him/her at a specific side and takes a step to the other side to make room, thereby terminating the interaction. This example also illustrates the autonomy of the interaction and the bi-directional influence of interactors and interaction: the mirroring coordination remains a property of the relational dynamic that makes the individuals face each other again and again, thereby remaining in interaction until the mirroring relation ‘breaks down’, e.g. when one party gestures where to move and steps to the other side (ibid).

Note that in contrast to the way the core concepts of enactivism in chapter 3.11, participatory sense-making particularly stresses the concept of *identity as an autonomous entity*. Since for enactivists living organisms are the paradigmatic example of cognisers, their organisational structure has certain properties that make them ‘cognitive’: “One such crucial property is the constitutive and interactive autonomy that living systems enjoy by virtue of their self-generated identity as distinct entities in constant material flux.” (De Jaegher & Di Paolo, 2008, pp. 34–35).

Autonomous systems are composed of several processes that *actively generate and sustain an identity* as a distinct, *operationally closed* entity under *precarious conditions* (ibid). It is the property of operational closure that allows the system to generate an identity. ‘Precarious’ pertains to the fact that if the system was not organised in terms of a network of its conditional processes, the system would tend to extinguish (ibid). This aspect is pivotal for their specification of social interaction, since social interaction processes themselves may take on a form of autonomy, and resulting relational dynamics are autonomous strictly in the sense of the definition provided above.

(1) there is a coupling, which is regulated so as to generate and maintain an identity in the relational domain. Thus, the resulting relational dynamics are autonomous in the strict sense of precarious operational closure given in this paper [definition as outlined just above] and define events and processes as either internal or external to the interaction.

And (2) the individuals involved are and remain autonomous as interactors. (De Jaegher & Di Paolo, 2007, p. 493).

If the interaction process itself is not (temporarily) autonomous, or if the autonomy of one interactor was destroyed (e.g. during torture¹⁸), the interaction would cease to be ‘properly’ social. While the autonomy of interaction may diminish or enhance the autonomy of one individual interactor, once the autonomy of one interactor is destroyed, the interaction process would be reduced to individual cognitive engagement. The interactor would simply become a tool or object in an *individual* sense-making process (ibid).

To provide an example: A dancing couple engaged in dyadic interaction moves each other and while one interactor is leading and the other is following, each interactor remains his/her autonomy while being affected by the movements of each other and the coordinated pattern making up the dance (De Jaegher & Di Paolo, 2007). If a dancer was to instead practice his/her movements with a dummy or a puppet, this would not qualify as social interaction in terms of a co-regulated coupling in the sense of coordination, and would be a case of individual sense-making. Crucially, this example also indicates how participatory sense-making differs from individual sense-making with regard to phenomenal experience. Consider the example of participatory sense-making of the dancing couple versus individual sense-making by the dancer who practices movements with a puppet. The puppet as a whole may be phenomenally perceived directly and become a ‘transparent’ object in itself due to the individual sense-making process. That is, via explorative sensorimotor activities and the perceptual experience of the whole puppet as either light or heavy, soft or hard, etc. However, an autonomous system (another human) as a dance partner is experienced differently, because in interaction he/she as ‘the other’ in contrast to the self (see

¹⁸ Note: In sociology, strong coercion or torture may still be considered a kind of social interaction, especially when considering power relations, see also chapter 4.2.1.1.

also self-other contingencies, chapter 3.11.3) is not fully constituted by individual sense-making activities:

We don't experience the other-in-interaction as totally obscure and inaccessible, nor as fully transparent (like an object fully constituted by my sense-making activity) [...] My actions contribute to define the other-in-interaction not so much as my squeezing contributes to the experience of softness of the sponge [Note: equivalent to the puppet in my dance example] but rather in ways that do not necessarily settle into a lawful relationship. I must alter my actions contextually in order to re-encounter the other and in the process, sometimes, be encountered myself when her sense-making unexpectedly modulates my own. This recursive effect on my actions describes the co-modulation of self-in-interaction and other-in-interaction. (De Jaegher & Di Paolo, 2007, p. 504).

Although all these examples and their (partially implicit) assumptions or claims presented here refer to human cognition, participatory sense-making is not restricted to humans. Participatory sense-making may also occur in non-human social interaction. According to De Jaegher and colleagues (De Jaegher et al., 2010a; De Jaegher & Di Paolo, 2007) many other social animals do indeed engage in patterns of coordination that builds up coherences of significance (or meaning). The discrimination between different significant or meaningful contexts through coordination, or how individual perception alters its meaning through social coordination and contingencies has also been demonstrated using simple evolutionary robotic models and minimal agents (Di Paolo et al., 2008; Di Paolo et al., 2013). Even direct cross-species interactions or the direct interaction with robots (cf. (De Jaegher et al., 2010a)) may also qualify as social interaction and/or participatory sense-making, as long as they fulfil the criteria of being *autonomous*.

The question whether the interaction between a human learner and an ITS can be a social interaction in the sense of participatory sense-making thus boils down to the question whether both interacting parties and their interaction qualify as autonomous systems. This matter, and if this framework can be (fruitfully) applied to this specific kind of interaction, will be discussed in chapter 5.6.

Part II Social Interaction with Intelligent Tutoring Systems

4 Social Interaction with Intelligent Tutoring Systems and the Role of Power and Related Affect – Proposal for a Research Design

Part II of the present master's thesis corresponds to RQ2 and describes a possible research design to address this question. Following a general introduction to the research field of ITS, including general ITS types, components/architectures, and behaviours, the research design, including the implementation with AutoTutor, as well as anticipated methods regarding data collection and analysis, are described. The employed ITS (AutoTutor) belongs to the class of conversational/dialogue-based ITS. Since the research interest of the present thesis lies on social interaction, this kind of ITS appears to be of particular interest. While all ITS types seem to achieve learning gains that are comparable to those achieved with human tutors (Graesser, Conley, et al., 2012a; Rus et al., 2013), this kind of ITS is not only interesting with regard to its educational performance, but particularly for its attempt to simulate a human tutor (Graesser, D'Mello, et al., 2012; Person et al., 2001). The conversational ITS-agent simulates specific human-like instructional roles while conversing with the learning in natural language, which could involve social interactional aspects. A strand of research regarding conversational ITS and/or pedagogical agents (e.g. (A. Baylor & Kim, 2005; Gulz et al., 2011; Haake & Gulz, 2006)) does in fact address the social dimension of learning with a pedagogical agent, including the attribution of social behaviours and norms, or human-like social characteristics like personality traits or stereotypes to pedagogical agents. Such research is concerned with the potential of an ITS-agent to become a *social* conversational agent to the learner (ibid). A more detailed description of this type of ITS, as well as the role of pedagogical agents and their instructional roles, and the concrete system AutoTutor, will be provided in chapter 4.1.3.

4.1 What are Intelligent Tutoring Systems?

Intelligent Tutoring Systems (ITS) are computer-based learning environments that incorporate computational models from the cognitive sciences, learning sciences, computational linguistics, artificial intelligence, mathematics, and other fields (Graesser, Conley, et al., 2012). These computer-based instructional systems are supposed to track the learner's psychological states, such as subject matter knowledge, skills, strategies, motivation, emotions, or other student attributes (student modelling), and provide adaptive instructions (Graesser, Conley, et al., 2012a; Ma et al., 2014). Furthermore, ITS – including, but not limited to conversational ITS (see types below) – can incorporate animated interface agents which may take on different instructional roles (A. Baylor & Kim, 2005). In systems featuring conversational agents, students/learners communicate with an ITS-agent through speech, keyboard, or other conventional input channels. In turn, the agents interact with the learner by generating speech, facial expressions, gestures, postures, and other (visually) embodied actions (Gulz et al., 2011; Gulz & Haake, 2006; Nye et al., 2014). In a meta-review on different ITS-types developed prior to 2013, Ma et al. (2014) define ITS in the following way:

An ITS is a computer system that for each student:

1. Performs tutoring functions by (a) presenting information to be learned, (b) asking questions or assigning learning tasks, (c) providing feedback or hints, (d) answering questions posed by students, or (e) offering prompts to provoke cognitive, motivational or metacognitive change.
2. By computing inferences from student responses constructs either a persistent multidimensional model of the student's psychological states (such as subject matter knowledge, learning strategies, motivations, or emotions) or locates the student's current psychological state in a multidimensional domain model.
3. Uses the student modeling functions identified in point 2 to adapt one or more of the tutoring functions identified in point 1. (Ma et al., 2014, p. 902)

4.1.1 Brief Historical Background of ITS

As a research field, ITS can be traced back to the early 1970ies (Gulz et al., 2011; Wenger, 1987). A crucial step in the development of ITS was the landmark paper by Carbonell (1970) in which the concept of human intelligence was ascribed to an artificial learning system (named SCHOLAR) that tutored students on South American geography. SCHOLAR implemented mixed-initiative dialogue (i.e. dialogue initiated by either the system or the student) and interacted with the learners via natural language to ask questions, provide feedback on learner contribu-

tions, or respond to student's questions (Ma et al., 2014). An architectural aspect that stood out at the time of publication was that the system represented domain knowledge separately from the natural language interface, and that this explicitly represented domain knowledge could serve as a basis for modelling student knowledge. This architecture – theoretically – allowed SCHOLAR to generate a large set of combinatory and diverse questions, as well as answers to a multitude of student questions (Carbonell, 1970; Ma et al., 2014). These architectural aspects were what distinguished the system from other computer-based instructional systems developed at that time. Those architectural components and their relations are also core components of many ITS designs that have been developed since (see also four component architecture below). SCHOLAR is thus sometimes regarded as the first ITS, although the term ITS was not used to describe the system in the original publication by Carbonell (Ma et al., 2014).

Another early example of ITS called 'BIP' (Barr et al., 1976) assisted students in learning programming by assigning individualised tasks to students that correspondent to their individual competencies, skill level, and learning needs, which were inferred by students' performance on a specific programming task. This was achieved by designing BIP with a domain representation that mapped goal skills to specific programming tasks, which served as a basis for the student model (ibid).

When the special issue 'Intelligent Tutoring Systems' edited by Sleeman and Brown (1979) appeared in the 'International Journal of Man-Machine Studies' and the associated edited book with the same name appeared in 1982, ITS had emerged as a new field of research on computer-based instruction systems (Graesser, Conley, et al., 2012b; Ma et al., 2014). The special issue as well as the associated book featured contributions from various disciplines, including artificial intelligence, cognitive science, and education, and most of them were concerned with student modelling in ITS (ibid). Some of the ITS featured in those works attempted to model strategies and characteristics of human tutoring or intended to simulate specific pedagogical models, or rational models of intelligent systems (ibid).

By the early 1990ies, two conferences that explicitly focused on the development and testing had been established: Intelligent Tutoring Systems, and Artificial Intelligence in Education (Graesser, Conley, et al., 2012b). In 1997, advances in ITS research had progressed to the point of ITS being used in schools (ibid). Furthermore, more work has been published on research that compares ITS performance and effectiveness to human tutoring and/or other instructional modes and systems¹⁹,

19 For details on other relevant landmark work in ITS research and its emergence as a research field, including ITS architectures, learning domains, and general ITS research from approximately 1980-2009 see (Woolf, 2009) for a meta-review on ITS research that compared ITS to human tutoring and other instructional systems up to 2013, see (Ma et al., 2014).

e.g. in the subject domains of algebra, computer literacy, physics, scientific reasoning, meta-cognitive skills, and reading comprehension across different educational levels, from elementary school to adult learning (Graesser, Conley, et al., 2012b). Different types of ITS have been developed since then, and the main types, as well as general ITS architectures and behaviours, will be briefly described in the following chapter.

4.1.2 General ITS Types, Architectures, and Behaviours

4.1.2.1 ITS Types

The most common types of ITS are:

- Cognitive tutors/model-tracing tutors
- Constraint-based tutoring systems
- Case-based reasoning tutoring systems
- Conversational/dialogue-based systems, also called expectation and misconception tailoring systems (Graesser, Conley, et al., 2012a; Ma et al., 2014)

Cognitive Tutors are systems that are inspired by cognitive research in psychology. Their architectural design is rooted in a model of human cognition and learning that focuses on memory processes. A widely known model of this type has been developed by Anderson (1990, 1993): ‘Adaptive Control of Thought’, abbreviated *ACT*, or – in its updated form – ‘Adaptive Control of Thought–Rational’; *ACT-R* in short. ACT-R is the most used symbolic computational architecture (Chipman, 2016). In short, it is a production system that differentiates between three types of memory structures: 1) a declarative memory of factual knowledge about the world, 2) procedural (long-term) memory of productions (if-then rules) that represent procedural task knowledge, and 3) a working memory via which production rules are activated, to perform a problem-solving task (Anderson, 1990, 1993; Anderson et al., 2018; Chipman, 2016; F. E. Ritter et al., 2019). In ACT-R, learning starts from factual and declarative knowledge and involves accessing the correct production rules and mastering them (ibid).

In cognitive tutors, the system’s pedagogical model (based on ACT-R) is designed to perform problem solving, i.e. to solve problems presented to a learner (Graesser, Conley, et al., 2012a; Ma et al., 2014). The knowledge or skill to be tutored is modelled by a set of production rules (if-then rules) that model correct steps in the solution of a problem, as well as ‘buggy’ or incorrect rules. That is, incorrect specifications along the problem-solving path that do not conform to the path to the correct solution (ibid). Production rules specify operations or actions, as well as the condi-

tions under which they are triggered: *if* a certain state occurs in the working memory, *then* a specific action is performed (ibid). During problem-solving, the ITS interacts with the student who selects a problem and solves it with the help of the tutor. Two processes inform this problem-solving and allow the ITS to model the student's knowledge and to produce adequate responses: *model-tracing* and *knowledge-tracing*.

The operations in the domain model are exposed in the interface where the student can select them to progress toward a problem solution. As the student selects operations a *model-tracing* [original emphasis] process maps them to a series of production rules in the domain model. If an error is detected, the student is given immediate feedback and allowed to choose a different operation. After the student's use of a production rule is identified by model tracing, a Bayesian procedure called *knowledge-tracing* [original emphasis] can be used to update an estimate of the probability that it has been correctly learned. Thus, the multidimensional student model in a knowledge-tracing cognitive tutor is constituted as probabilities assigned to production rules in the domain model. (Ma et al., 2014, p. 904).

Examples of such model-tracing systems are the CognitiveTutor (e.g. (S. Ritter et al., 2007)) and the Pittsburgh Advanced Cognitive Tutor (PACT, e.g. (Corbett et al., 2000)). Both ITS provide tutoring of well-defined mathematical problems and/or Algebra and are developed at Carnegie Mellon University (Corbett et al., 2000, p. 200; S. Ritter et al., 2007). Another model-tracing system the ITS Andes, which is used in the learning domain of physics (VanLehn et al., 2005).

Constraint-based tutoring systems are fundamentally different from cognitive (Graesser, Conley, et al., 2012a; Ma et al., 2014). Rather than modelling each step on the path to an optimal solution of a problem, these systems model the declarative structure of a 'good' solution. In this sense, knowledge is modelled as *constraints* relative to each constraint in states that could arise in the solution (ibid). Constraints consist of:

1. a relevant condition indicating when the constraint is applicable,
2. a satisfaction condition for the constraint which tests student contributions, and
3. a feedback message informing the student of an error that is activated when the student's solution state does not match the satisfaction condition of a constraint (Ma et al., 2014; Ohlsson & Mitrovic, 2007).

If a student's response violates a constraint, this is identified as an error and the ITS provides feedback. Conversely, if no violations of constraints are detected, then the ITS infers that the student is on the correct path to a good solution (Ma et al., 2014; Ohlsson & Mitrovic, 2007). This modelling is based on Ohlsson's (1994, 2016) cognitive theory of learning from performance errors, that was later revised by Ohlsson and Mitrovic (2007). An example of a constraint-based system is KERMIT (Suraweera & Mitrovic, 2004), which tutors database design and used database design principles as constraint.

Case-based reasoning systems differ from the other ITS types in one relevant regard; while model-tracing and constraint-based systems are adequate for tutoring semantically well-defined and well-formed problems (e.g. mathematics), systems informed by case-based reasoning can accommodate verbal reasoning and ill-defined domains (Graesser, Conley, et al., 2012b). This kind of system is based on different premises and paradigms stemming from learning theories rooted in psychology, the educational sciences, and cognitive science, and emphasises that learning involves case-based reasoning in order to solve problems (Aleven, 2003; Graesser, Conley, et al., 2012a; Shiri A et al., 1998). These systems are often integrated into other classroom and learning activities and are meant to assist the student in forming his/her own case-based reasoning with regard on a certain problem (Graesser, Conley, et al., 2012a). Usually, the learning process involves the formulation of a problem or question with respect to the present case, generating a hypothesis regarding the solution, retrieving and analysing existing cases as input, and generating and analysing the result (Aamodt & Plaza, 1994; Aleven, 2003; Graesser, Conley, et al., 2012a). This implies that memory, which is retrieved during learning and solving problems, is organised around cases. In this view, solving a problem generally involves the following four ‘r-steps’:

1. *retrieving* similar cases from memory,
2. *reusing* those cases in order to solve the current problem,
3. *revising* the derived solution, if necessary, and
4. *retaining* the solution as a part of a new case in the memory (ibid).

This is best explained by providing an example. A fairly well-known example ITS that is informed by theories of case-based reasoning is the law-based learning system CATO (Aleven, 2003), which tutors students on legal argumentation. CATO uses a case library on the domain to organise multi-case legal arguments in order to retrieve and analyse the cases, reason about their differences, and determine which case(s) are the most adequate to apply to the current situation (Aleven, 2003; Graesser, Conley, et al., 2012a). In their interaction with CATO, students have to predict a line of legal argumentation (a hypothesis) for a hypothetical case, retrieve relevant similar cases from CATO, and then evaluate and potentially revise their hypothesis based on the reviewed cases (ibid).

Conversational/dialogue-based systems (or *expectation and misconception tailoring systems*) are based on natural language interactions between student and learner. Systems involving natural language have been a challenging goal within ITS-research since SCHOLAR (Graesser, Conley, et al., 2012a; Ma et al., 2014). These systems usually incorporate an animated conversational agent that may take on different instructional roles (A. Baylor & Kim, 2004; A. L. Baylor &

Plant, 2005; Graesser, Conley, et al., 2012b; Y. Kim & Baylor, 2006a, 2006b; Ma et al., 2014). The agent interacts with the student by holding a conversation with him/her and is intended to simulate human tutorial interactions in social contexts. This kind of ITS heavily relies on natural language processing and thus models student knowledge by comparing a student's input (e.g. text typed in via a keyboard or generated via speech recognition) to text passages that represent specific expectations (e.g. learning goals, ideal answers to questions, or ideal solutions to problems) and misconception in the domain (D'mello & Graesser, 2012; Graesser, Conley, et al., 2012a; Ma et al., 2014; Nye et al., 2014). Comparing and matching actual contributions to expectations and misconceptions in the domain model often involves statistical methods (e.g. latent semantic analysis, see also chapter 4.3.2.1). Based on the outcome of this statistics-based comparison process, the ITS delivers an adequate response, e.g. positive feedback to a correct answer, or a supportive hint in case the student provided an incorrect answer (ibid).

Examples of such systems are AutoTutor (D'mello & Graesser, 2012; Graesser, Conley, et al., 2012a; Nye et al., 2014), Why2 (Vanlehn et al., 2002), GuruTutor (Olney et al., 2012), or DeepTutor (Rus et al., 2013), which have been used to tutor various subject domains across different academic levels and age groups.

Note that these four general types of ITS appear to be among the most common (Graesser, Conley, et al., 2012a; Ma et al., 2014), but of course other types or hybrid-types exist. To name a view, other systems are example-tracing tutors (Aleven et al., 2009), or systems based on Bayesian network modelling, which can include aspects of knowledge-tracing as well as constraint-based modelling to create complex student models (cf. (Johnson & Lester, 2016; Ma et al., 2014; Sabourin et al., 2011)). Rather than attempting to cover and briefly describe all potential kinds of ITS, the following two chapters will provide generalised ways to describe and characterise the architecture of ITS (four component architecture), as well as their behaviour (two-loop framework), across different ITS types.

4.1.2.2 Four Component Architecture

Despite the differences between the types of ITS and their functioning in terms of computational mechanisms, ITS, including the types described above, are often considered to have four major components (some of which have already been mentioned above) that comprise their architecture (Anderson et al., 2018; Graesser, Conley, et al., 2012a; Ma et al., 2014; Nkambou et al., 2010;

Pavlik Jr. et al., 2013). This *classic component view* thus represents a generalisation of architectures across ITS-types. The four major classical components are:

1. the domain model or expert knowledge,
2. the student model,
3. the tutoring model, also called tutor model or pedagogical model, and
4. the tutor-student interface model, also called interface model or simply interface (Nkambou et al., 2010; Pavlik Jr. et al., 2013).

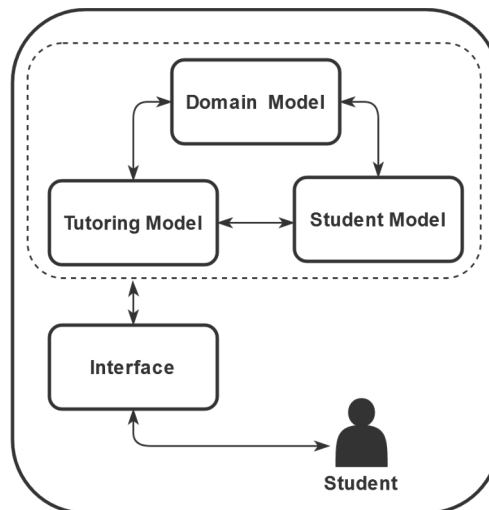


Figure 2: *Four-component architecture; this illustration is an adapted version of [Nkambou et al. 2010, p.4., fig. 1.1] that has been modified for the purposes of this master's thesis*

The *domain model* (also called *expert knowledge*) contains knowledge and/or a set of concepts, rules, or strategies of the subject matter that is being tutored (ibid). It contains expert knowledge, but can also contain misconceptions students might exhibit (Pavlik Jr. et al., 2013). This model can serve different purposes; as a source of expert knowledge it serves as a standard for evaluating possible student responses, or serve the detection of errors or misconceptions (Nkambou et al., 2010; Pavlik Jr. et al., 2013). Ideally, the domain model is a representation of all possible student states in the domain, e.g. regarding content, or psychological states, that are relevant for selecting possible pedagogies within the domain (ibid). The domain can be organised into a *curriculum*, which defines a specific structure including all knowledge aspects and their sequence, which may be modelled dynamically, along hierarchies, semantic networks, production rules, or other methods.

The *student model* is the core component of the ITS; it is the pivotal component that distinguishes ITS from other computer-based instructional systems (Ma et al., 2014). The student model comprises cognitive, affective, or other psychological states that are inferred based on data that is generated over the course of tutoring (Nkambou et al., 2010). The student model should contain information on the student's knowledge state, as well as its development and progression over time, relative to the domain model (e.g. via knowledge-tracing, see above) in order to generate didactic strategies that correspond to the student's knowledge state. The student model implements several functions. Nkambou et al. (2010) refer to two early landmark contributions on the foundations of ITS to specify these functions: the three functions proposed by Wenger (1987), and the six functions specified by Self (Self, 1990). The three main functions of the student model put forth by Wenger (1987) are that it 1) gathers explicit and implicit/inferred data from and about the student, which it 2) uses to represent the student's knowledge and learning process over the duration of tutoring, and it 3) accounts for the data by performing a diagnosis or evaluation of the student knowledge and/or state to select the optimal pedagogical strategy for presenting domain information to the student (Nkambou et al., 2010).

According to Self (1990), the student model has six functions:

1. a *corrective* function to root out bugs or misconceptions in the student's knowledge,
2. an *elaborative* function that corrects incomplete student knowledge,
3. a *strategic* function to initiate changes in the tutorial strategy (other than 1, 2),
4. a *diagnostic* function that allows the ITS to diagnose student errors,
5. a *predictive* function to determine and anticipate likely responses of the student after a specific tutorial action, and
6. an *evaluative* function in order to assess either student or ITS (cf. (Nkambou et al., 2010).

Note that these functions have been expanded or adapted since the emergence of ITS as a research field. Nonetheless, they still account for relevant student model functions.

The *tutor model* (also called *pedagogical model*) receives its input from both the domain model and the student model in order to make decisions regarding instructional strategies or actions, such as what content to deliver in which way (Ma et al., 2014; Nkambou et al., 2010). The pedagogical model selects the next tutoring actions to deliver to the student in order to move the student to a more optimal state in the knowledge domain (Nkambou et al., 2010; Pavlik Jr. et al., 2013). Such pedagogic decisions should consider and match the interaction with the student with respect to the student's state and his/her responses, e.g. by providing feedback, hints, or assertions

based on the student's contributions (ibid). In ITS with mixed-initiative dialogue (such as AutoTutor, see chapter 4.3.2.1), actions may also be initiated by the student, e.g. by asking a question to which the tutor has to generate an answer or select a different response to (Pavlik Jr. et al., 2013).

Interactions of this kind usually occur via some kind of interface; the *tutor-student interface model*. The interface model interprets student contributions through input media (e.g. speech, typing, clicking) and produces output via different media (e.g. text, pictures, speech). In addition to traditional human-computer interaction devices such as interaction via keyboard, some ITS also feature dialogue-based natural language interaction and/or speech and emotion recognition (cf. (D'Mello et al., 2011; Johnson & Lester, 2016; Nkambou et al., 2010; Nye et al., 2014; Pavlik Jr. et al., 2013)).

Note that this component view is an abstraction or generalisation that cannot account for all (current) types of architectures. Not all ITS have these four conceptual architecture component, some types might subsume different component and/or their purpose in one domain or knowledge base (cf. (Graesser, Conley, et al., 2012b; Ma et al., 2014; Nkambou et al., 2010; Pavlik Jr. et al., 2013)). For example, in the system BIP that was mentioned earlier – as well as in various other systems that have been developed since – the student model is an overlay or subset of the domain model (Barr et al., 1976). As Nkambou et al. (2010) point out, different tutoring and/or pedagogical aims, strategies, and philosophies “emphasize different components of the learning process: domain, student or tutor. The architectural design of an ITS reflects this emphasis, and this leads to a variety of architectures, none of which, individually, can support all tutoring strategies” (Nkambou et al., 2010, p. 5). Nonetheless, it appears to be a useful classification scheme for viewing and understanding the principles of ITS architectures for the purposes of the present thesis.

4.1.2.3 Two-Loop Framework

In the 2-loop framework of VanLehn (2006), ITS are described as having two loops: an outer, and an inner loop:

The outer loop executes once for each task, where a task usually consists of solving a complex, multi-step problem. The inner loop executes once for each step taken by the student in the solution of a task. The inner loop can give feedback and hints on each step. The inner loop can also assess the student's evolving competence and update a student model, which is used by the outer loop to select a next task that is appropriate for the student. (Vanlehn, 2006, p. 227)

This framework is – just as the classical four component view has been – a generalisation of different ITS types. But while the four component view is meant to describe the architectural design and the internal structure of ITS, the two-loop framework is meant to capture the behaviour of ITS. As VanLehn (2006) points out, although ITS do differ greatly in their task domains, software structures or architectural designs, as well as their user interfaces, and the like, their behaviours are actually quite similar. For instance, all example systems presented in chapter 4.1.2.1 (ITS types) can be viewed within the two-level structure of inner loop and outer loop (cf. (Graesser, Conley, et al., 2012a).

In this view, tutoring is essentially concerned with providing guidance for solving a set of tasks, whereby each task can be solved in multiple steps. The outer loops describes the ITS behaviour over tasks, while the inner loop executes over each step:

The main duty of the outer loop is to select a task for the student that will help the student learn. The inner loop is responsible for both eliciting the right steps from the student and helping the student to learn [...] The outer loop executes once per task, while the inner loop executes once per step. (Vanlehn, 2006, p. 230)

In a nutshell, tutoring systems with an *inner loop* provide *within-problem assistance* and guidance, and not only *end-of-problem feedback* once the problem-solving task is finished (Aleven et al., 2009; Graesser, Conley, et al., 2012a; Vanlehn, 2006). There are many ways to organise and implement an inner loop of an ITS. For example, the previously mentioned model-tracing systems ANDES and the Algebra CognitiveTutor provide feedback on any student step, i.e. whenever the student enters a contribution into the system during solving a problem (or task). But while Andes only provides hints on demands (e.g. when the student presses a button), CognitiveTutor provides hints when asked to, but also delivers unsolicited hints when the student takes two incorrect steps in a row (ibid).

AutoTutor's inner loop operates along pump-hint-prompt-assertion cycles when assisting a student in solving a problem (Aleven et al., 2009; Graesser, Conley, et al., 2012b; Vanlehn, 2006). It provides feedback on each student steps, generates pumps for more information and/or hints in case of (partially) incorrect student contributions, which may be followed by a prompt as well as assertions in case the student again enters a (partially) incorrect response into the system (see chapter 4.3.2.1).

The *outer loop* organises which task to select, i.e. what kind of problem or question the student should work on next (upon which the problem is worked on in the inner loop). There are various

ways to organise task selection in different ITS. VanLehn (2006) lists four common types of outer loops:

1. The task might be *selected by the student* via an interface menu (like in Andes), or
2. the ITS assigns the tasks in a *fixed sequence* (e.g. in some versions of AutoTutor). Other common outer loop organisations are
3. *mastery learning*, in which the curriculum is structured in a sequence with increased difficulty, or
4. in the form of *macro-adaption*, which is a complex pedagogy in which tasks are assigned based on the tasks knowledge components in relation to an estimated degree of mastery the student has concerning this knowledge. This method of task selection is used by the Algebra CognitiveTutor (Graesser, Conley, et al., 2012a; Vanlehn, 2006).

Some tutoring systems also model incorrect knowledge or misconceptions and can select a task based on an estimation of the students incorrect knowledge in order to clarify misconceptions (ibid). In addition to selecting a task in the outer loop, the ITS may also select a mode in which the task is to be worked on, e.g. by providing a hint on each student contribution, or to let the student work on the problem on his/her own without providing unsolicited hints (ibid).

Of course, seeing the outer and inner loop as entirely separate behavioural mechanisms is a simplified view on ITS behaviour; in fact, the two loops – or, rather, their functions or responsibilities – might intersect. For example, tracking students knowledge (student modelling) and providing adaptivity in tutor-responses across the entire tutoring session – which involves several tasks, for which steps have to be taken in order to solve them – are necessary at both the outer and the inner loop (Graesser, Conley, et al., 2012a).

4.1.3 Dialogue-Based ITS incorporating Conversational Agents

Dialogue-based ITS (also called ‘conversational ITS’, or ‘expectation and misconception tailor-ing systems’), are a special kind of ITS – and, more general, a type of educational technology and/or computer-based instructional systems – that are (often) informed by constructivist theories of learning and collaborative problem-solving activities based on human tutoring (Rus et al., 2013). This type of ITS is explicitly based on observations and investigations of human tutoring strategies and its computational modelling focussed on natural language dialogue (Pavlik Jr. et al., 2013). Conversational ITS combine task-oriented instruction, human-like pedagogical principles, and dialogue-based tutorial strategies (e.g. 5-step tutoring frame, EMT-dialogue, see chapter 4.3.2.1) with natural language to hold a conversation with the learner (Graesser, Conley,

et al., 2012b; Gulz et al., 2011; Rus et al., 2013). Natural language that simulates human tutorial dialogue is used by the ITS-agent to ask the student questions, ask him/her to elaborate on answers, to provide feedback, hints, correct misconceptions, and the like (Graesser, Conley, et al., 2012b; Gulz et al., 2011; Rus et al., 2013). Some systems also implement *mixed-initiative dialogue*, meaning that dialogue is not only initiated by the ITS, but also by the student, e.g. by asking the tutor a question (Graesser, Conley, et al., 2012b; Gulz et al., 2011; Nye et al., 2014).

4.1.3.1 Embodied Pedagogical Agents as Human-Like Tutors

Crucially, conversational ITS do not only provide computational learning environments that offer the student subject-specific content via different media (e.g. texts, pictures, videos), but build their *student-tutor interface* around an agent the learner can actively interact with. These systems often incorporate an anthropomorphic animated interface agents (also called ‘conversational agent’, ‘pedagogical agent’, or ‘embodied conversational agent’ and ‘embodied pedagogical agent’, respectively) that are *visually embodied* in the sense that they are capable of generating speech, facial expressions, gestures, or other (visually) embodied actions (Graesser et al., 2005; Graesser, Conley, et al., 2012b; Gulz et al., 2011, 2011). Embodied pedagogical agents are thus not only capable of addressing a learner through written or spoken natural language, but may also use facial expressions or gestures to cue the learner, e.g. a particular head movement and hand gesture may afford or prompt the learner to start, stop, or continue talking (Gulz et al., 2011; Haake, 2009; Veletsianos & Miller, 2008; Veletsianos & Russell, 2014). An often cited example for such a dialogue-based ITS that features a conversational agent and implements mixed-initiative dialogue is *AutoTutor* (Graesser et al., 2005; Graesser, 2016a; Graesser et al., 2014; Nye et al., 2014), which will be described in detail in chapter 4.3.2.1. Furthermore, these pedagogical agents may take on specific *instructional roles* that are based on studies of *human* tutors and their behaviours (Baker, 2016; A. Baylor & Kim, 2004; Graesser, Conley, et al., 2012a; Gulz et al., 2011; Y. Kim & Baylor, 2006b; Person et al., 2001; Veletsianos & Russell, 2014).

4.1.3.2 Instructional Roles of Pedagogical Agents

Pedagogical agents are often explicitly designed to represent human instructional roles, such as tutor (D’mello & Graesser, 2012; Graesser et al., 2005; Person et al., 2001), mentor (A. Baylor & Kim, 2005), expert (A. Baylor & Kim, 2005; Liew et al., 2013a; Olney et al., 2012), motivator (Y. Kim & Baylor, 2006b), peer (Graesser, 2016a; Y. Kim & Baylor, 2006b; Liew et al., 2013a), learning companion (Y. Kim & Baylor, 2006a) or even a social role model (A. L. Baylor & Plant,

2005). Instructional roles are operationalised along different dimensions of pedagogical agent design, including the agent's look (visual appearance including age, gender, race), their voice, as well as the instructional content they deliver and how they deliver it (e.g. involving affect), or what kind and extend of knowledge they are modelled to have (for a broad overview see e.g. (Gulz et al., 2011; Gulz & Haake, 2006; Haake, 2009; Veletsianos & Miller, 2008; Veletsianos & Russell, 2014)). For example, an expert might be modelled as a professional-looking agent in a suit that is older than the student, has an authoritative voice, speaks in a formal manner, shows little or no affect, and strictly delivers accurate information in a way that presents the agent as possessing profound and deep knowledge about the subject matter, e.g. in form of a lecture (A. L. Baylor & Plant, 2005; Liew et al., 2013a). In comparison, a motivator may be designed as a younger agent, who does not possess as much knowledge as an expert, but is eager to express, work on, and reflect on ideas, while encouraging the student to do the same by addressing him/her in a more informal and colloquial manner while speaking enthusiastically and showing a range of affective states (ibid). On the other, hand, a peer might be modelled to match the student's attributes in terms of look, knowledge state, and behaviour (A. Baylor & Kim, 2004; Baylor, Amy L, Kim, 2005; Ozogul et al., 2013). Some systems (e.g. some versions of AutoTutor, see (Graesser et al., 2016a; Graesser, 2016a) also include multiple agents that have different instructional roles, like having both an expert tutor, as well as a peer/student who interact with the human learner in order to address different perspectives on the subject matter and/or to offer different levels of guidance, support, and motivation. Furthermore, some instructional roles may merge or combine instructional roles; Kim and Baylor (2005) developed the instructional role of 'mentor', which combines characteristics of expert and motivator agents.

Crucially, pedagogical agents do not only adaptively address the content/knowledge-dimension of tutoring, but (potentially) also the social dimension of learning. Through their appearance and behaviour (including e.g. eye gaze, gestures, facial expressions, non-verbal feedback, conveying and/or eliciting emotions) and the way in which they perform along an instructional role, pedagogical agents may not only be perceived as tools or computer artefacts, but as characters or 'personas' (Gulz et al., 2011). By being designed to have a human-like persona – as well as being perceived this way by students – social contexts may be simulated better in order to promote a more natural learner-tutor interaction (Baylor, Amy L, Kim, 2005; Graesser, 2016a; Gulz et al., 2011). Thus, pedagogic agents may not only result in higher learning gains or a more engaging and motivating learner-tutor interaction, but also promote aspects of sociality in this kind of human-com-

puter interaction, thereby (potentially) rendering it a social kind of interaction (Gulz et al., 2011; Y. Kim & Baylor, 2016; Morishima et al., 2004).

4.1.3.3 Computers as Social Actors and Social Conversational Agents

According to the *Computers As Social Actors* paradigm, *CASA* in short, (Nass et al., 1994a, 1995; Nass & Moon, 2000; Reeves & Nass, 1996) humans tend to act socially towards computers and similar media. For example, it has been shown that humans can perceive computers as ‘trustworthy’, ‘selfish’, or ‘caring’ (Kulms & Kopp, 2018; Lee & Nass, 2010), or that they can feel flattered by the feedback of a computer in a similar way they would response to flattering feedback received from other humans (Fogg & Nass, 1997). This behaviour has been described as *the media equation*, which, essentially, postulates that humans somewhat mindlessly apply social rules, norms, and expectations that are key to human-human social interactions and relationships to human-computer interaction (Lee & Nass, 2010; Nass & Moon, 2000), including the interaction with anthropomorphic embodied conversational agents and pedagogical agents (Gulz et al., 2011; Schaumburg, 2001). This is of particular interest when considering conversational ITS and pedagogical agents, since they are intentionally designed to simulate human characteristics, actions, and behaviours. In fact, it has been shown that humans perceive virtual agent and/or pedagogical agents as ‘knowledgable’, ‘nice’, and ‘friendly’ (Gulz et al., 2011; Schaumburg, 2001) or ‘polite’ (Hoffmann et al., 2009; Y. Kim, 2007; Krämer et al., 2013). Furthermore, human learners tend to apply social expectations they would have of human instructors to pedagogical agents (Gulz et al., 2011; Y. Kim, 2007), and pedagogical agents – and their appearance and voice in particular – can trigger social expectations and stereotypes (Haake & Gulz, 2006; Veletsianos, 2010). Computer agents are also perceived as having more ‘intelligent’ and ‘appealing’ personalities if they match human expectations (Gulz et al., 2011; Nass et al., 1995; Ozogul et al., 2013). This indicates that embodied agents can – potentially – become social actors, rather than merely serving the purpose of a tools during learning activities (Gulz et al., 2011; Haake, 2009; Nass et al., 1995; Nijholt, 2003; Wachsmuth, 2015).

Note however, that the *CASA* framework and related research have been criticised, contested, and qualified (cf.(Formosa, 2021; Gambino et al., 2020; Goldstein et al., 2002; L. M. Pfeifer & Bickmore, 2011)). Essentially, the original *CASA* and the media equation postulates that humans tend to mindlessly apply social qualities and characteristics to computer and computational agents (Nass et al., 1994a; Reeves & Nass, 1996). This is due to process of anthropomorphism and ascription, i.e. treating computational agents as if they were human actors/agents, by drawing on

experiences and applying ‘social scripts’ from human-human interaction to the context of human-computer interaction (Gambino et al., 2020). However, the original CASA is almost three decades old and the way in which humans use and interact with computers, and technologies in general, is hardly comparable with the way human-computer interactions occurred during the time of CASA’s original publication. The interaction with technologies today is, on the one hand, more fine-grained and increasingly complex, and, on the other hand, far more accessible and integrated in day-to-day live. The usage of and interaction with computers was relatively rare at the time the original CASA was formulated, while today humans frequently and routinely interact with computers and other technologies. Some individuals even grow up using all kinds of technologies from early childhood on, e.g. via smartphones or tablets that feature animated interface agents in serious games or voice assistants/smart assistants (e.g. Alexa, Siri, Cortana), as well as more ubiquitous technologies such as smart home technologies (e.g. a smart Nest Thermostat and Google Home). As Gambino and colleagues (2020) point out in their ‘extended CASA’ theory: “[...] such changes represent a shift in the sociocultural context in which CASA is applied, that these changes are ongoing, and that they drive the need to extend the CASA paradigm” (Gambino et al., 2020, p. 75). In a nutshell, Gambino et al. (2020) argue that about 30 years ago social scripts rooted in human-human interactions were the only available social script to apply to human-computer interactions. Today, due to the acceleration of technological advances and their integration into society and daily live, humans may develop specific social scripts for interaction with computers/computational agents (‘media agents’, as referred to by (Gambino et al., 2020)). Such specialised scripts may be based on or appear to be similar to human-human social scripts, but are, in fact, distinctly different from human-human social scripts (ibid). That is not to say that this kind of human-technology interaction is not a social kind interaction, but that it is not (solely) based on human-human social scripts. Social scripts are developed and employed through and for those kind of interactions *specifically*. Neither does this imply that all interactions with computers and/or computational agents are social interactions. Furthermore, interactions with the same kind of technology may become or cease to be social over a prolonged time of repeated interactions. Empirical studies have shown that while relationships similar to human-human relations may develop over time (Bickmore & Picard, 2005), trust and feelings of partnership develop over a longer period of time (two months of interaction in case of smartphone/voice agents, cf. (D. Kim & Lim, 2019)), but social aspects may also fade after repeated and prolonged interactions, e.g. children tended to show decreased social responses in the case of social robots used for tutoring after repeatedly interacting with them (Serholt & Barendregt, 2016). Gambino et al. (2020) – who

refer to the above mentioned studies (see (Gambino et al., 2020) for a detailed discussion) – interpret those results as an indication that even if humans may apply human-human social scripts when initially interacting with computers/computational agents, these scripts change after continued usage, i.e. human-human social scripts fade and are replaced by specific human-technology social scripts (ibid).

Another relevant factor in modern human-technology in the context of the extended CASA paradigm (ibid) is the concept of social affordances (see chapter 3.8.2.4). Today's computational agents have a different affordance than the computer-based systems 30 years ago had. In a nutshell, affordances are opportunities to act in a certain way, i.e. a cup may afford grasping. Social affordances provide opportunities for or prompt specific social actions and interactions (Borghini, 2018; Rietveld et al., 2013; Rietveld & Kiverstein, 2014). In the context of computers and/or computational agents, social affordances that are perceived by humans may inform the development and refinement of technology-specific social scripts; the computational agent is perceived as providing an opportunity for social interaction or even as potential social interaction partners (Gambino et al., 2020; Gulz et al., 2011; Wachsmuth, 2015). However, the perception of artificial agents as displaying social affordances, including ascriptions or anthropomorphism, is only 'one side of the coin', so to speak. The 'other side of the coin' is that computational systems and artificial agents are also intentionally designed by developers, researchers, and engineers to display affordances that provide possibilities for social interactions. Socio-technological scripts are not only perceived and developed through continuous human-technology interaction by the human user, they are, in a way, also 'produced' and 'prompted' by the human developers. This is also the case for current ITS development.

Early ITS, such as SCHOLAR or BIP, were primarily task-oriented. Despite their capacity to provide individualised and adaptive instructions to learners, these systems were:

[...] strictly performance-based and not aimed at establishing a social relationship with the student. Personal interaction is absent, and therefore, students do not develop personal relationships with virtual tutors. The classic ITS is an impersonal, non-social and abstract pedagogical agent whose sole purpose is to tutor (Gulz et al., 2011, p. 130).

Learning with and from others is, however, an inherently social activity and social interaction is considered to be important for learning (including tutoring situations) and motivation (Gulz et al., 2011; Y. Kim, 2007; Y. Kim & Baylor, 2006a).

While early ITS did not address the social dimension of tutoring, technological advances – which occurred during the late 1980ies and early 1990ies – changed this aspect drastically (Gulz et al., 2011). Significant developments in video techniques and animation (Gulz et al., 2011; Y. Kim et al., 2007; Y. Kim & Baylor, 2006a), as well as a major shift in computational linguistics and natural language processing (cf.(Schubert, 2020)), led to new achievements in the domains of visual embodiment and animation of virtual characters, as well as dialogue-based systems.

A new generation of pedagogical agents entered the arena. This development can be ascribed to significant technological developments within two domains: the visual embodiment area and the dialogue systems area. These agents have been characterized as personal and relational artifacts. [...] Present day pedagogical agents are – in contrast to classical tutoring systems – visually embodied, and often animated. They have a face and often a body or torso and are usually humanlike in their appearance. This makes a modern pedagogical agent more tangible and less abstract than a classical ITS, which in turn increases the likelihood of learners approaching agents as social entities (Gulz et al., 2011, p. 130)

Thus, today’s dialogue-based ITS and pedagogical agents can be a “powerful means for engendering affordances for social interaction” (Gulz et al., 2011, p. 131) and may become *social* conversational agents. In this sense, pedagogical agents can expand the functionalities of ITS into the social domain: “pedagogical agents can be designed to support the social-cognitive aspect of learning in computer-based environments, playing well-defined instructional roles, following specified social conventions” (Y. Kim & Baylor, 2006a, p. 571). While early ITS were concerned with learning in the sense of ‘purely’ content-based and task-oriented instruction to maximise learning gains, modern-day ITS have the potential to address the social dimension of learning as well. Dialogue-based ITS incorporating conversational agents may be particularly suited to have learners engage in social interactions with them.

4.2 Research Design for Investigating the Role of Power and Related Affect in the Social Interaction with ITS

The following chapters are a more detailed and comprehensive description of the research design that has already been developed, partially implemented, and reported in 2016 (Roetzer, 2016), and was presented at the International ITS Conference 2016. Informed by social science theories and thinking with regard to the concept of social interaction, the present research proposal is an attempt to approach the phenomenon of social interaction with educational technologies, specifically, ITS, which are a sub-field at the intersections of computer science and AI, human-computer and human-technology interaction, psychology, educational sciences, and cognitive science.

This proposal aims to investigate the interaction between a human learner and an ITS (namely,

AutoTutor) as a *social* interaction, and, particularly, the role of power and related affect in this interaction.

Concretely, the research proposal is meant to address the following research question:

- RQ2: How do power and related affect influence and shape the social interaction with ITS regarding: i) social roles; ii) interaction and the (situational) structures it is embedded in; and iii) social relations?

Note that the main research interest is thus not about the learning content and the learning goals (chapter 4.3.2.2), or if those goals are achieved, but on the social interaction during the tutoring session. For example, whether the student provides a correct answer is not relevant on the content level or terms of learning goals, but rather how the student reacts when told that the answer is correct/incorrect, or to positive/negative feedback from the tutor, to being addressed by his/her name, to (not) being able to interrupt the tutor and/or being interrupted, and the like. Not if the student actually learnt the content is in the focus of the present thesis, but how he/she acts and reacts in terms of social interactions (which are, of course, related to the situational context, including the learning content).

This focus on social interaction allows to consider a broad range, but not all possible aspects of social interaction. It served as a starting point for designing the research setting and to make decisions regarding methods of data analysis and collection (see chapter 4.3).

Furthermore, note that RQ2 was strongly informed by a social science perspectives on the matter, particularly sociological theories on power and affect (e.g. Kemper (1991a, 2006a) and his relational reading of Collins and Goffman (Kemper, 2011)), as well as some social psychological theories (e.g. (Boiger & Mesquita, 2012; Burr, 2015; Wetherell, 2012, 2015)). The designed research setting and the reasoning behind it is, thus, based on several assumptions (see chapter 4.3.1), that are drawn from these theories.

Additionally, two more strands of research motivated the planned study.

On the one hand, the CASA paradigm and the media equation (Lee & Nass, 2010; Nass et al., 1994a, 1995; Reeves & Nass, 1996), which postulate that humans tend to (mindlessly) apply social norms, rules, conventions, and resulting expectations of human-human interaction to the interaction with computers, including the interaction with virtual and/or pedagogical agents (*ibid*). This research indicates that humans can indeed perceive computers, and particularly embodied agents, as if they were social entities in the sense of social actors, and as characters that have a

personality, rather than a computer artefact or tool.

On the other hand, there are strands of research that address one or all of the following points:

1. The social dimension of learning and an understand of learning in general as a cognitive, but also a social process based on psychological and/or sociological theories. This includes Vygotsky's (1978), understanding of social interaction and learning in the context of the zone of proximal development, Banduras's (Bandura, 1986, 2001) social cognitive theory, and/or Hutchins' (1995a, 2000, 2006) distributed cognition perspective on social interaction (like, for example, in the work of (A. Baylor & Kim, 2004; Y. Kim et al., 2007; Y. Kim & Baylor, 2006b), or Goffman's (1956) dramaturgical approach (in (Prendinger & Ishizuka, 2001))
2. Explicitly design pedagogical agents (e.g. (Baylor, Amy L, Kim, 20005; Gulz et al., 2011; Olney et al., 2012)) as well as conversational dialogue (e.g. (D'mello & Graesser, 2012; Nye et al., 2014; Rus et al., 2013; Veletsianos & Miller, 2008)) along human-like instructional roles and tutor and/or pedagogic strategies, respectively, based on the study of human-human (social) roles and interactions, and/or
3. explicitly incorporate human-like social capacities, characteristic, or qualities in the design of intelligent artificial and/or pedagogical agents,
 1. including – quite rigorously – the affective dimension of interaction, often in terms of either the recognition and tracking of human emotions and/or the generation of artificial emotions through visual embodiment and/or dialogue (see e.g. (D'mello & Graesser, 2012; Graesser et al., 2016a; Hernández et al., 2015; Y. Kim et al., 2016; Y. Kim & Baylor, 2006b; Morishima et al., 2004; Padrón-Rivera et al., 2016; Paiva et al., 2017; Qu et al., 2014),
 2. and – less often than affective aspects – social power dynamics (e.g. (Hashemian et al., 2018; Mascarenhas et al., 2013; Pereira et al., 2015, 2016)

4.2.1 Social Interaction with Intelligent Tutoring Systems: Why Investigate Power and Related Affect?

Research in the area of human-technology interaction, human-computer interaction, and intelligent systems/agent design shows that humans tend to treat artificial agents, including pedagogical agents, as if they were not 'mere' tools, but social entities (Gulz et al., 2011; Schaumburg, 2001; Strasser, 2015; Wachsmuth, 2015), and that these agents can furthermore be intentionally designed to simulate human-like roles and associated (social) characteristics, actions, and beha-

viours (ibid). Interactions which such agents may thus be considered to be of a *social* kind, i.e. be classified as social interaction. The understanding of social interaction within the present research design is strongly informed by sociological theory, in which some strands of research do explicitly acknowledge that non-humans and/or artefacts can, in fact, have agency or be treated as social actors/actants/interactants, e.g. in the context of actor-network theory (Cerulo, 2011; Müller, 2015; Sayes, 2014; Thürmel, 2015), ethnomethodology (L. Suchman, 2006; L. A. Suchman, 1987), or new materialism (Clarke, 2003, 2018; Clarke et al., 2015).

Social interaction is one of the key theoretical concepts and topics of investigation within sociology, as well as social science more generally (Bierstedt, 1950; Haugaard & Clegg, 2009; Kemper, 2006a). Power, as well as affect, are among the most puzzling and disputed concepts of the social sciences, and have been investigated on all levels of social interaction and society; the micro-level of direct social interaction, as well as the meso-level pertaining to specific social groups or communities, and the macro-levels of social structures, practices, institutions, and the like, on a societal level. Some authors even declare power as ‘the’ central concept of the social sciences (Haugaard & Clegg, 2009). Affect, too, is a central, somehow ‘re-discovered’, area of investigation in the social sciences. Wetherell (2012, 2013, 2015) and Heaney (2011, 2013, 2019) even speak of an ‘emotional’ or ‘affective turn’ within the social sciences. View concepts in the social sciences, and sociology in particular, are as manifold and complex as power and affect, spanning across a multitude of different theoretical positions and paradigms which include various – competing and partially even contradictory – theories on both subjects, viewed either in isolation from, or in relation to each other. As a consequence, Power as well as affect have been defined in different, sometimes even contradictory, ways²⁰. There exists no consensus or unified understand-

20 The following chapters describe conceptions of power and affect in terms of a broad generalisation across different social science disciplines, including conception stemming from the disciplines of sociology, political sciences, communication and media studies, as well as social psychology.

For a detailed overview of different theoretical approaches to power, including power in the interaction with human and non-human actors (e.g. Actor-Network Theory), see e.g. ‘The SAGE Handbook of Power’ edited by Haugaard and Clegg (2009).

For a general overview of conceptualisation of power and how the different approaches may be categorised and differentiated see e.g. (Göhler, 2009).

For a detailed review of emotion and affect within the sociology of emotions see e.g. (Bericat, 2016; Turner, 2009), and for a philosophical overview (in relation to psychological conceptions), see e.g. (Ben-Ze’ev, 2009).

For a more general debate of psychological theoretical positions on affect, ranging from basic emotions of a ‘natural kind’ to emotions as being socially constructed see e.g. (Frijda, 2008), and for an overview of competing discourses regarding affect within the social sciences (including social psychology) see e.g. (Wetherell, 2012) Concerning different approaches to and perspectives on emotions and affect in computer science and HCI, see e.g. (Beale & Peter, 2008; Picard, 2002). For a cross-disciplinary perspective of approaches (including computer science and AI, psychology, and philosophy) see (Cowie et al., 2011).

For a review of different philosophical and social science approaches to power and affect, and how they relate to each other, see e.g. (Heaney, 2011)

ing concerning a definition for neither the concept of power, nor of affect (Göhler, 2009; Haugaard, 2003; Haugaard & Clegg, 2009; Li, 2015; Wetherell, 2012). However, despite their differences, many theories appear to converge towards the consensus that social interaction crucially involves power as well as affect, and that power and affect occur in and influence *all* kinds of social relationships (ibid).

4.2.1.1 Defining Power

Already in 1950, Bierstedt notes that view problems in the social sciences are as perplexing as the issue of power. Conceptualisation of power can differ greatly, and some authors conceptualise power along and/or relate it to other analytical concepts, such as such as dominance, influence, prestige, or (social) order (Bierstedt, 1950), which renders any attempt of a clear-cut definition problematic. Regarding the adequate definition and investigation of power, he states:

In the entire lexicon of sociological concepts none is more troublesome than the concept of power. We may say about it in general only what St. Augustine said about time, that we all know perfectly well what it is-until someone asks us. (Bierstedt, 1950, p. 730)

Almost 60 years later, in their introduction to ‘The SAGE Handbook of Power’ Haugaard and Clegg (2009) point out that power is a core concept within the social sciences for as long as there has been an investigation of social relations, but they also note that, despite being such a central issue, power still is one of the most difficult concepts to make sense of within the social sciences. Concerning their understanding of power the editors explicitly note that the handbook is based on the premises that: “there is no single correct interpretation of power; thus, they do not seek to impose one.” (Haugaard & Clegg, 2009, p. 22)

Later, Haugaard (2014) similarly states:

There is no single uncontested definition of power, but many theorists converge on a general characterization of power either as the capacity to make another do something that that person would not otherwise do or as the exercise of such a capacity. [...] concepts, including power, have a specific meaning in relation to a theoretical context, which manifests itself in *local language games*. Thus the theoretical context is key to understanding a *particular local usage* of the concept. There are too many power debates, with particular usages, to be able to cover them all. (Haugaard, 2014, p. 2965)

However – as Haugaard (ibid) points out – there are macro-debates that structure the field of power as a whole, which offer some generalised perspectives on the theories of powers, as well as a framework for classifying and distinguishing them. On a general level, power can be understood as either

1. A Capacity in the sense of a ‘*power over*’ (e.g. in (Kemper, 2006a; M. Weber, 1978)) or – alternatively – a ‘*transitive power*’ (cf. (Göhler, 2009)) conceptualisation (Göhler, 2009;

Pitkin, 1972). That is, two parties are in *conflict* and one exercises power over the other in order to make the other do something which he/she would otherwise not (*ibid*). A classical definition is found in the work of Weber (1978), who defines power as enforcing one's will over another person in a social relation despite his/her resistance, which can be achieved using violence. As a relational phenomenon, power only exists in relation to others and is the probability of an actor to carry out his/her will despite facing resistance. In this sense, 'power over' is to prevail over others (Göhler, 2009).

2. Having the capacity, ability, or '*power to*' (as in (Arendt, 1958)) do or achieve something independently of others, i.e. to have the power to act in a certain way (Göhler, 2009). However, such a capacity may also refer to power as *consensus*, e.g. in the sense of 'empowerment', where actors and their actions and/or behaviours mutually empower each other, which may also be referred to as '*intransitive power*' (Göhler, 2009; Pitkin, 1972).

In summary, on one side of the spectrum within the range of theories on power, we find 'power over' or 'transitive' relations (*ibid*). That is, to exercise power by enforcing one's will over others even when facing resistance, which may involve violence or coercion (Göhler, 2009; Kemper, 2006a; M. Weber, 1978). In such a conceptualisation, power means that a gain for one individual leads to a loss in power for the other individual, and vice versa. On the other side of the spectrum, there is 'power to' (Arendt, 1958; Göhler, 2009; Haugaard, 2014; Pitkin, 1972). That is, having the ability or capacity to achieve something independently of others, or to – in an 'intransitive' conception – achieve something with others in the absence of violence and through communication and 'acting in concert' (cf. (Arendt, 1958)) to empower each other. Power in this sense means that a gain or rise in power can also lead to a gain for others (Göhler, 2009; Pitkin, 1972). Along this spectrum, one may generalise that 'power to' produces the social relations in which the individual and his/her field of potential actions arise due to some sort of capacity or ability. 'Power over', on the other hand, limits the array of possible actions, while at the same time opening up the field of action for the one exercising power (*ibid*). Power is a kind of configuration or medium in social relations that structures fields – or possibilities – of action (*ibid*). When power is exercised or emerges, specific options to act become available while others become unavailable (Göhler, 2009; Haugaard, 2014; Pitkin, 1972). Of course, there exists a kind of 'in between'. In more complex social relations 'power to' and 'power over' may not easily be separated and the inability or ability to act as an individual cannot be understood without contextualising it in a social relation and its (situational) structure. When working together towards a common goal, for

example, ‘power over’ and ‘power to’ may merge (Göhler, 2009). A comprehensive understanding of power should encompass both a power over/transitive and power to/intransitive relations (ibid).

Within the context of this thesis, power is broadly conceived of as a relation and social circumstance; it is a configurational property of social relations and structures. Power only exists in relation with and to other individuals. It is – in one form or another (e.g. power over or power to, transitive or intransitive) – present in all social relationships and shapes social interactions by enabling or closing off possibilities for certain behaviours, action, and interactions.

4.2.1.2 Defining Affect

Within the social science, and sociology in particular, affect has been used to refer to certain phenomena, theories, or concepts in differently, and even contradictory ways. Furthermore, affect is seldom defined clearly (Turner, 2009; Wetherell, 2012). In each discipline of the social sciences, there exists a multitude of different theoretical approaches to and conceptual investigations of affect, that are rooted in different theoretical traditions and premises: “Because of its use as a term across disciplinary and theoretical frameworks, affect remains open to mutually incompatible definitions” (Paasonen et al., 2015, p. 6).

Rather than reviewing all major theoretical approaches or traditions regarding the phenomenon of affect, the following chapter is meant to illustrate some conceptualisations of affect that appear relevant regarding direct social interaction and how affect in interaction may be approached. Considered are those conceptions of affect that appeared to be adequate to show the range of different conceptions, while focusing particularly on those of interest to the study of direct social interaction with regard to power. Furthermore, in reviewing the range of (some) conceptualisations of affect, two aspects will be discussed. On the one hand, conceptions of affect vary in their definitions with regard to related concepts such as emotions, sentiments, and feelings. On the other hand, that they seem to vary regarding the ‘where’ of affect (personal versus social, inside versus outside, and the like).

Concerning affect and related concepts, it seems problematic to clearly differentiate the terms of affect, emotion, and feeling (Hemmings, 2005; Li, 2015; Turner, 2009; Wetherell, 2012). Some scholars seem to equate one or the other concepts (or all of them), or to use them to define each other. Affect appears to be closely related to – and is sometimes even explained in terms of –

emotion. But conceptualisations of emotions too vary greatly and sometimes define emotions by means of feelings. E.g. emotion are often depicted as comprising specific components (cognitive, motivational, and feeling, cf. (Ben-Ze'ev, 2009)), with feelings being a component of emotion. Alternatively, affect can be conceived of as a pre-personal force, while feelings are personal and biographical, and emotions are social (e.g. (Shouse, 2005)).

More generally, emotion theories range from the assumption of general and universal 'basic' emotions (e.g. (Ekman, 2005; Kemper, 1991a; Tomkins, 2008)), to the view that emotions (and feelings) are neither something that is pre-given nor something universal, because they are constructed in and through social interactions, structures, discourses, or practices (e.g. (Boiger & Mesquita, 2012; Burr, 2015; Mesquita & Frijda, 1992; Wetherell, 2012)). In his attempt to define emotions with regard to the sociology of emotions and its theoretical arguments in relation to those of other fields, Turner (2009) states that:

Generally emotions are defined in terms of other terms like 'feelings' and 'affect' that are themselves defined in terms of each other. [...] much of the problem is that emotions operate at many different levels of reality—biological and neurological, behavioral, cultural, structural, and situational; and depending upon which aspects of emotions are relevant to a researcher, a somewhat different definition will emerge. For example, if the neurological aspects of emotions are emphasized, then emotions are the arousal of body systems; if culture is stressed, then the ideologies, rules, and vocabularies of emotions are seen as critical; if cognitive aspects of emotions are seen as critical, then conscious feelings along some set of dimensions will be part of the definition. (Turner, 2009, p. 341)

On the other hand, some researchers use the concept of affect as a kind of umbrella that comprises emotions, feelings, and other concepts, with all of them being social, meaning that they relate to social interactions and situations. For example, Cvetocic (Cvetkovich, 2012) – who's approach is informed by social constructionist research – uses affect in a more generic sense that encompasses emotions and feelings, as well as impulses or desires, that are constructed based on present and past social encounters. In their expected states theory of emotion, Correll and Ridgeway (2006) similarly use affect as an umbrella term. In their work (*ibid*), the concepts of affect, emotion, feeling, and sentiment, as well as their relation, are defined explicitly: Affect is understood to encompass emotions, expressive behaviours, and sentiments towards others, whereby affect may vary from intense flashes of feelings to more enduring attitudes towards others (e.g. liking or disliking someone). Emotions refer to the actual subjective experiences of feelings in a particular situation, which may vary in duration and intensity and might be accompanied by bodily sensations (*ibid*). In the sense of attitudes, sentiments are a summary of past feelings that reflect expectations for future feelings and reflect a more stable affective appraisal, e.g. towards a person or a specific relationship (*ibid*). When interacting with others, affect encompasses the engage-

ment in observable behavioural displays and expressions of emotions or other ‘socio-emotional’ behaviour (ibid).

Note that the interesting point here is not necessarily (only) how scholars relate and define affect, emotions, feelings, and the like, but also that their conceptions of affect encompasses ‘outside’ observable aspects during the interacting with others. However, conceptions of affect vary greatly in the ‘inside-outside’ regard. In broad terms, affect may be understood as a pre-personal kind of force (e.g. (Massumi, 2002; Thrift, 2004)) that provides the ability to affect and be affected (e.g. (Shouse, 2005)). Lindgren (2007), on the other hand, defines affect (in the context of digital media) as a kind of attachment; “a force that attaches people to certain communities, networks, topics, tools and so on” (Lindgren, 2007, p. 136). In alternative conceptions, affect has been defined as an unconscious, pre-personal or non-subjective potential or intensity (e.g. (Massumi, 2002; Shouse, 2005)), or as hard-wired or innate biological system in humans (Ekman, 2005) in which affect is part of a physical input and output system in relation to identifiable emotions such as joy or disgust (Tomkins, 2008). Affect has also been described as embodied meaning-making, involving subjective and intersubjective aspects (e.g. (Wetherell, 2012, 2015)), or as a kind of force or intensity that exists somewhere ‘in between’ embodied processes, sensoral experiences, and the naming of an emotion (Cho, 2015).

What appears to be of interest in these conceptualisations is that there seems to be a kind of dispute regarding the ‘where’ of affect. Affect is conceptualised either as something unconscious or pre-personal that is void or independent of personal significance and meaning, as something that is personal, subjective, and/or and consciously experienced ‘inside’, or as something that is expressed and observable on the ‘outside’ of the (human) body and/or the brain, i.e. something that is (to some extent) relational and social (Ahmed, 2004; Cromby & Willis, 2016; Leys, 2011; Li, 2015; Wetherell, 2012). On the other hand, some authors (e.g. (Ahmed, 2004)) place affect neither on the inside, nor the outside, since affect is not a pure inside-property, or expresses and owned by an individual, but formed and presented in dialogue with the outside. In this sense, affect – or at least some aspects of it – can be observed as social or relational phenomena during social interactions.

This is an important point because conceptualising affect as something unconscious or pre-personal, that is strictly distinguishing affect, emotions, and feelings as something either social or non-social, is problematic for social scientists, and particularly empirical field and/or qualitative researchers, who mean to investigate affect in social interactions (Li, 2015; Wetherell, 2012).

Concerning affect as a phenomenon, Wetherell (2015) points out that “the unit of analysis for social and cultural research on affect, is not some kind of inarticulable, momentary, spurious, hard-to-detect, pre-conscious judder.” (Wetherell, 2015, p. 152). In an analysis of affect in sociology, Li (2015) notes something along similar lines:

the abstract ontological distinction between affect, emotion, and feeling does not help in translating affect to field-work research: how does an ethnographer observe affect? How does a researcher ask informants to articulate something said to exist before consciousness and beyond signification? (Li, 2015, p. 20).

That is to say, a socio-cultural empirical researcher has to apply a conceptualisation of affect that overcomes the dichotomy of affect and feeling, and affect and emotion, as well as dichotomous debates about the relation of mind and body, and the inside and the outside of affect (Cvetkovich, 2012; Li, 2015; Wetherell, 2012, p. 20).

With this in mind, affect is used as an umbrella term encompassing emotions and feelings within the context of this thesis. Given the research topic and proposed methods (chapter 4.3.4) of the present thesis, affect has to be conceptualised as something that is empirically accessible and (at least) partially observable from the outside. Thus, affect is conceived of as encompassing both positive and negative affect as observable aspects of emotions and feelings in social interactions, whereby observable means either third-person observations (e.g. of emotional expression in form of facial expressions, gestures) or via subjective first-person reporting in interviews (e.g. reflection on feelings).

4.2.1.3 Power and Affect

Despite the apparent differences in the above conceptualisations of both power and affect, most theories regarding power, as well as affect, seem to converge towards the consensus that both aspects are relevant in and for social interaction. In fact, several scholars have – theoretically and empirically – related the concepts of power and affect with regard to social interaction and claim that they are strongly intertwined. Heaney (2011), for example, refers to power and affect (in the sense of emotion) as “conceptual twins in need of serious theoretical reunion” (Heaney, 2011, p. 259). In his argumentation for a conceptual integration of theories on power and emotion, Heaney (2011) criticises that many theories of affect – with only a few exceptions – do not adequately address power relations in their conceptual framework. Conversely, many power theorists fail to incorporate emotions within their theoretical analysis (ibid). Heaney’s (2011) main point appears to be that, if it is the case that all social relations are to be seen as power relations, then they

should also be seen to imply relations of affect because “relations of power are emotionally valenced” (Heaney, 2011, p. 259).

There are some theories that do address both power and affect in social interactions. One of the most comprehensive theories concerning power and affect is probably Kemper’s power-status theory of emotion (Kemper, 1991a, 2006a, 2011).

Kemper's theory describes a model for analysing social interactions and relations in which emotions are systematically derived from the structures of social relationships along the dimensions of power and status. According to Kemper (Kemper, 1991a, 2006a, 2011), these are the dimensions along which actors perceive, organise, and understand their relationships. In his later collaboration with Collins (Kemper & Collins, 1990), Kemper expanded this theory to include the aspect of expected states, which has later also been refined by Thamm (2004) .

In a nutshell the power-status theory of emotions (Kemper, 2006a, 2011) states that developments and changes regarding power (tied to status) are accompanied by changes in affective behaviours and displays, including expression and experience of emotions (in the sense of the basic emotion paradigm after (Ekman, 2005)).

When experiencing a gain in status, individuals will experience positive emotions, such as well-being or satisfaction (Kemper, 2006a; Kemper & Collins, 1990; Turner, 2009). They will also ‘give off’ these emotions to others, which in turn, increases the flow of emotions between those receiving and those giving status (ibid) But when individuals lose status, they experience negative emotions. For example, if individuals see themselves as the reason for this loss, they might experience shame or embarrassment, or feelings of sadness (ibid). If they believe that others are the reason for their experienced loss of status, they will feel aggression and try to re-gain power (ibid). Furthermore, expectations regarding gains or losses of power (in terms of status) greatly influence affect (Kemper & Collins, 1990; Thamm, 2004; Turner, 2009). Individuals who expect to gain power, but do not, experience fear or anxiety, or a loss of confidence (ibid). But when they experience a gain in status although they have not expected it, they will feel positive emotions and gain self-confidence. Expectations furthermore influence the intensity of relational dynamics and the actions and behaviours towards others with regard to expected gains/losses of status and emotions. For example, if an individual has expected a gain in power, but did not experience it, and blames him/herself for this, then he/she will experience more intense feelings of sadness or shame (ibid). If an individual expected to gain a higher level of status, but did not, and blame others for this circumstance, then they experience more intense levels of anger. Conversely, if indi-

viduals did not expect a gain in status, or receive a greater gain than expected, their feelings of satisfaction will be more intense and they express positive emotions towards those who caused their unexpected gain in status (ibid).

Other social scientists too have addressed the relation of power and affect, and how power and affect influence action – or possibilities for action – in social relationships. For example, in their expectation states theory of emotions Correll and Ridgeway ((2006), see previous chapter), too point out that social interactions depends on (socially) expected states, e.g. based on an actor's socio-cultural environment, to determine possible actions and interactions along the relations of power and affect. They note that interpersonal status hierarchies have powerful affective implications on the participants in social interactions, whereby affect encompasses emotions, expressive behaviours, and sentiments towards others (ibid).

In a similar vein – but based on a social constructionist understanding of social interaction – Abu-Lughod and colleagues (1990) state that power relations determine what can and cannot be said about the self and affect. Via talk and dialogue (about emotions) power and status differences are asserted, reinforced, or challenged (ibid).

Wetherell (2012) too points out the strong relation of power and affect, and their crucial role in social interactions: “power works through affect, and affect emerges in power” (Wetherell, 2012, p. 16). Her understanding of affect is, however, informed by social practice theories, which she positions within social psychology (ibid). When relating affective practice to habitus (Bourdieu, 1977), Wetherell (2012) points out that affect can maintain, increase, and/or diminish power.

In summary, despite the similarities and differences regarding the specifics of power and affect presented in the above chapters, most approaches seem to converge towards the consensus that social interaction involves both power, as well as affect, and that power and affect are related to each other. Thus, questions about social interaction lead to issues regarding power and affect, and all kinds of social interaction involve power as well as affective dimensions. In this sense, power and affect should be meaningful analytical dimensions to investigate and describe social interactions.

Concluding, the above theoretical insights pertaining to power and affect and their role in social interaction in conjunction with research in relation to the CASA paradigm and the study of pedagogical agents (see chapter X), suggest that human and (potentially) also non-humans are social entities that can take part in social interaction processes. Since social interactions involve affective dimensions, to engage in social interaction means to also be involved in affective processes.

Furthermore, all social interactions are organised along and involve dimensions of power. To be in a social interaction means to be involved in power dynamics. Finally, power is related to affect and all social interactions thus involve both power and affect to some extent. For the phenomenon of social interaction with ITS this consequently means that, if this interaction is of a social kind, then an investigation of this interaction must reveal instances of power and affect to some degree – at least from a social science perspective.

4.3 Methodology

4.3.1 How to Investigate Power and Related Affect?

When asking how power and related affect are involved in social interaction RQ2 addresses three aspects: social roles, interactions and their situational structure, and social relations in terms of ongoing social interactions. Within this context, the interaction between a human learner and the ITS represents a direct, dyadic interaction in terms of a micro-level social interaction. That is, an interaction consisting of bi-directional social actions in which these actions are oriented towards each other, i.e. towards the other social actor. Of interest are the participants (the actors and their roles) of this interactions, their actions, and behaviours along the dimensions of power and related affect. The interest in the social role is based on sociological theory, e.g. role-theoretical and dramaturgical approach from Goffman, including ‘face work’ and the presentation of self when either maintaining/saving or losing face with respect to social order and expectations (Goffman, 1956), as well as Kemper’s relational reading of said theories of Goffman, as well as his interpretation of the work of Collins with respect to power (Kemper, 2011). In essence, dramaturgical approaches suggest that social actors strategically ‘play’ out their social role in interactions to be perceived in a way that is in accordance with their role and associated social norms, rules, or expectations ((Goffman, 1956; Kemper, 2011)). The interest in the social role also stems from research on pedagogical agents and their instructional roles (see chapter 4.1.3), and the idea that humans may apply (role) expectations – and associated actions and behaviours – they would have of human tutors to these agents (A. Baylor & Kim, 2005; Gulz et al., 2011; Veletsianos & Russell, 2014). Of interest is how these roles are perceived and what kind of actions or behaviours are (not) expected of them, as well as what happens – in terms of power and related affective dynamics – when such expectations are (not) met. Since social action and social structure are intertwined (Giddens, 1984, 2006), the structure in which this moment-to-moment interaction is em-

bedded (i.e. in the sense of a micro-structural approach), is also of interest. At this intersection of action and structure, power and related affect should structure the possibilities (or lack of possibilities) for particular actions and interactions. Relative power between two actors along interpersonal hierarchies and power dynamics during the course of tutoring, as well as how this relates to affect, might be revealed. Furthermore, since past social interaction influence present interactions (Boiger & Mesquita, 2012; Burr, 2015; Mesquita & Frijda, 1992; Wetherell, 2012), the change or development of social roles, actions, interaction (and embedded structures) is of interest. This aspect is – in the context of this thesis – referred to as social relationship in the sense of ongoing social interactions. This could reveal insight regarding the change and development of the social interaction over time (i.e. across the three tutoring sessions) and how power and related affective dynamics are established, changed, or (de-)stabilised.

Considering the vast amount of existing – and contradictory – theories on both power and affect, as well as their relation, the present study aims to ‘take a step back’ and approach the phenomenon of social interaction with ITS in an open and exploratory manner, based on the qualitative methods provided by Grounded Theory (GT). For the purposes of the planned study, three consecutive tutoring sessions about the history of the University of Vienna were designed using the dialogue-based ITS AutoTutor, which also incorporates a conversational agent (chapter 4.3.2). Data collection (including the observation of the tutoring session and interviews), as well as data analysis, is based on the constructive approach to Grounded Theory by Charmaz (2006), and Situational Analysis ((Clarke, 2018; Clarke et al., 2015), see chapter 4.3.4).

In line with constructivist versions of grounded theory, it is important to make underlying and/or prior assumptions (see first paragraphs, this chapter) with regard to the phenomenon of interest explicit to ensure a reflective and critical stance towards these ‘extant’ concepts (Charmaz, 2001a, 2006, 2008). Although the researcher is explicitly not considered to be an objective ‘blank slate’ without prior knowledge, one must not mindlessly impose or project such pre-existing theoretical constructs onto the data. Grounded theory is meant to be used to construct mid-range theories about a phenomenon, whereby this theory is ‘grounded’ in the data (ibid). However, the researcher may critically relate such extant concepts – which requires knowledge about different theories as well as one’s own theoretical assumptions – to the constructed grounded theory in later stages of the analysis; if they ‘match’ the data (see chapter 4.3.4, theoretical coding).

Thus, RQ2 and the aspects of social interaction it focuses on are supposed to serve as a starting point for the planned grounded theory analysis. It functions as a kind of initial focus and orientation in line with the ‘theoretical playfulness’ (in the sense of (Charmaz, 2006; Thornberg & Charmaz, 2014)), openness, and reflective practice that constructivist grounded theory allows for. However, how exactly power and affect manifest in and shape the interaction with the ITS, how these concepts relate to each other, as well as how they influence the interaction partners and the interaction itself, is to be determined throughout the process of data analysis and theory construction. The resulting grounded theory could differ from the assumptions – as well as the theories in chapter 4.2.1 – outlined above. Hypothetically, this includes the possibility of power and/or affect playing a different role in social interaction than the assumptions outlined above suggest. It also includes the possibility of power and/or affect playing a minor or no role, not being relevant, or only occurring in relation to other, more prominent aspects or dimensions of interaction.

4.3.2 Research Setting: Tutoring Session Design

Dialogue-based ITS that incorporate a conversational agent – which are explicitly modelled along human-like dialogue structures and tutor moves – appear to be an interesting ‘test bed’ for an investigation of social interaction, which is why *AutoTutor* was chosen for this investigation. The ITS *AutoTutor* (Graesser, D’Mello, et al., 2012; Graesser et al., 2005) has been used for the purposes of designing a pilot study within the context of the present thesis. Using the 2015 version of the *AutoTutor* Authoring Tool (ASAT, see e.g. (Nye et al., 2014) for general information on ASAT), provided by the University of Memphis, *AutoTutor* has been used to implement three consecutive tutoring sessions about the history of the University of Vienna. Each tutoring session has been designed to last about twenty minutes. Each session is based on the previous one(s), contents and questions get more related/complex with each session.

Note that these tutoring sessions have been planned/designed, but were only partially implemented and debugged (session 3 not finished). Furthermore, it is currently not possible to use those exact tutoring sessions for purposes of the proposed study (see also chapter 6.1). While the curricular content and session structure for all three sessions has been finished and (except for session three) and been implemented via *AutoTutor*/ASAT (XML scripts), all sessions have neither been fully debugged, nor have they been pre-tested with participants. Consequently, the data collection and analysis as it will be described in chapter 4.3 has not taken place.

The following chapters will provide a detailed description on the main aspects of AutoTutor, including its performance and dialogue on a behavioural level of learner-student interaction and dialogue moves, as well as the technological level and its concrete computational mechanisms. Furthermore, the concrete way in which AutoTutor was used to create tutoring sessions in terms of curriculum content and tutoring session structure, will be described.

4.3.2.1 AutoTutor

AutoTutor has been developed in 1997 by the Tutoring Research Group at the University of Memphis (Graesser et al., 2001, 2014). AutoTutor is an ITS that incorporates a conversational pedagogical agent capable of producing speech, facial expressions, and some rudimentary gestures. In AutoTutor, student-tutor interaction is based on natural language. Several versions of AutoTutor exist. AutoTutor has been used to implement 1-on-1 dialogue with an ITS-agent and one human student (Graesser et al., 2001, 2014), as well as triologue-based settings in which two ITS-agents, e.g. a tutor and a student peer agent, interact with the human learner (D'mello & Graesser, 2012; Graesser et al., 2001, 2005; Person et al., 2001). AutoTutor has been implemented to assist different target groups (e.g. high school students, college students, adults) learn difficult materials across different learning domains including physics, computer literacy, critical thinking, reasoning, and reading comprehension (Graesser et al., 2016a, 2014). While other ITS mainly base their *tutor-student interaction* on problem-solving approaches, e.g. CognitiveTutor (Aleven et al., 2005, 2009) or Andes (VanLehn et al., 2010), AutoTutor organises its interactions on the basis of *dialogue-based strategies* that are modelled along human tutoring situations (Graesser et al., 2016a, 2014). The ITS simulates a human tutor by employing ideal pedagogical strategies and dialogue moves that are also commonly used by human tutors (ibid). As a conversational/dialogue-based ITS, the pedagogical agent holds a conversation with the learner. During tutoring, the learner thus interacts with AutoTutor via natural language (in most cases by typing on a keyboard) and the conversational interface agent delivers certain dialogue moves (see following chapter for details) by generating speech, facial expressions, and/or gestures.

On an abstract level, AutoTutor can be considered to include all of the *four major classical components* of ITS (see chapter 4.1.2.2) and/or architectural elements and computational mechanisms that serve the component's main purposes. Note, however, that its architecture appears to be more fine-grained and differs from the traditional component view (for details on the AutoTutor architecture see e.g. (D'mello & Graesser, 2012; Nye et al., 2014). As pointed out earlier, an ITS' architecture mirrors the pedagogical strategy or philosophy, and systems that are based on con-

constructivist theories of learning and heavily draw on natural language and human-like tutoring strategies, such as conversational ITS, thus emphasise different components than e.g. cognitive tutors, which is represented in their architectural design. In this sense, AutoTutor thus differs from both other types of systems, as well as from the classical component view (see illustration below, figure 3).

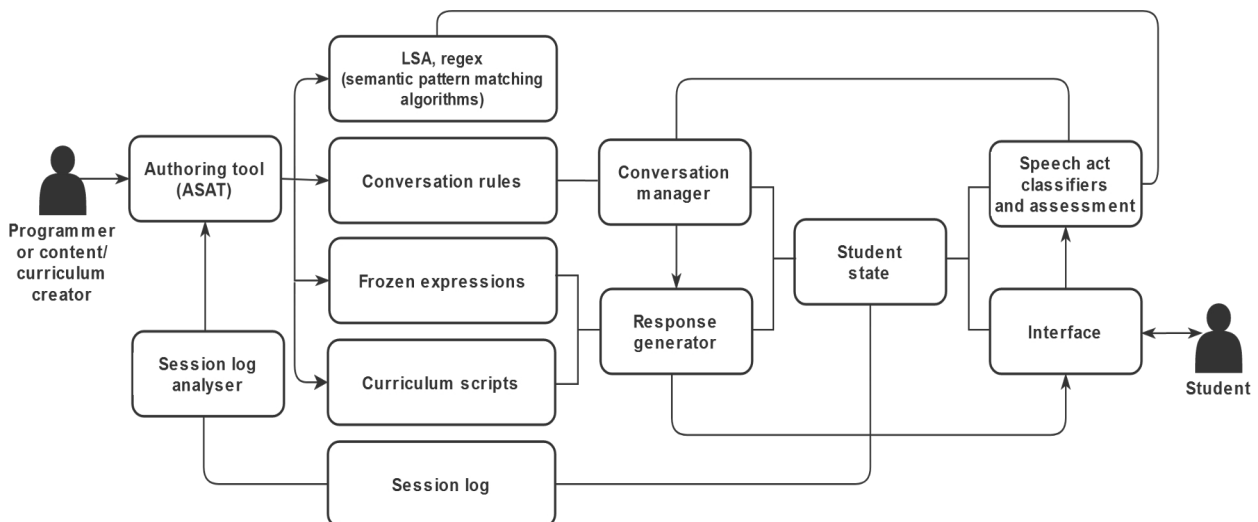


Figure 3: AutoTutor architecture; this illustration is an adapted version of [D'Mello & Graesser 2012, p. 23:12, fig. 2] that has been modified for the purposes of this master's thesis

Nonetheless, AutoTutor does have *domain knowledge* of the subject matter represented via a *curriculum script* which includes specified expectations and misconceptions (see next chapter) that – from a *conceptual* point of view – comprise a multidimensional *domain model* (D'mello & Graesser, 2012; Nye et al., 2014). It furthermore employs a *student model* that track's the student's knowledge, and it evaluates student contributions that are entered via an *interface* (through typing on a keyboard) relative to each expectation and misconception in the domain model (including the curriculum scripts) in order to adaptively select, generate, and deliver adequate *pedagogical strategies* like giving feedback or hints (ibid; for detailed information on fig 3, please see also following chapter). Since AutoTutor is a conversational/dialogue-based system, much modelling is based on dialogue strategies and/or natural language contributions by the learner and its correct classification and evaluation (with respect to the domain) by the ITS based on Latent Semantic Analysis (LSA) and other computational and/or mathematical methods (ibid). How exactly AutoTutor evaluates student contributions and generates adequate tutorial responses (dialogue moves) via *expectation and misconception tailoring* will be reviewed in the following chapter, which describes some of AutoTutor's behavioural and computational mechanisms.

As an ITS, AutoTutor can also be characterised using VanLehn's (2006) *two-loop framework* on a behavioural level (see chapter 4.1.2.3 for general information on the two-loop framework). In AutoTutor, the *outer loop* consists of a set of didactic lessons and associated main questions or problems (D'mello & Graesser, 2012; Rus et al., 2013). This also includes the order of lessons (e.g. mini-lectures, see next chapter) and associated questions/problems, as well as corresponding hints or prompts, which can be dynamically selected based on the student's knowledge or abilities (i.e. the student model, see below). However, the sequential order of lessons and questions is fixed in most AutoTutor versions, including the one used for the purposes of this master's thesis. The interactive dialogue occurs when tutor and student collaboratively construct an answer to a main question, not during the didactic delivery of information (ibid). The correct answer to a question usually requires several aspects and/or sentences of information that together comprise an *ideal* answer (ibid). The tutor assists the student in constructing the ideal answer based on the student's initial response (5-step tutoring frame and EMT-dialogue, see below). In the *inner loop*, AutoTutor consists of this collaborative interaction in which the student answers a question while the tutor helps the student by providing missing information, correcting misconceptions, and/or answering to student questions (D'mello & Graesser, 2012). In the inner loop, AutoTutor tracks the student's performance through embedded assessment, updates the model of the student's understanding/knowledge (the student model) accordingly, and uses this model to provide adequate scaffolding, e.g. via hints (Rus et al., 2013). In a dialogue-based system such as AutoTutor, this modelling heavily relies on the assessment of natural language utterances that occur during tutorial dialogue (see semantic pattern matching algorithm, next chapter). "It is this inner loop that is the distinctive hallmark of AutoTutor." (Graesser, Conley, et al., 2012a, p. 171). The specifics of this inner loop dialogue, as well as some outer loop aspects, will be described in more detail in the following chapters.

Main Dialogue Structure and Dialogue Moves

AutoTutor's dialogue structure and its concrete dialogue moves are inspired by explanation-based constructivist theories of learning and strongly informed by research on the behaviour of human tutors (Graesser et al., 2005, 2014). Essential dialogue structures that are prominent in human tutoring and are also implemented in AutoTutor are: on a general level, the constructivist 5-step tutoring frame, and, more specifically, the *expectation and misconception tailored dialogue* (EMT-dialogue), and a 3-slot conversational turn management.

In the *5-step tutoring frame*, a challenging main question or a problem are posted and worked through in a specific fashion. In a first step, the tutor presents the question relating to a specific subject matter (Graesser, Conley, et al., 2012a; Nye et al., 2014; Pavlik Jr. et al., 2013). In step two, the student gives an initial answer to the tutor's question. In a third step, the tutor provides feedback on the student's response. In step four, the tutor and the student collaboratively improve on the answer in a turn-by-turn dialogue (see also conversational turn management below), and the tutor encourages the student to construct his/her own knowledge. In the last fifth step, the tutor evaluates whether the student understands, e.g. by asking directly if the student understands, or by posing a follow-up question (ibid).

According to Graesser et al. (2014), most human tutors furthermore follow a systematic conversational structure that is referred to as expectation- and misconception-tailored dialogue (EMT-dialogue). By employing EMT-dialogue

Human tutors anticipate particular correct answers (called "expectations") and particular misconceptions when they ask the students challenging questions (or present them with challenging problems) and trace the students' reasoning. As the students express their answers, which are distributed over multiple conversational turns, tutors compare these contributions with these expectations and misconceptions. The tutors give feedback to the students that is sensitive to how well the students' contributions match the expectations or misconceptions. The tutors also produce dialogue moves to encourage the students to generate content and to improve their answers to challenging questions or problems. (Graesser et al., 2014, p. 375).

EMT-dialogue is a pedagogical means to scaffold good student answers. Most human tutors, including expert tutors, use the following dialogue moves in EMT-dialogue, which are also implemented in AutoTutor:

- *Short Feedback* which can be either
 - *positive* (e.g. 'Yes', 'Very, good', and/or a positive gesture like a head nod, and/or a positive facial expression like a smile), or
 - *negative* (e.g. 'No', 'not quite', or gestures like a head shake and/or frowning facial expression), or
 - *neutral* ('Hmm', 'Uh-hm', 'I see..', and the like)
- *Pumps* in which the tutor makes non-directive statements to get the student to talk about or elaborate on a specific topic, e.g. 'What else?' or 'Can you tell me more about this?'
- *Hints* by the tutor guide the student towards specific answers or action to direct the student along a conceptual path. Hints can be generic statements like 'What about ... ' or 'Try to think about ... ', or concrete speech acts that lead the student to a specific answer. In this

sense, hints are scaffolding moves to promote student learning by directing the students focus towards relevant aspects of the subject domain.

- *Prompts* are leading questions or statements that are meant to prompt the student to articulate a specific word or phrase.
- Assertions, in which the tutor expresses a fact or state of affairs (ibid).

Similar to a human tutor, Auto-Tutor can use *pump-hint-prompt-assertion cycles* in EMT-dialogue to interactively assist a human learner in formulating the correct answer to a specific problem (Graesser et al., 2014). During this process, other dialogue moves also occur frequently. Students might ask questions (either on-topic or off-topic), which the tutor answers immediately, and also occasionally express misconceptions, which too are immediately addressed and corrected by the tutor; either by providing a short answer, a summary, or a kind of mini-lecture (ibid). Additional tutor dialogue moves of this kind that are implemented in AutoTutor are:

- *Answers* to the learner's questions
- *Corrections* of student errors and/or misconceptions
- *Summaries*, in which the tutor briefly summarises previously stated information that is pivotal to the subject matter
- *Mini-lectures* through which the tutor presents didactic content and more detailed information on a particular topic relevant to the subject matter
- *Off-topic comments*²¹ in which the tutor expresses information unrelated to the subject matter at hand (Graesser et al., 2014)

These dialogue moves and the questions/problems AutoTutor can provide tutoring for are stored in the *curriculum scripts* (D'mello & Graesser, 2012; Rus et al., 2013). That is, the tutor covers a curriculum with specific didactic content, which includes information on the subject matter (*expert knowledge* on the domain), as well as a set of questions/problems that address the didactic content. This content can be presented in a mini-lecture. The curriculum script also includes specified ideal answers, expectations, misconceptions, as well as hints and prompts associated with each question, as well as other 'inner loop' information in the sense of a multidimensional *domain model* (Graesser, Conley, et al., 2012a; Ma et al., 2014; Nye et al., 2014). The tutor then

²¹ All of these dialogue moves, except for off-topic comments, have been implemented in the designed tutoring session. Off-topic input by the learner only results in the tutor responding with "I don't see how this connects to today's topic. Let's get back to our session" (or a response of similar wording), followed by whatever tutor response was planned to be selected based on the previous student input.

works on these questions together with the student until all content and the associated questions and expectations are covered, which involves EMT-dialogue and the corresponding dialogue moves.

During EMT-dialogue, AutoTutor can detect misconceptions and expectations in terms of similarly to an expected ideal answer (stored in the curriculum script) and adaptively responds with one of the dialogue moves listed above. Concerning the technical level and concrete implementation and computational mechanisms of AutoTutor, all the components of EMT-dialogue are implemented in AutoTutor by means of computational linguistics and statistical representations (D'mello & Graesser, 2012; Graesser et al., 2014; Nye et al., 2014). Using syntactic parsers and Latent Semantic Analysis (LSA, cf. (Graesser, Conley, et al., 2012a)). AutoTutor is capable of *classifying speech acts* produced by the learner (i.e. student answers or other contributions typed into the AutoTutor *interface*) into specific *speech act categories*, such as questions, statements, meta-cognitive expressions (e.g. 'I don't know'), acknowledgements, or short responses (ibid). In terms of *pedagogical modelling and tutorial strategy selection* AutoTutor adaptively responds based on the student input (see conversation management and response generation below), whereby student contributions are evaluated based on a *semantic pattern matching algorithms* that compares the student input to each expectations and misconceptions associated with a pre-defined ideal answer to the posted main question/problem in the domain:

Semantic pattern matching algorithms determine the match between student input and the expectations or misconceptions, whereas semantic pattern completion mechanisms generate the AutoTutor's discourse moves to guide students to fill in missing information. User input can be compared to expectations and misconceptions through a variety of semantic analysis techniques, including latent semantic analysis (LSA), regular expressions, word-frequency overlap techniques, and logical entailment (Nye et al., 2014, p. 451).

LSA has played an important role in robustly comparing student's input to AutoTutor's expectation and associated misconceptions (Graesser, Conley, et al., 2012a; Nye et al., 2014). Conceptual semantic similarity between the student input and AutoTutor expectations is calculated with an LSA-based *similarity matching algorithm*; a vector representing the semantic content of the student input is created and compared to the vectors representing the semantic content of AutoTutor expectations and misconceptions and produces a *matching score* between 0 and 1 based on the cosine of the vectors (D'mello & Graesser, 2012; Person et al., 2001). When an expectation is missing – e.g. when the student answer does not include a certain expectation and is not or only partially correct – AutoTutor attempts to get the student to articulate it via hints and prompts to fill in the missing information, e.g. a specific world or propositions. Associated hints and prompts are selected to maximise the LSA cosine match score, i.e. to maximise pattern completion. When

all the student contributions regarding a question and each of its expectations are matched, and if all cumulative comparisons exceed a certain pre-defined *threshold*, the expectation is considered covered. During this process, selection of expectations is based on the pedagogical principle of the *zone of proximal development or frontier learning* (Vygotsky, 1978), meaning that AutoTutor builds on what the student managed to provide as an answer and uses scaffolding (e.g. via hints) to assist the learning in moving towards a more ideal stage in the domain in order to construct an ideal answer (Graesser, Conley, et al., 2012a). Formally, AutoTutor selects the expectation that has the highest match score and has not yet been covered (that is, the expectation has a match-score below the specified threshold), i.e. expectation selection is not based on a prescribed sequential order (ibid). If the student input matches a misconception with a high enough score, AutoTutor corrects the misconception and continues (ibid).

These semantic pattern matching and completion algorithms also inform *student modelling*²² (see also ch X, component view of ITS) across conversational turns of the tutoring session. i.e. across different tutor questions, student answers, tutor hints, etc. until the correct answer is articulated by the human students and/or the tutor.

If the student model was completely rebuilt on each student turn, the semantic matches would wildly vary, representing the vicissitudes of the student's ability to provide content in response to different tutor dialog moves. [...] Therefore, the semantic match is calculated using both the current student response alone (local assessment) and the current response concatenated with all previous student responses for the current problem (global assessment). This global assessment represents AutoTutor's model of student knowledge across turns for a particular main problem. (D'mello & Graesser, 2012, p. 23:11)

During tutorial dialogue, including EMT-dialogue across several turns, tutors strategically employ *conversational turn management*. Human tutors generally tend to systematically organise their conversational turns along a 3-slot structure that has also been implemented in AutoTutor (ibid). Any tutor turn has three slots or constitutive parts (ibid). In the first slot/part, the tutor gives short feedback on the previous student/learner turn, which can be positive, negative, or neutral (see dialogue moves listed above). In the second part of the turn, the tutor advances the interaction by delivering either a prompt for specific information, or hints, assertions with correct information, or corrections of misconceptions, or answers to student questions (see dialogue moves above). In the third and final turn slot, the tutor cues the shift in conversation from the tutor to the learner. Specifically, AutoTutor ends a turn with a question or a gesture that is meant to signal the student to start talking (ibid).

²² For specifics regarding the general AutoTutor architecture and details regarding student modelling, as well as more concrete information regarding the semantic pattern matching algorithms see (D'mello & Graesser, 2012; Graesser et al., 2005; Nye et al., 2014).

AutoTutor's *conversation management and response generation* depends on the speech acts expressed by the learner, i.e. a certain student input and its correct classification. If the student asks a question via typing it into the *interface*, AutoTutor either provides an answer (if it has an answer) or generates other dialogue moves, e.g. pumps. If the student provides a meta-cognitive response (e.g. 'I'm lost'), the tutor acknowledges it and provides a hint. AutoTutor generally generates pump-hint-prompt-assertion cycles (see above) during EMT-dialogue and student inputs in terms of speech acts are thus often responses to tutor's questions, hints, or pumps, which are evaluated on quality (semantic pattern matching and completion algorithms). AutoTutor's *conversation manager* can be represented by a "set of "if <state> then <action>" production rules" (D'mello & Graesser, 2012, p. 23:13) or a "finite state transition network", which passes information to the *response generator*, which then generates the actual text of the selected dialogue move to be expressed via the *tutor-student interface*, i.e. by the conversational interface agent via written text and/or speech (ibid).

Note that AutoTutor *cannot* correctly classify, interpret, or match all speech acts that students produce. However, it can simulate human EMT-dialogue sufficiently well to hold a conversational dialogue during tutoring. Similar to human tutors, AutoTutor too only has an "approximate understanding" (Graesser et al., 2014, p. 375) of what the student supposedly knows and of what he/she expresses. Some of the student's speech acts may furthermore be vague, semantically ill-informed, or formulated in an ungrammatical manner. AutoTutor can nevertheless match the student's contributions over several conversational turns accurately enough to the expectations and misconceptions by applying the semantic pattern-matching and pattern-completion algorithms described above. AutoTutor's evaluation of student contributions is comparable to the evaluations of intermediate human experts of the subject domain (ibid). Furthermore, the dialogue quality of AutoTutor is reasonably coherent. Its developers state that "a bystander who reads the tutorial dialogue in print cannot tell whether a particular tutor turn was generated by AutoTutor or by an expert human tutor" (Graesser et al., 2014, p. 376) and that learning gains when using AutoTutor are comparable to those of trained human tutors (ibid).

4.3.2.2 Learning Content and Curriculum

In 2016, using the 2015 version of the AutoTutor Authoring Tool (ASAT), provided by the University of Memphis, the ITS AutoTutor has been used to implement three consecutive tutoring sessions (i.e. three curriculum scripts, using XML) about the history of the University of Vienna.

The reason for choosing the topic of the history of Vienna was the 650 year anniversary of the University in 2015, and the fact that, partially due to this event, a vast amount of literature, video, and other media content was available²³. Furthermore, it was assumed that the participants (see chapter 4.3.3), were university students and/or employees, who already knew about the anniversary, and might be interested to learn more about the University's history.

Note that AutoTutor has been applied, among other fields, to physics, scientific methods, and computer literacy, and is generally geared to provide tutoring sessions that promote conceptual thinking and deep reasoning (e.g. complex problem solving, analysing causal relations, making inferences), rather than subject-specific facts (Graesser, Conley, et al., 2012a). However, there is nothing in AutoTutor's design and architecture to prevent the ITS from implementing sessions in which the tutor assists students with learning domain-specific facts (ibid).

In the concrete case of the implemented curriculum, the content does include historical facts and dates, but also addresses how certain historical events came about, as well as what they caused, i.e. how they relate to later events and what role they played in a bigger socio-historical and/or political context. This is, within this master's thesis, referred to as 'historical continuity'.

The concrete curriculum and the anticipated learning goals are depicted in the table below.

23 The following material (including websites, official videos and pictures from the University, as well as literature) was used to create the content of the tutoring sessions and to design the curriculum:

- Gall, Franz, 1965. Die Insignien der Universität Wien (Graz-Köln: Studien zur Geschichte der Universität Wien im Überblick, Bd. 4, 1965).
- Hauch, Gabrielle, 1998. Die Wiener Achtundvierzigerinnen. In: Walter Öhlinger 1998. "Wien 1948. "Das tolle Jahr". Chronologie einer Revolution. 44-51.
- Häusler, Wolfgang, 1979. Von der Massenarmut zur Arbeiterbewegung. Demokratie und soziale Frage in der Wiener Revolution von 1848 (Wien: Jugend & Volk, 1979).
- Heindl, Waltraud und Tichy, Marina, 1990. "Durch Erkenntnis zu Freiheit und Glück ..." Frauen an der Universität Wien (ab 1897) (Wien: Schriftenreihe des Universitätsarchivs, Universität Wien, Bd. 5, 1990).
- Maisel, Thomas, 1998. Alma Mater auf den Barrikaden. Die Universität Wien im Revolutionsjahr 1848 (Wien: WUV-Universitätsverlag, 1998).
- Maisel, Thomas, 1995. "Bellum Latinum". Eine studentische Rebellion des frühen 16. Jahrhunderts in Wien. Historische Anthropologie. 1995;3:389-411.
- Mikoletzky, Juliane, 1998. "- um der Rettung der Freiheit willen!" Das Jahr 1848 und die Folgen am k. k. polytechnischen Institut in Wien (Wien 1998).
- Uiblein, Paul, 1995 Mittelalterliches Studium an der Wiener Artistenfakultät. Kommentar zu den Acta Facultatis Artium universitatis Vindobonensis 1385-1416 (Wien: Schriftenreihe des Universitätsarchivs, Universität Wien, Bd. 4, 2. Aufl., 1995).
- Referat Frauenförderung und Gleichstellung der Universität Wien, 2007. Gender im Fokus : Frauen und Männer an der Universität Wien. Wien: Referat Frauenförderung und Gleichstellung der Univ. Wien ; Erscheinungsverlauf: [1.]2007 - [Hrsg.: Referat Frauenförderung und Gleichstellung der Universität Wien. Für den Inh. verantw.: Sylwia Bukowska. Texte, Diagramme: Sylwia Bukowska ; Bettina Kubicek] ; Wien: Referat Frauenförderung und Gleichstellung der Univ. Wien.
- Official Website "650 Plus" of the University Vienna Archive on the history of the University [<https://geschichte.univie.ac.at/en/>].
- Official University of Vienna videos "Welcome to the University of Vienna" [<https://www.youtube.com/watch?v=fb3T4CUGHTI>] and "University of Vienna – Imagefilm" [<https://www.youtube.com/watch?v=VY5iQUME4-M>].

	Content/Topic	Learning Goals
Session 1: University foundation and its position between monarchy, church, and state	<ul style="list-style-type: none"> • Foundation of the University • Relation of university, church, and state • Reformations of the University under the Jesuit era and the Austrian Monarchy 	<ul style="list-style-type: none"> • Knowledge of the University's founding, significant historic events leading to the 1848 student uprisings (= session 2) • Knowledge of status of the university in the context of church and state • Understanding university structure and reformations • Understanding historical continuity, relations of historic events
Session 2: University power structures, student protests and uprisings in the history of the University	<ul style="list-style-type: none"> • University organisation and power structure • 1848 Vienna Revolution and student uprisings in 1848 • Changes in university structure and • Other student uprisings and/or demonstrations (focus on protests of 1968, 2009) 	<ul style="list-style-type: none"> • Knowledge of university structure and organisation • Understanding the reforms under the Austrian monarchy • Monarchy reforms and their effects on the 1848 uprising • Understanding contexts, motives, and reasons of student demonstrations, protests, and/or uprisings • Understanding historical continuity, relations of historic events
Session 3: History of women at the University of Vienna	<ul style="list-style-type: none"> • Women's participation in the 1848 uprisings and its consequences • Founding of the Vienna Democratic Women Association • Women's right to study at the University of Vienna 	<ul style="list-style-type: none"> • Knowledge of role of women in 1848 Vienna Revolution • Understanding consequences for traditional gender roles and norms, and women's right movement • Knowledge of history of women's right to study • Understanding historical continuity, relations of historic events

Table 6: Overview of the curriculum

4.3.2.3 Tutoring Session Structure

For the purposes of the pilot study, three consecutive tutoring sessions have been designed using the 2015 version of ASAT provided by the University of Memphis. In these sessions, the male tutor interacts with the learner via natural language (dialogue), while the learner types his/her replies or questions via a keyboard. In general, each tutoring session is structured to have an introduction, in which the tutor announces the session's topic and agenda. This is followed by two to four 'mini-lectures' and repetition questions concerning the respective lecture (see instructional modes below). Each session contains between two to four lectures. After a lecture, the tutor asks two or three questions – and may provide two to three hints and/or prompts per question, if re-

quired – and evaluates the student’s answer, upon which the tutor will provide either positive or negative feedback, a concluding/ideal answer (depending on the degree of correctness of the student’s answer), and then proceeds to transition to next lecture. In case there are no more lectures, the tutor will conclude the session by summarising key topics, thanking the student for coming, stating the topic of the next session, and that he looks forward to seeing him/her in about a week. Each session ends with a final conclusion by the tutor. Every session is designed to take approximately 20-30 minutes and has the same overall structure.

Note that this session design is intended to represent an ‘average’ or ‘typical’ tutoring session, which also features different tutor and student roles and interactional dynamics, and, thus, (potentially) also different power-dynamics and associated affective behaviours and actions. During the sessions, the tutor is intended to display both expert-like and motivator-like qualities and characteristics (see below). Each tutoring sessions features two different instructional modes: 1) a lecture-like mode, in which the tutor provides information and cannot be interrupted by the learner, and 2) a repetition-like, dialogue-based mode following each of these ‘mini-lectures,. During the lectures, the tutor talks and provides historical facts while also using and referring to pictures or videos during his speech. Afterwards, the tutor asks the learner questions to repeat the lecture’s contents. Here, the learner answers questions concerning the preceding lectures, while the tutor gives feedback, and provides pumps, hints, and prompts in an engaging and supportive manner, as well as concluding remarks (EMT-dialogue, see chapter X). The two instructional formats, i.e. the lecture mode and the question-repetition-mode, are repeated until all questions regarding the previous lecture have been answered.

The design of the sessions structure and the two instructional modes are based on AutoTutors general design and capabilities, as well as work on optimal (human) tutoring strategies and behaviours, the design of instructional agents, and operationalisations of pedagogical roles. The structure of the *lecture mode*, in which the tutor is intended to predominantly show attributes associated with expert tutors, as well as the overall structure, (opening including introduction to the session and agenda setting, lectures, conclusions, closing/session end), draws upon work on (human) tutors, which indicate that expert and highly effective tutors, who posses and display domain-specific and general and subject-specific and general pedagogical knowledge, tend to set an agenda, provide direct instruction and knowledge via holding lectures, and provide conclusions (Cade et al., 2008; Lepper & Woolverton, 2002; Xu, 2012), as well as work on expert versus non-expert dialogue moves and interaction patters (Cade et al., 2008; Lepper & Woolverton, 2002; Lu et al.,

2007; Xu, 2012); (Lehman et al., 2012). Additionally, the tutor interactions and tutor moves during the lecture mode are informed by work on computational models of expert tutoring (Eugenio et al., 2006; Lu et al., 2007; Olney et al., 2012), and studies on expert-like computer tutors and their qualities or characteristics (Baylor, Amy L, Kim, 2005; Y. Kim, 2007; Y. Kim & Baylor, 2006b; Liew et al., 2013a); see also chapter ped. roles/instructional roles). This mode was implemented using rigid packs (i.e. to deliver contents in which a student input does not interrupt the tutor's introduction, lecture, correct answers and conclusions, or transitions) and frozen expressions (e.g. fixed positive/negative responses like "mhm, yes [...]") or fixed meta-cognitive responses like "do you need more time?" or "I am not sure what you mean, can you repeat that, please try to be as specific as possible."), and fixed gestures and facial expressions (e.g. looking down) in ASAT/AutoTutor's curriculum script.

The *question-repetition-mode*, in which the tutor is meant to predominantly display motivator-like qualities, is informed by the work of Baylor, Kim, and others (A. Baylor & Kim, 2004; Baylor, Amy L, Kim, 2005; Gulz & Haake, 2006; Veletsianos et al., 2010) studies on instructional/pedagogical agents as motivators and motivator-like qualities of tutorial agents (A. Baylor & Kim, 2004, 2005; Y. Kim & Baylor, 2016), studies on computational models and roles of pedagogical agents as mentors (incorporating both expert-and motivator-like qualities), peers and learning companions (A. Baylor & Kim, 2005; Y. Kim & Baylor, 2016), and collaborative learning strategies in human and computer tutors (Baylor, Amy L, Kim, 2005; D'Mello et al., 2010, 2011; Y. Kim & Baylor, 2016) were considered during the design of the tutoring sessions. This instructional mode is implemented via AutoTutor's EMT-dialogue, including hints, pump, prompts, and feedback functions, as well as frozen expressions and fixed (emotional) facial expressions and gestures, like smiling, nodding, pointing, or eye gaze.

The following illustration depicts the session structure in more detail.

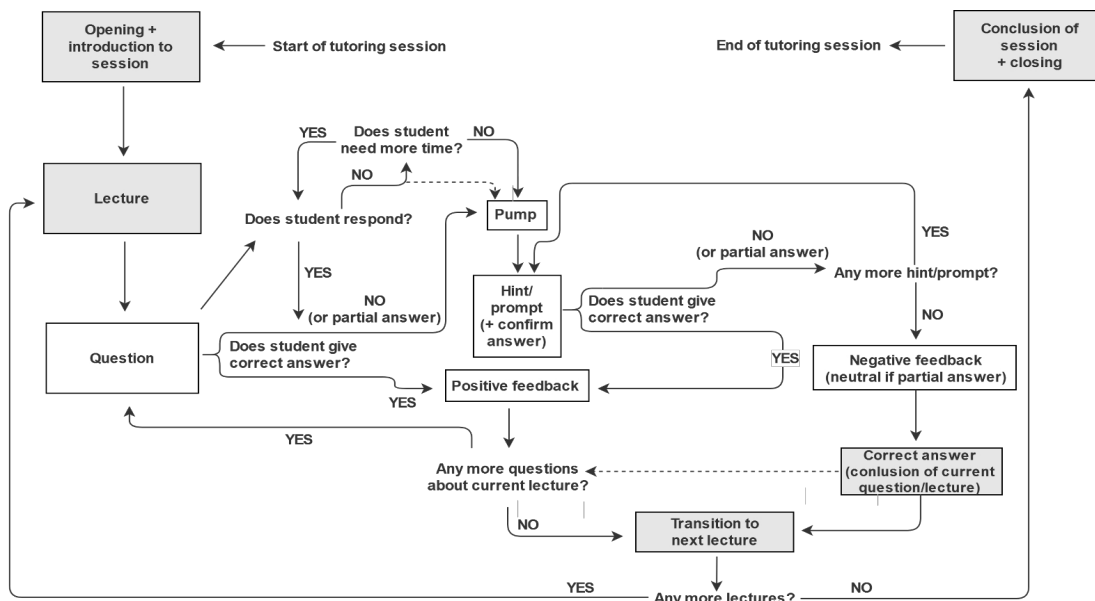


Figure 4: Simplified tutoring session structure; design of tutor actions and dialogue moves. Grey fields represent modes in which the tutor cannot be interrupted by the student (lecture mode). The elements comprising the question-repetition-format (based on EMT-dialogue) are depicted in white.

As shown in the picture above (figure 4) the tutoring session starts with a short opening by the tutor, in which the tutor welcomes the student/learner, sets the agenda, and announces the topic and contents of the present tutoring session in a brief introduction. Before the session, the student can also enter his/her name into the system, which will be used by the tutor to address the student throughout all three tutoring sessions. In the first session, the tutor will also introduce himself to the student as his tutor (preliminary named Eddy Smith). In the second and third session, the introduction also includes a short repetition of the main facts from the previous tutoring sessions that are relevant for the current topic. This introduction is followed by a lecture, in which the tutor tells the student specifics about the current topic, supported by pictures and short videos to aid the student's understanding. The tutor then asks the student questions regarding this mini-lecture to repeat the content of the lecture. During this question-repetition-mode the tutor also provides hints and prompts in case the student does not know the correct answer. In case the student gives an answer to the question within 20 seconds, the tutor will evaluate the answer. If the student's answer is correct, the tutor will provide positive feedback and proceed to next question. The following example dialogue from the first session design illustrates this. The following is an excerpt from a session log of the first tutoring session that stems from the implementation and testing phase of the tutoring sessions. It shows the first question concerning the first lecture, as well as the following student and tutor responses:

Tutor: Now, let's see if you have been paying attention. In which year was the University of Vienna founded? [Main question]

Student: the University of Vienna was founded in 1365 [Correct answer]

Tutor: That's correct. Well done! [Positive Feedback]

If the answer is incorrect or only partially correct, the tutor will assist the student to find the correct answer by initially providing a prompt and hint. If the student then still does not give the correct answer, the tutor provides another hint or a prompt, respectively. If the student does not or only partially know the answer despite all the hints or prompts, the tutor will give negative (or neutral, in case of partial answers) feedback and provide the correct answer. If there are no more questions, the tutor proceeds to transition to the next lecture.

Tutor: Now, let's see if you have been paying attention. In which year was the University of Vienna founded? [Main question]

Student: 13 something. [Incorrect answer]

Tutor: Okay, I'm sure you know this. Please try to be as specific as possible. [Pump]

Tutor: Let me help you a little, Katharina [student name]. Together we can find the correct answer. The University of Vienna was founded in the second half of the 14th century. Can you remember the foundation year now? [Hint]

Student: the University of Vienna was founded in 1335 [Incorrect answer]

Tutor: Let's try this. The first university building was opened in 1385. That was 20 years after the University's foundation. This means that the University of Vienna was founded in ___? [Prompt]

Student: 13something [Incorrect answer]

Tutor: No. Sorry, Katharina [student name] that is not correct. [Negative Feedback]

Tutor: The University of Vienna was founded in 1365. [Tutor provides correct answer]

Tutor: Okay, make sure to remember this. For now we should proceed with our session. [Transition to next lecture]

If the student does not respond at all for 20 seconds after the tutor has asked a question, the tutor will ask if the student needs more time to think about his/her answer. In case the student does then – again – not respond in time when providing an answer, the tutor will use a pump and hint (if he has not provided a hint yet), or another hint (after the initial hint), or a prompt (if the tutor has provided all hints for the correct answer) to engage and assist the student. The tutor will tell the learner the complete and correct answer if the student provides an incorrect or partial answer

to the question based on the hint or prompt. This is illustrated by the following example dialogue from designing the first session.

- Tutor: Let's see if you understood this. Why did the University lose prestige and what did King Ferdinand do to counteract this development? [Main question]
- Tutor: Do you need more time to think about it? [After 20 seconds of no student input]
- Student: the reformation attacked the catholics and the pope [Answer is partially correct]
- Tutor: Okay, what else can you say about it? Can you be more specific? [Pump]
- Tutor: I am sure you know the answer, Katharina [student name]. Just think about what happened during the Reformation. How could this affect a papal institution such as the University of Vienna? [Hint]
- Student: the university suffered a great loss of prestige [Answer is partially correct]
- Tutor: Alright. Due to the Reformation the University of Vienna as a papal institution suffered a great loss of prestige because the Pope, the Catholics and their beliefs were questioned [Assertion. Tutor confirms/provides correct answer based on hint]
- Tutor: Okay, let's repeat. The University of Vienna was a papal institution, but the Catholics, including the Jesuit Order, faced a great loss of prestige during the Reformation. [Summary]
- So, if King Ferdinand wanted to ensure the Catholics' power over the university, what could he do? [Hint]
- Student: he installed the Jesuit order at the university [Answer is correct]
- Tutor: Alright. That is correct, Katharina [student name]. [Positive Feedback]
- Tutor: Good. But due to continuing tensions between the Jesuits and the university the Emperor did something to ensure the Jesuits' position. Can you remember what this was? [Next/follow-up question]

4.3.3 Participants

Three to five subjects, native English speakers aged 20-35, should participate in all three sessions and corresponding interviews. Informed by criteria and characteristics specified in the work of Andreou and Galantomos (2009) and Aneja (2014), a native speaker of English is classified as an individual that:

- has been born and raised in an English-speaking country (with English as dominant/first language) and/or has spoken only English from infancy on, or

- has been born in a non-English-speaking country, but was raised by at least one English-speaking parent, i.e. a native speaker of English, and learned English from early childhood on, starting before age 6, or
- began learning English before or at age six (e.g. at a school with English as primary language of instruction) in a non-English-speaking environment, and communicates primarily in English (in private and business contexts), or
- is highly educated in and speaks English with high proficiency, e.g. someone working as an official translator.

Furthermore, the participants should have a relation to the University of Vienna (e.g. University employees and/or students) and have a general interest in and motivation to learn about the University's history. No monetary or other (e.g. ECTS) incentives are planned as a remuneration for the subjects' participation.

4.3.4 Means of Data Collection and Analysis: Grounded Theory Methodologies

Data collection and analysis is based on Grounded Theory Methodologies (GTM), which are “inductive methodology that provides systematic guidelines for gathering, synthesizing, analyzing, and conceptualizing qualitative data for the purpose of theory construction.” (Charmaz, 2001a, p. 6396). GTM offers a flexible set of strategies for qualitative data collection and analysis and generally aims to construct and conceptualize mid-range theories about a phenomenon. Theory construction is based on several phases of data collection and its analysis through coding and a repeated comparison of data with data on different levels of analysis: the ‘constant comparative method’ (cf. (Charmaz, 2006; Glaser, 1978, 1998)). In this way, a grounded theory emerges from systematic analyses and interpretation of empirical data – it is ‘grounded’ in it. Rather than being based on existing theories and conceptions regarding a phenomenon, it is derived from empirical data. Like other qualitative methods, GTM aims to generate theories and hypothesis, rather than testing them (Charmaz, 2001a, 2006; Thornberg & Charmaz, 2014). It thus appears to be a fruitful means of generating insights into a new phenomena – or phenomena on which a multitude of contradictory theories exist, e.g. power in social interaction – in order to provide basic orientation and directions for future research.

Note that GTM is not characterised by one concrete form of data analysis and collection. Different forms of grounded theory – which are based on different epistemological premises – exist. Despite their differences regarding their philosophical roots, as well as their concrete application, all versions of GTM:

1. Conduct data collection and analysis simultaneously in an iterative process.
 2. Analyse actions and processes rather than themes and structure.
 3. Use comparative methods.
 4. Draw on data (e.g. narratives and descriptions) in service of developing new conceptual categories.
 5. Develop inductive categories through systematic data analysis.
 6. Emphasize theory construction rather than description or application of current theories
 7. Engage in theoretical sampling.
 8. Search for variation in the studied categories or process
 9. Pursue developing a category rather than covering a specific empirical topic.
- (Thornberg & Charmaz, 2014, p. 155).

For the purposes of the proposed pilot study, a *Constructivist Grounded Theory* (CGT) approach that is complemented by its “methodological sister” (Whisker, 2018, para. 1) *Situated Analysis* (SA, also called situated grounded theory) has been chosen.

Constructivist versions (Charmaz, 2006, 2008) adopt the grounded theory method without adhering to objectivist and postivist assumptions underlying the traditional grounded theory approaches (Glaser & Strauss, 1968), including Straussian Grounded Theory (Strauss, 1987), Glaserian Grounded Theory (Glaser, 1978, 1998), or the variant qualified by Strauss and Corbin (Strauss & Corbin, 1991).

A constructivist grounded theory attends to data collection closely. It does not assume that data simply exist in an external world, awaiting discovery. Nor does it assume that an observer can enter the research scene as a *tabula rasa* [original emphasis]. Rather, what observers see and hear depends upon their prior interpretative frames, biographies, and interests as well as the research context, their relationships with research participants, and modes of generating and recording data. The questions that observers ask of the empirical world shape data that they select from it. Similarly, constructivists view their conceptual categories as constructed through their interpretations of the data rather than emanating from them. Thus, they assume that their resulting theoretical analyses provide interpretive renderings of a reality rather than an objective reporting of the reality. (Charmaz, 2001a, p. 6397)

Constructivist Grounded Theory has been used in different areas of psychological, educational, and nursing research . It is a theory-generating inductive method of qualitative data collection and analysis(Charmaz, 2006, 2008). CGT is well-suited for studying social actions and processes because it places actions and how they are created in social situations and social structures in the focus of its analysis (ibid). CGT can be used to analyse a variety of qualitative data, including field notes and observational data from ethnographic research settings, as well as interview data, or documents. Data are analysed by systematically employing different forms of coding: “Coding is the pivotal link between collecting data and developing an emergent theory to explain these data.”(Charmaz, 2006, p. 46) . CGT-coding has an emphasis on analytic development rather than description and comprises at least two main phases: initial coding and focused coding. Addition-

ally, researchers may engage in the sophisticated process of theoretical coding (for details, see description of concrete process of data collection and analysis below).

Situational Analysis provides different perspectives on and means of data analysis and seem to fruitfully extend and complement CGT. In contrast to CGT, SA focuses on situational aspects (Clarke, 2005, 2018; Clarke et al., 2015). SA supplements the analysis of actions and basic social processes of CGT with an ecological focus on the situatedness of actions, i.e. it centres on the situation as a conceptual structure to focus its analysis. A crucial aspect in regard to the present thesis is that SA explicitly addresses the role of non-humans in social situations (Clarke, 2003, 2018; Clarke et al., 2015). Explicitly taking the non-human into account is in fact an important but sometimes overlooked aspects in the study of social situations and that:

fresh methodological attention needs to be paid to *nonhuman objects* [original emphasis] in situations—things of all kinds. These may include cultural objects, technologies, animals, media, nonhuman animate and inanimate pieces of material culture, and the lively discourses that also constitute the situations we study—from cups and saucers to lab animals to tv programs.”(Clarke et al., 2015, p. 91).

Similar to the way non-human elements are viewed to participate in social interaction in Actor-Network Theory (e.g.(Cerulo, 2009, 2011; Sayes, 2014)), SA assumed that non-human elements can have agency and become actors in the sense of non-human actants (Clarke, 2018; Clarke et al., 2015). Based on work in the area of Science and Technology Studies, SA assumes that human and non-human, including their interactions and engagements, co-constitute each other. In the context of SA, non-human actants are assumed to structurally condition the interactions with the situation through their specific (material) properties and through engagements with them: “Their agency is everywhere” (Clarke, 2015, p. 91). By providing concrete analytical means to elucidate non-human elements (including technologies) in the situation of enquiry, as well as their relation in the situation of concern (including power relations, cf. (Clarke, 2018, p. 91), SA appears to be a fruitful means to investigate the interaction with ITS. Processes of co-construction and co-constitution can be studied through using the situation as the locus of analysis an explicitly including all pertinent non-human elements (Clarke et al., 2015), i.e. animate technologies such as ITS. SA thus seem to complement CGT’s focus on action and social processes and a combination of CGT and SA should adequately capture the social interaction between learner and ITS by accounting for both the human and the non-human ITS as possible actors, their actions, interactions, as well as their relations and embedding in situational structures.

Concerning the concrete application of SA, this kind of grounded theory may also contain aspects of coding (Clarke, 2015, 2018; Clarke et al., 2015), but its main advantage is its focus on diagramming and the creation of cartographic visualisations. Similar to CGT, SA can be used to ana-

lyse different kinds of data and is particularly suited for multi-site and/or multimodal research that includes textual data, visual data (e.g. pictures or videos), archival materials, as well as ethnographic materials such as interview data, observation transcripts or field notes. However, while CGT mainly analyses such data by means of coding, SA draws on the method of mapping. SA comprises three kinds of cartographic approaches:

1. *Situational maps* lay out all major human, non-human, discursive, and other elements in the research situation of concern (Clarke, 2003, 2018; Clarke et al., 2015).
2. *Social world/arena maps* are ecological cartographies of major (collective) sites of action and relations in the situation and lay out collective actors, non-human elements, and the arenas of commitments they are engaged in in ongoing negotiations and relations (ibid).
3. *Positional maps* plot positions that are (not) taken or articulated in the situation (Clarke, 2005, 2018). They map out the full array of positions (including contradictory or competing positions) that are taken – or not taken – in the data vis-a-vis discursive axes surrounding issues in the situation (ibid).

All three kinds of maps contribute to an analysis of the relational ecology of the situation. Taken together, the three kinds of maps answer the following basic questions:

What and who are involved in the situation? What is going on in the situation? What is at stake in the situation for the different entities involved? What conditions or possibilities are offered by the situation? What is rendered essentially impossible? What and who are rendered invisible or marginalized [...] Why and how do these conditions *matter* [original emphasis]? To whom?" (Clarke, 2018, p. 104).

The maps can be used throughout the entire research process. In early stages of the research process, situational maps can aid the researcher in mapping out all relevant elements and/or inform choices regarding data collection, or to capture relevant aspects or influencing factors regarding decisions with respect to the research design (ibid). Later in the research process, situational maps are used to map the relations of all relevant elements in the research situation of concern – a process Clarke (ibid) refers to as ‘relational mapping’. Social world/arena maps help to specify the relations of these elements and can offer interpretations of the broader situation (ibid). They can furthermore offer new perspectives onto the data and provide a tool to provoke and analyse potential analytical directions in a more abstract manner (ibid). In the later stages of analysis, positional maps may offer additional insights by systematically mapping out positions in the situation along two specific axes of concern to the situation (ibid). Positional maps raise the level of analysis onto the level of related discourses and chart out position regarding relevant discourses vis-a-vis the positions taken or articulated in the data to capture analytic complexities regarding the

research situation of concern more explicitly (ibid). All three maps can inform CGT by providing additional analytic means to make sense of data on a situational level by considering all its key elements and their relations.

How CGT and SA are employed in the proposed study is depicted in figure 5 below.

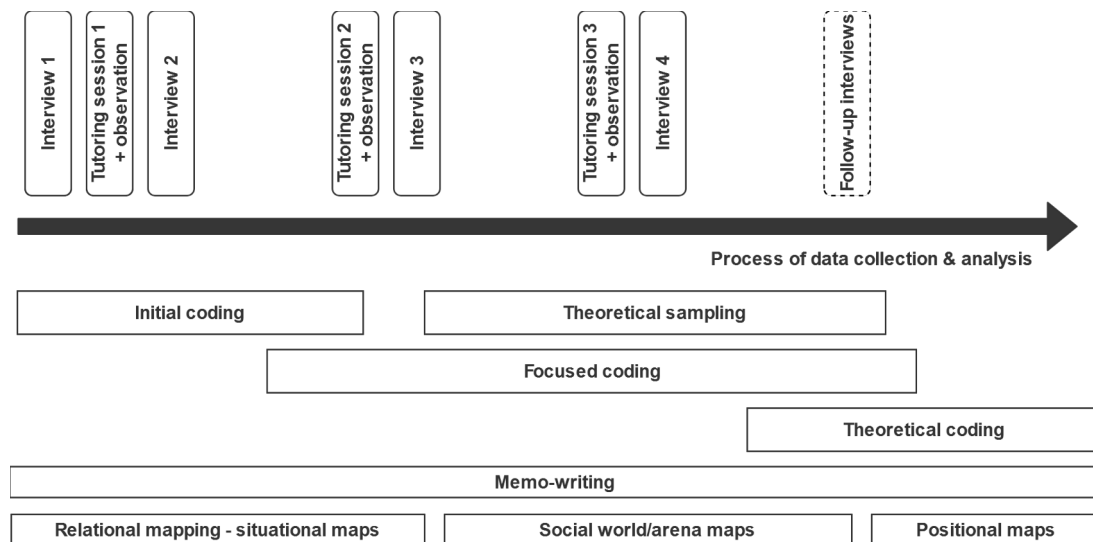


Figure 5: Process of data collection and analysis (simplified)

In line with grounded theory, data collection and analysis are conducted simultaneously in an iterative process (Charmaz, 2008; Thornberg & Charmaz, 2014). Right before the first tutoring session, a short informal conversational interview (Allen, 2017; Roulston, 2008; Spradley, 1979) takes place with the participant (interview 1 in figure X) to enquire about his/her expectations about the tutoring sessions, as well as previous experiences and with and knowledge about ITS or similar systems. The following tutoring session (tutoring session 1) is observed by means of open, exploratory observation based on Spradley's (1980) descriptive observation framework, during which observational notes about the interaction between the participant and the ITS are taken along different dimensions, including space, actors, activities, objects, time, and feelings (ibid). Additionally, the tutoring sessions are recorded (video recording). The ITS-software also automatically generates a session log that records the tutoring session, i.e. tutor actions (as shown in figure X session structure) and the inputs by the participant, i.e. his/her answers to the tutor's questions. The first tutoring session is followed by another informal, unstructured interview (interview 2) immediately after the tutoring session in order to ask the participant about his/her experiences when interacting with the ITS. After a period of one to two weeks – during which the collected data is analysed to inform subsequent interview guides and observation schemes – the

next tutoring session (tutoring session 2) takes place, again, immediately followed by an interview (interview 3). In a similar fashion, data are collected and analysed iteratively, followed by the next session and interview (tutoring session 3 and interview 4, see also theoretical sampling below).

Data analysis will mainly use the interview transcripts and observation notes. However, in case analytically relevant instances are identified in the data, they can be enriched by complementary analysis of the corresponding video recording and/or session log. For example, if a participant reports his/her experiences regarding a specific moment during the tutoring session, e.g. a specific tutor action, this can be investigated in-depth by analysing the corresponding moment in the video recording and/or session log by means of coding and/or mapping and memo-writing. This procedure should allow for gathering 'rich' (Charmaz, 2006) and its 'thick' (Clarke, 2003, 2018) description in terms of data analyses, interpretation, and theory construction.

The first phase of grounded theory coding (initial coding) will comprise mainly the transcripts of interview 1 and 2, as well as the observational notes from the observations of the first tutoring session of all participants. This may be complemented with the session log of the first tutoring session, as well as the video recording of the session. In *initial coding*, the researcher does not apply preconceived codes or categories (e.g. codes based on existing theories, prior studies, literature research, etc.) to the – primarily written/textual – data. Instead, initial coding is meant to closely capture and describe the data at hand by means of open and inductive coding. The resulting codes are constructed in the sense that the researcher actively uses them to name segments of data. During initial coding, the researcher remains open to exploring different theoretical possibility that may be discerned in the data. Initial coding is done rather quickly and spontaneously and sticks closely to the data and data is codes as action, i.e. codes are named in the form of action patterns that are kept short and precise. Questions that can guide this coding phase are. "What process is an issue here and how can it be defined?", "How does the process develop?", "How does the participant act in the process?", or "Why and when does the process change?", "What are the consequences of the process?", and the like (c.f. (Charmaz, 2006; Thornberg & Charmaz, 2014)). Usually, this kind of coding is done line-by-line, i.e. each line – not each sentence – is named by linking a code to it. Keeping the data segment to be coded during initial coding short helps to focus on what is in the data and apply codes that stay close to the data, instead of projecting one's own assumptions and preconceptions onto it. Thus, initial coding may lead to indications of power and related affect in social interaction, but should also reveal other relevant aspects, ac-

tions and processes since the researcher theoretically open. This aspect is crucial for grounded theory coding: “the grounded theory method itself contains correctives that reduce the likelihood that researchers merely superimpose their preconceived notions on the data. Line-by-line coding provides an early corrective of this type” (Charmaz, 2006, p. 51). This kind of coding provides a detailed view into the data that allows to identify explicit statements as well as implicit concerns and their nuances, helps to identify relevant topics, patterns, or themes, and provides a basis for both further data analysis as well as collection (ibid). Initial coding informs the separation of data into codes that can later develop into broader categories to identify – potentially competing or contradictory – conceptions of actions and related social processes in the data (ibid).

In this stage of data analysis, *situational maps* are used to map all relevant elements in the research situation, i.e. the first tutoring session, in order to complement the process of initial coding by offering a situational perspective in relation to the action-based codes. Throughout initial coding – as well as during focused coding (see below) – these maps are updated to include relations of all relevant elements in the situation (relational mapping) as more data is collected and analysed. Additionally, insights gained from these *relational maps* can inform data collection, i.e. theoretical sampling (see below).

Crucially, coding in CGT always involves the *constant comparative method* (Charmaz, 2006; Glaser, 1978); right from the beginning of coding and throughout the entire process of theory construction. During coding, data is compared with data, and later, data is compared to codes, codes are compared to codes, codes to categories, and so on. For example, codes from the line-to-line coding of interview 1 of a participant are compared to the interview 1 when read in its narrative entirety, or codes from interview 1 of one participant are compared to those from another participant, codes from one participant are compared between interview 1 and 2, and so on. This basic comparative approach is maintained throughout the *entire* coding process.

Through coding, the researcher makes sense of the data by comparing and relating it analytically, and further coding leads to the development of theoretical categories during *focused coding*. After establishing analytical directions through initial coding and relational mapping, the most significant or frequent codes from the initial coding phase are used to analyse larger amount of data. Focused codes are more directed, selective, and conceptual than the previously described initial codes. Charmaz (2006) described focused coding as follows:

Focused coding means using the most significant and/or frequent earlier codes to sift through large amounts of data. One goal is to determine the adequacy of those codes. Focused coding requires decisions about which initial codes make the most analytic sense to categorize your data incisively and completely.

But moving to focused coding is not entirely a linear process. Some respondents or events will make explicit what was implicit in earlier statements or events. An 'Aha! Now I understand,' experience may prompt you to study your earlier data afresh. [...]

You act upon your data rather than passively read them. Through your actions, new threads of analysis become apparent. Events, interactions, perspectives come into analytic purity that you had not thought of before. Focused coding checks your preconceptions about the topic [...] Through focused coding, you can move across interviews and observations and compare people's experiences, actions, and interpretations. (Charmaz, 2006, p. 57)

Thus, focused codes are developed and refined throughout the coding phase. Through comparing data to data and comparing data to codes, the focused codes are refined and become richer and clearer. Note that focused coding is not entirely a linear processes. Developing a cohesive theory requires the researcher to engages in *theoretical sampling* (ibid). Once focused codes are developed, the researcher collects additional data to refine the emerging conceptual categories – and, thereby, the emerging grounded theory – to specify their properties, boundaries, causes, consequences, etc. (Charmaz, 2006; Thornberg & Charmaz, 2014). According to Charmaz (2001) “theoretical sampling builds precision, density, and complexity into the emerging theoretical statements and keeps them grounded in data.” (Charmaz, 2001a, p. 6398). Thus, iterative analysis (during initial and focused coding phases) will inform the observational scheme for tutoring session 1 and 2, as well as the interview guides for the semi-structured interviews 3 and 4 and, if necessary, additional follow-up interviews. Through developing focused codes by analysing fresh data, related codes are developed into conceptual categories.

During focused coding, *social world/arena maps* fulfil a similar purpose as situational maps during initial coding, but they capture the relations of involved elements and or actors more explicitly and concretely. In combination with the development of conceptual categories via means of coding, social world/arena maps can help to flesh out the actor's relations and the situational structures they are embedded in. This kind of mapping too can inform data collection during focused coding, as well as theoretical coding (see below). Crucially, CGT coding and SA involve *memo-writing* throughout the entire research process (including memo-writing on maps) and this is of particular relevance when defining and refining conceptual categories:

Through memo-writing, grounded theorists fill out their codes and identify gaps in them. They define the code, delineate and analyze its properties, specify conditions under which it exists and changes, demonstrate its relationship to other codes, and weigh its significance for processes discovered in the field (Charmaz, 2001a, p. 6398).

Memo-writing means to take a step back, reflect, and write down the process of analysis to make maps, codes, categories, and related thoughts and reflections manageable: “It leads the researchers to explore and scrutinize their codes, categories and emerging GT [Grounded Theory]” (Thornberg & Charmaz, 2014, p. 163). Memo-writing allows the researcher to keep track of and document analytical interpretations and ideas, as well as possible theoretical meanings and relations of codes and categories. Memo-writing is pivotal for developing codes and conceptual categories, as well as their relations. This is a prerequisite for the next phase of CGT-coding: *theoretical coding* (Thornberg & Charmaz, 2014). Theoretical coding is a more sophisticated and abstract process than initial and focused coding (ibid). The purpose of theoretical coding is to identify and describe the relations among focused codes and categories in order to further the analysis towards a coherent theory (ibid). In this process, theoretical codes serve as analytical tools to organise and conceptualise the conceptual categories and codes grounded in the data into a coherent grounded theory (Charmaz, 2006; Thornberg & Charmaz, 2014).

theoretical codes specify possible relationships between categories you have developed in your focused coding. [...] Theoretical codes are integrative; they lend form to the focused codes you have collected. These codes may help you tell an analytic story that has coherence. Hence, these codes not only conceptualize how your substantive codes are related, but also move your analytic story in a theoretical direction. [...] If you use them skilfully, theoretical codes may hone your work with a sharp analytic edge. They can add precision and clarity-*as long as they fit your data and substantive analysis* [emphasis added]. They can aid in making your analysis coherent and comprehensible. Depending on the data you have [...] you may clarify the general context and specific conditions in which a particular phenomenon is evident. You may be able to specify the conditions under which it changes and to outline its consequences. You might learn its temporal and structural orderings and discover participants' strategies for dealing with them. (Charmaz, 2006, p. 63).

In the phase of theoretical coding the researcher analyses how the categories and codes constructed from the data may relate to each other as sets of hypotheses that are integrated into a theory (Charmaz, 2006; Thornberg & Charmaz, 2014). When engaging in theoretical coding, researchers also critically consider what Charmaz (2006) refers to as *extant* concepts or theories:

Initial and focused coding generate data-driven and empirical codes and categories by building on constant comparisons of data, data and codes, and codes and codes. In contrast, theoretical codes consist of ideas and perspectives that researchers import to the research process as analytic tools and lenses from out-side, from a range of theories. Theoretical codes refer to underlying logics that could be found in pre-existing theories. (Thornberg & Charmaz, 2014, p. 159).

Theoretical coding requires knowledge about theories from different disciplines in order to critically investigate and employ them as potential perspectives or analytical lenses to refine and con-

textualise the emergent grounded theory²⁴. Theoretical codes could originate from theories on (human) social interaction and social cognition (e.g. theories reviewed in 2.1, 3.5-3.11), and particularly theories on power and related affect and its role in social interaction stemming from the social sciences and cognitive sciences. The emerging grounded theory has to be critically and systematically analysed and discussed with respect to existing extant theories about social interaction and cognition – if applicable. However, the researcher must not project his/her preconceptions or favoured theoretical constructs onto the data without critical analysis through constant comparison. Such extant theories must ‘earn their way’ into the grounded theory:

adopting and applying theoretical codes poses similar risks of preconceiving the analysis [...] theoretical codes must *earn their way* [emphasis added] into the analysis by the work of careful and constant comparisons between theoretical codes, data, empirically generated codes and categories, and memos [...] Theoretical codes must work, have relevance, and fit the data and generated and refined categories (Thornberg & Charmaz, 2014, pp. 160–161).

In this stage of analysis, *positional maps* may be used in several ways. Firstly, to identify the positions represented in the emergent grounded theory based on the conceptual categories derived from focused coding. While focused coding is well-suited to capture the relevant actions and social processes that lay the basis for theoretical coding, positional maps can inform theoretical coding on a more abstract level that is “disarticulated” (Clarke et al., 2015, p. 14) from the concrete site of research. For example, positional map could contrast position taken or not taken by the participants, e.g. when mapping along the axes of negative or positive affect (from low to high) and the treatment of the tutor as an actor or subject. Positions of power can be mapped in a similar fashion – as can any other relevant aspects identified during CGT-coding. Applied this way, the positional map would still be strongly related to the micro-level interaction that are of interest with regard to the phenomenon of social interaction with ITS. On the more abstract level of discourse with respect to the research situation, such positional maps may provide additional insights into discourses that frame the observed situation and its structure, e.g. topics like the use of ITS or intelligent agents for purposes of learning or assistance more generally. As such, positional maps may inform decisions on which areas of research, disciplines, and/or concrete extant theories to consider during theoretical coding. Additionally, positional maps may be modified in order to chart the assumptions and hypotheses of the emergent grounded theory against extant

24 Note that in this sense CGT involves abduction as well as induction and allows the researcher to explicitly employ his/her prior knowledge in a reflective and systematic manner: “constructivist grounded theorists admit and use the analytical power of the constant interplay between induction (in which they are never tabula rasa) and abduction [...] In contrast to Glaserian GT (Glaser, 1978; 1998) which argues for delaying the literature review in the substantive area of the actual GT study until the analysis is nearly completed, constructivist grounded theorists [...] take advantage of knowing and using the literature, not for forcing the research into preconceived categories but as multiple possible lenses. [...] The ability to draw good abductive inferences is dependent on the researchers’ previous knowledge, rejection of dogmatic beliefs and development of open-mindedness” (Thornberg & Charmaz, 2014, pp. 162–163)

theories as positions in a general theoretical discourse on the research topic, and thereby aid the researcher in keeping a critical stance towards the extant theories and their premises, as well as integrating them.

Note that despite the way the research process has been illustrated in figure 5, neither CGT-coding nor SA-mapping is a linear process: "coding is not a linear process, but in order to be sensitive to theoretical possibilities, researchers move back and forth between the different phases of coding." (Thornberg & Charmaz, 2014, p. 156). Different phases of coding may overlap or merge into each other. The act of changing from a mode of initial coding to focused or theoretical coding is a fluent process. Depending on the richness of data collected, the researcher may even be prompted to return to earlier phases of coding and/or to collect new data and start out coding afresh. Data might have to be analysed by means of focused coding, then theoretical coding, and then back to focused coding in order to flesh out conceptual categories more concretely, and the like.

Theoretical playfulness allows us to try out ideas and to see where they may lead. Coding gives us a preliminary set of ideas that we can explore and examine analytically by writing about them. Grounded theory coding is flexible; if we wish, we can return to the data and make a fresh coding. We can go forward to writing about our codes and weighing their significance. (Charmaz, 2006, p. 71)

Similarly, mappings are always provisional; they are updated, changed, and revised during the research process as more data is being analysed (Clarke, 2015, 2018), e.g. whenever a new tutoring session takes place and data on it are collected. All kinds of maps are meant to be used as analytical tools to further the process of theory generation and represent a certain point in time during the research process, not a final result or outcome.

Concerning the concrete technical and practical aspects of employing CGT and SA as described in the present chapter, the iterative data collection and analysis is to be done with ATLAS.ti²⁵. The

25 Note that although some authors (Friese, 2016, 2019a) point out that ATLAS.ti can indeed be used for conducting GT, the use of software for purposes of GT analysis is a disputed topic (see for example (Friese, 2019a; Gorra, 2019; Woods et al., 2016). I, i.e. the author of the present master's thesis, have however used ATLAS.ti for GT in the past, e.g. to investigate social and epistemological processes during collective knowledge-creation among university students (Peschl et al., 2014), as well as to capture and analyse affect in terms of feelings of uncertainty when engaging in knowledge-creation and innovation projects (Hartner-Tiefenthaler et al., 2018) and to investigate differences regarding those processes in online versus offline settings in the context of COVID-19 (Bottaro et al., 2021). For purposes of transparency, I would also like to state that I am working as a software trainer for different qualitative and quantitative data analysis tools/software packages, including ATLAS.ti, and have also held workshops/courses on how to use ATLAS.ti for CGT. Although potentially being biased, I do claim that the use of software like ATLAS.ti is not only legitimate, but also necessary to ensure high quality analyses that is in accordance with scientific standards of qualitative methods and research practice. Additionally, the use of software makes the entire data collection and analysis process – which, due to the complexity of any GT, may not always be easily comprehensible or reproducible – transparent and accessible to others, i.e. the research community.

Qualitative Data Analysis Software (QDAS) ATLAS.ti supports systematic analysis through coding and diagramming in form of networks, which can visualise codes, data segments, comments, memos, as well as their relations (Friese, 2016, 2019b). The software supports the researcher in managing, organising, and coding different media and data types, e.g. video and audio recordings, interview transcripts, observational notes, images and photographs (ibid). This QDAS is well-suited for handling larger amounts of qualitative data and supports the analysis of different media through coding, as well as linking and relating different media/data, specifying their relations, and writing memos and/or comments on them. ATLAS.ti also supports the process of memo-writing and (initial) coding particularly well, which is crucial for employing CGT and SA. Furthermore, different types of media/data can be linked and synchronised, e.g. the video recording of a tutoring session could be synchronised with the corresponding session log and/or a video transcripts to link and analyse particularly interesting moments across the different media/data types. Similarly, a particular moment in the session log could be linked to a passage in an interview transcript or of the observational notes that makes reference to this instance, which aids the systematic and in-depth coding, retrieval, and interpretation of the data on particular instances across different media/data types and levels of abstraction. In this sense, the software's functions can support the researcher in employing the principle of 'constant comparison'. The software does seem to lend itself very well to the process of sorting and organising 'rich' data (Charmaz, 2006) and analysing and interpreting those data in the sense of a 'thick' analysis process (Clarke, 2015, 2018).

Part III Bridging Part I and II: Situated Cognition Perspectives on the Social Interaction with ITS

5 Discussion across Part I and II

5.1 *Human Learner-ITS Interaction as Social Cognition: Social Actors, Cognitive Agents, or Socio-Cognitive System?*

In the following, RQ 3 will be discussed:

- RQ3: Which of the conceptions from RQ1 are relevant for and can be fruitfully applied to the case of social interaction with ITS?

Within the context of the present master's thesis, 'fruitfully' means that a concept may lack truth value (because it is not the truth value that is being assessed), but is, in reference to Merton (1968) and Sohlberg and Leiulfstrud (2018), generating new insights or research questions that may inform future theoretical models and/or empirical research. Concepts that make the empirical analysis difficult are less fruitful, while concepts that aid the empirical analysis or solve problems encountered during the empirical analysis are considered to be more fruitful (Hagen, 2017). By comparing perspectives (i.e. specific concepts or theories from the perspective of a situated cognition paradigm) on the phenomenon of interest (i.e. the social interaction between a human learner and an ITS), insights into their *potential* empirical applications may be revealed (Sohlberg & Leiulfstrud, 2018). Whether a conception is 'correct' or 'true' is thus not the primary concern of the conceptual analysis in the following chapters; conceptions are assessed in terms of fruitfulness.

For this purpose, a 'conceptual analysis road map' (see figure 6, in the following also referred to as 'analysis road map') has been developed. This road map serves as a kind of blueprint for the conceptual investigation and discussion of concepts with regard to their (fruitful) applicability to the context of social interaction between a human learner and an ITS. I.e. this analysis road map guides the conceptual discussion.

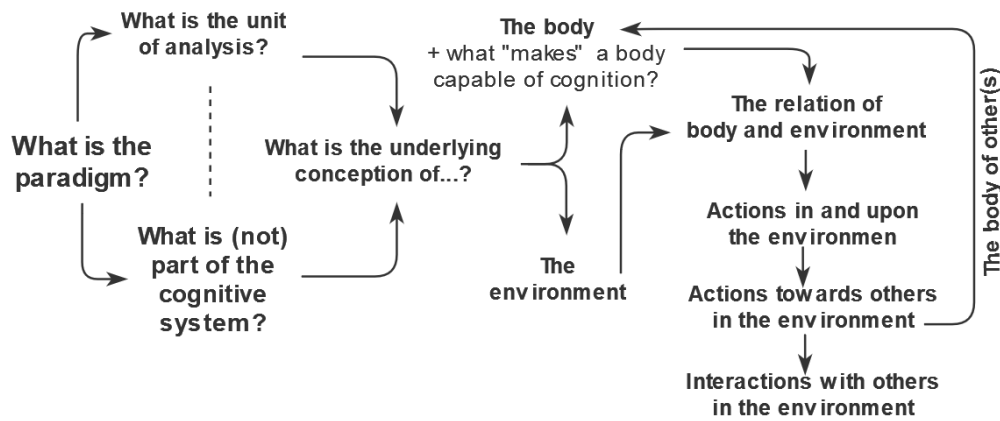


Figure 6: Conceptual Analysis Road Map

Specifically, this analysis road map (see figure 6) is used to discuss – on a conceptual level – whether a given concept can be fruitfully applied to the context of human-ITS interaction in the sense of a social interaction. Although the focus is on the interaction, this interaction does involve two entities – the human learner and the ITS. These two entities may not partake in the interaction to the same degree. Note that the human learner will primarily be discussed in the questions of the analysis road map, e.g. about the body, while the ITS-agent will (if adequate) explicitly be considered as ‘the other’, because this appears to be aligned with most of the theories/concepts described in part I of this master’s thesis.

5.2 Social Interaction with ITS from an Embodied Cognition Perspective

As stated in chapter 3.6, conceptions in which

- cognition is *co-dependent upon extra-cranial/bodily processes* are associated with *weak embodiment*.
- cognition is *co-constituted by extra-cranial/bodily processes* capture the idea that cognition is *strongly embodied*.

(Newen et al., 2018; S. Walter, 2014b)

Thus, for the purpose of conceptual considerations and reasoning within the context of this master’s thesis, one may postulate that:

A conception in which

- cognition occurs in a social environment in which cognition is *co-dependent upon social extra-cranial/bodily processes* may be interpreted as being *weakly socially embodied* in the sense of social embodiment effects (as in (Barsalou et al., 2003)).

- cognition occurs in a social environment in which cognition is *co-constituted by social extra-cranial/bodily processes* may be interpreted as being *strongly socially embodied* in the sense of bodily actions and (dynamic) functions (as in (Lindblom, 2015b, 2015a)).

5.2.1 Social Embodiment Effects

Summary:

Barsalou et al. (2003) define embodiment as “states of the body, such as postures, arm movements, and facial expressions, arise during social interaction and play central roles in social information processing” (Barsalou et al., 2003a, p. 43) and state that “embodied theories represent knowledge as partial simulations of sensory, motor, and introspective states” (Barsalou et al., 2003, p. 44).

Barsalou et al. (2003) propose four ‘social embodiment effects’:

1. Perceived social stimuli don’t only produce cognitive states but also bodily states, e.g. when adopting a more upright posture when receiving a compliment.
2. Observing bodily states in others causes mimicry of bodily states in the observer, e.g. when automatically mimicking someone’s facial expressions or body posture during conversation.
3. Bodily states in the self also produce affective states in the self, e.g. experiencing more pride in an upright position than in a slumped one.
4. Performance in social interaction is modulated by the compatibility of bodily and cognitive states, e.g. nodding when agreeing with someone in relation to a positive preference towards someone or something.

What is the paradigm?

The theory/concept under consideration is associated with embodied cognition, specifically, social embodiment.

What is the unit of analysis; what is (not) part of the cognitive system?

Cognition is body-based and involves cognitive and bodily states. This social cognition conception, i.e. in this sense of social embodiment, essentially means to perceive social stimuli, which produce cognitive and bodily states. Even if social stimuli are perceived by means of observing another human (e.g. during a conversation), this other human is involved in, but not necessarily a part of the cognitive system. The unit of analysis is the individual and its cognitive system.

What is the underlying conception of the body and what ‘makes’ a body capable of cognition?

The concept of social embodiment effects is concerned with human cognition and human-human interaction (human-centredness). In this sense, the body is a human or a highly human-like body, i.e. possessing bodily capacities that are similar to humans. However, this is not specified within the work of Barsalou et al. (2003). Ziemke (2003) points out that social embodiment could be viewed as “orthogonal” (Ziemke, 2003, p. 1309) to other notions of embodiment, but that it seems compatible with the notions of organismoid embodiment and body-based offline cognition. It may also be compared to embodiment as structural coupling if the environment is a social environment. Thus, an artificial body that has highly human-like capabilities can also be considered capable of cognition in the sense of the notions of embodiment above.

Concerning the body of ‘the other’, see questions regarding the environment and actions towards others in the environment.

What is the underlying conception of the environment?

Essentially, the (physical) environment is a ‘source’ of information to be perceived and processed. This environment also includes social stimuli, which produce cognitive and bodily responses in the observer. The body of ‘the other’ is a part of this environment, i.e. it is an extra-bodily/environmental ‘source’ of social stimuli (potentially ‘richer’ or more likely to produce social stimuli than non-human artefacts or objects).

What is the underlying conception of the relation between body and environment?

Social embodiment effects seems to imply a (co-)dependency relation. The body is the means by which the environment is perceived and to produce actions upon the environment. It seems that the body is conceptualised as a kind of vehicle that ‘mediates’ the internal and external, particularly with regard to social stimuli and social bodily actions. The notion of the body in relation to the environment would probably best be placed in-between what Kiverstein (2012) calls ‘body-functionalism’ (or Gallagher’s (2011) notion of embodied functionalism) and Kiverstein’s (ibid) ‘body-conservatism’. That means, the conception depicts the body as implementing the computational processes that underpin cognition; the body links external (social) information/stimuli to the internal “biological machinery” (Kiverstein, 2012, p. 740), but also admits the relevance of the body for social cognition and views the body as a means to contribute to information-pro-

cessing by ‘supplying’ inputs to the brain, as well as executing motor-instructions ‘sent out’ by the brain.

What is the underlying conception concerning actions in and upon the environment?

The underlying conception of action in and upon the environment appears to be leaning towards ‘body-conservatism’. The body contributes to the processing of (social) information; actions are outcomes of the processing of social stimuli and resulting motor instruction that produce motor functions (meaning bodily actions, e.g. facial expression, nodding).

What is the underlying conception concerning actions towards others in the environment?

Conceptually, there does not seem to be much difference between actions towards the environment and actions towards others in this environment. They are the result of internal processing of social stimuli which involves cognitive and bodily aspects. However, ‘others’ may be a ‘richer’ or more likely source of social stimuli than inanimate objects and may result in mimicry. Bodily aspects of actions towards others (eye gaze, body posture, gestures, and the like) are, in this sense, outcomes of ‘inner’ mental processes.

What is the underlying conception concerning interactions with others in the environment?

The concept of social embodiment effects suggest a strong relationship between cognitive and bodily, as well as affective, states in social situations, especially when perceiving others in direct (social) interaction. Performance in social interaction is modulated by the compatibility of bodily and cognitive states, e.g. nodding when agreeing with someone in relation to a positive preference towards someone or something (e.g. positive feedback). Such actions may be voluntary or involuntary (e.g. automatic facial mimicry, mimicry of body posture).

Considering that the concept of social embodiment effects seems to be based on an understanding of cognition as information-processing, ‘the other’ appears to be an extra-cranial/bodily ‘source’ of social information, which is perceived and processed by the individual cogniser /cognitive system during the course of direct interaction (methodological individualism). Interacting socially with others, thus, means that (at least) two individual cognitive systems are involved in co-dependent processes of perceiving, processing, and producing social stimuli in the sense of bodily actions or behaviours. Agents relate to each other through the use of explicit internal representations (and/or simulations).

Having conceptually investigated the theory/concept on the basis of the analysis road map, the issue of fruitfulness can be addressed.

In summary: What, if any, aspects of interest with regard to the phenomenon of social human-ITS interaction are focused on and/or can be adequately captured when applying this theory/concept as a framework?

The focus appears to be on the individual cogniser/cognitive agent (methodological individualism) and its perception of social stimuli (possibly produced by other cognisers/cognitive agents), which include facial expressions, gestures, or bodily movement and composure (e.g. posture) during social interactions, as well as the production of such socially embodied actions or behaviours (i.e. facial expression, gestures, etc.). On a conceptual level, the framework adequately captures social bodily actions and behaviours of the human learner. Of interest are the learner's social bodily action and behaviours towards or in response to the ITS, assuming that stimuli coming from the ITS-agent (e.g. language, animated facial expression and gestures) can be considered to be social stimuli (as-if mode).

Finally: In what manner, if any, can this theory/concept be fruitfully applied to the case of social interaction between human and ITS?

The focus would be on the human learner and the ITS would likely be considered a source of (social) information or stimuli located in the environment. The concepts of social embodiment effects could be employed to investigate the human's social actions or responses, possibly including mimicry and imitation of body postures or facial expression of ITS-agent's expressions and gestures. When the learner responds socially (in the sense of social embodiment effects) when receiving positive or negative feedback from the ITS-agent, this might indicate that he/she perceives the ITS-agent as 'social' (in the sense of an as-if mode).

5.2.2 Socially Embodied Functions and Actions

Summary:

The approach to embodied social cognition proposed by Lindblom (Lindblom, 2015b, 2015b, 2020) is based on a shift in perspective from an information-transmission metaphor to a dance metaphor. That is, (social) cognition is not viewed as internal (social) information that is externally transmitted back and forth between agents, but instead focusses on the emergence of inform-

ation ('meaning') in dyadic interactions. Embodied actions are no linear or discrete 'mapping' processes, but dynamically emerging in a particular context of interactive social situations.

Lindblom (2015a) generalise the role of the *human* body and its sensorimotor processes in social contexts to four fundamental functions:

- The body functions as social resonance mechanism
 - The body functions as a means and end in social interaction
 - Embodied actions and experiences function as a 'helping hand' in shaping, expressing and sharing thoughts
 - The body functions as a representational device
- (Lindblom, 2015a, p. 149).

Focussing on embodied actions in social interaction and cognition, the most relevant aspects – as specified by Lindblom (2015a) – are that:

- Embodied actions are viewed as a general cross-modal interaction system rather than being considered separate parts of facial expressions, gesture, bodily positions and so on.
 - Embodied actions are a set of movements that have meaning for the actual person or agent, suggesting a move from the experienced body to the enacted body.
 - Embodied actions are always in relation to others.
 - The nature of social interaction is relational and meaning and intentions are emergent phenomena of these social interactions.
 - Embodied actions are situated in cultural and material contexts.
 - Interacting socially through embodied actions is a dynamical process that unfolds with a temporal horizon.
 - Socially embodied action is not pre-given, since the social mind develops through the social body in motion.
 - The unit of analysis is situated embodied actions in co-regulated social interaction.
 - The main focus of analysis is embodied actions in social interaction and cognition
- (Lindblom, 2015a, pp. 155–156)

What is the paradigm?

The theory/concept under consideration is associated with embodied social cognition, specifically, social embodiment.

What is the unit of analysis; what is (not) part of the cognitive system?

The unit of analysis comprises visible – and thus observable – embodied actions and the (co-regulated) interactions between agents and/or their environment. Embodied actions are always in relation to others, and those must be considered with respect to the cognitive system. Hence, the study of embodied actions is not limited to the interaction between agents, but also includes their environment. Social cognition involves the entire body, particularly gestures. But examining embodied actions apart from "their context of use" (Lindblom, 2015a, p. 255) is not considered to be meaningful (ibid).

What is the underlying conception of the body and what ‘makes’ a body capable of cognition?

Cognition is body-based and (social) cognition is dependent on the body (involving the mirror neuron system, simulations, or direct and/or online perception). Particular gestures may also have a constitutive role because they can ‘shape’ and ‘share’ thinking.

Firstly, the body in this perspective on embodied social condition is a *human* body (human-centredness). Secondly, this suggests a ‘living’ body in the sense of enactivism. Thirdly, this seems to be compatible with notions of embodiment that rely on strong interpretations of structural coupling, e.g. historic embodiment (the self and the mind/cognition develops through interaction with the social world), and organismoid embodiment, if the organism is highly human-like or has human-like capacities. With respect to non-humans within this conceptualisation of social interaction and cognition, employs the method of natural inquiry to investigate embodied action in the interaction between two humans and a horse (Lindblom, 2015a). However, she explicitly addresses interactive technology and artificial agents, and does (currently) not seem to consider them as being capable of social cognition in her sense. The ‘social’ would be in the eye of the observer (Lindblom, 2015a; Lindblom & Ziemke, 2003), meaning the human interaction partner (as-if mode). The conception of social embodiment under consideration may also – more restrictively – be interpreted as organismic, i.e. excluding artificial agents.

What is the underlying conception of the environment?

The physical and social environment (including others, but also e.g. cultural practices) provide the context for social interaction and cognition, suggesting that it is involved in cognition as (at least) a contextual factor.

What is the underlying conception of the relation between body and environment?

The relation appears to be a co-dependency of brain and body (including the brain) and the environment. Cognition is body-based and relies on actions, which includes the social environment as contextual factor with regard to the body during social interaction and cognition. This is because a particular embodied action is only ‘meaningful’ with regard to a) others, and b) the social environment and cultural context, practices, and the like.

What is the underlying conception concerning actions in and upon the environment?

Actions are cognition. Embodied actions towards rely on the environment with reference to their

‘meaning’, which is relational to the particular social, cultural, and material context. However, there is a difference in the dynamics between actions towards the physical environment and the social environment, i.e. actions towards or interactions with other agents in terms of complexity (see also next question).

What is the underlying conception concerning actions towards others in the environment?

Simply put, actions are a form of cognition, and actions towards others are a form of social cognition. However, other agents are more complex and diverse than the physical environment. Actions towards others are thus more complex and more heavily rely on the expression, sharing, or emergence of meaning. Embodied actions towards others may involve simulation and the mirror neuron system, but also pragmatic practices and direct perception of the other.

What is the underlying conception concerning interactions with others in the environment?

The interaction with others – through embodied actions – involves ‘pragmatic embodied practices’ with regard to the expression, sharing, and/or emergence of meaning in those social interactions. Cognitive processes of social (cognitive) interaction are grounded in embodied representations that have representational content. Here, ‘representational’ seems to refer to meaning or ‘significance’ for the agents. Embodied actions are representational similarly to the representational qualities of language. A certain word refers to or represent (in its meaning) an ‘actual’ object that may be significant to the agent in a specific way. The gesture of opening a jar is not ‘just’ bodily movement, but also has a symbolic function because it stands for the *actual* opening of a jar (cf. (Lindblom, 2015a)).

Having conceptually investigated the theory/concept on the basis of the analysis road map, the issue of fruitfulness can be addressed.

In summary: What, if any, aspects of interest with regard to the phenomenon of social human-ITS interaction are focused on and/or can be adequately captured when applying this theory/concept as a framework?

The focus would be on the human’s embodied action, particularly on gestures, during his/her interaction with the ITS, assuming that the tutoring session can be assumed to provide a social environment or setting. The virtually/animated embodied actions of the ITS-agent (facial expressions, head movements, gestures) may also be captured in the sense of an ‘as-if’ mode.

Finally: In what manner, if any, can this theory/concept be fruitfully applied to the case of social interaction between human and ITS?

One important aspect with regard to fruitfulness is that this approach has already been used empirically and that theoretical and empirical insights have been (partially) synthesised (e.g. in (Lindblom, 2015b, 2015a)). As a framework, socially embodied actions and the functions of the body in social interaction and cognition are fruitfully applicable to the human learner and his actions towards or responses to the ITS-agent. However, is not the case with regard to the ITS-agent. Nevertheless, this is also why it may yield novel insights or research questions with regard to *minimal* approaches of socially embodied actions and/or gestures.

5.3 Social Interaction with ITS from an Embedded Cognition Perspective

As stated in chapter 3.6, conceptions in which

- cognition is *co-dependent upon extra-bodily processes* (i.e. the natural, social, and/or technological environment) are generally referred to as *embedded cognition* (Newen et al., 2018; S. Walter, 2014b).

Thus, for the purpose of conceptual considerations and reasoning within the context of this master's thesis, one may postulate that:

A conception in which

- cognition occurs in a social environment in which cognition is co-dependent on social extra-bodily processes may be interpreted in the sense of social scaffolding or social affordances.

5.3.1 Social Scaffolding

Summary:

Scaffolding generally refers to a specific kind of (actively) using resources in the environment, i.e. a metaphorical 'scaffold' that is used to reduce the workload of information processing when solving cognitive problems (Clark & Chalmers, 1998a, 2010a; S. Walter, 2014b). The concept of scaffolding is based on the idea that the environment of a cognitive system is of a specific nature; a different type of environment therefore leads to a different possibility of scaffolding in the sense

of exploiting an external structure. If the external structures are social structures (a social environment), then social interactions and social and/or pragmatic practices (e.g. specific gestures) can *socially scaffold* cognition (S. Gallagher, 2017; Lindblom & Ziemke, 2005; D. Ward & Stapleton, 2012).

What is the paradigm?

The theory/concept under consideration is associated with embedded cognition, specifically, social variants of embedded cognition (socially embedded cognition).

What is the unit of analysis; what is (not) part of the cognitive system?

The unit of analysis – and the cognitive system – comprise the (human) agent, including his/her brain and body, the environment, and others in this environment. Specifically: the agent and the particular object, artefact, or action and behaviour of ‘the other’ that provides the social scaffold and/or is exploited as a scaffold.

What is the underlying conception of the body and what ‘makes’ a body capable of cognition?

The conceptualisation of the (human) body seems comparable with embodiment as structural coupling and physical embodiment. Seeing that conceptions of scaffolding may include functionalist/computational, as well as dynamic interpretations, the underlying concept of the body may be conceptualised in line with either body-functionalism/embodied functionalism or (if the body is given a prominent role of relevance) body-conservatism, and as body-enactivism, respectively.

What is the underlying conception of the environment?

The environment is a resource to be exploited; environmental structures can be actively created, manipulated, and used to aid cognitive processes and/or the environment (including others in the environment) can ‘provide’ such structures.

What is the underlying conception of the relation between body and environment?

The relation between the body and the environment appears to be a kind of co-dependency that may be based on notions of (weak) embodiment containing functionalist/computational conceptions, or be based on strongly embodied notions that imply dynamic and/or enactive conceptualisations. In the latter case, some aspects in the body-environment-interactions of either one or two agents as a ‘cognitive unit’ may also be of a co-constitutional nature.

What is the underlying conception concerning actions in and upon the environment?

Action are a form of embodied cognition. Depending on the underlying perspective (functionalist and/or computational vs. dynamic and or enactive) action in and upon the environment may be leaning towards body-conservatism, or towards a dynamic conception of embodied actions. In the first case, the body contributes to the processing of (social) information, i.e. environmental structures, and actions are outcomes of the processing of (social) information and resulting motor instruction that produce motor functions. In the second case, actions are better construed as a kind of embodied cognition and/or perception in which environmental structures (including social structures) occur in the context of dynamic body-environment interactions.

What is the underlying conception concerning actions towards others in the environment?

In short, actions upon the environment, including others, are a means to reduce the agent's cognitive load through environmental exploitation. (See also previous question)

What is the underlying conception concerning interactions with others in the environment ?

Depending on conception of the body, functionalistic/computational as well as dynamic interpretations may be applied to the concept of (social) scaffolding, with regard to the relation to the external (social) resource that is either used as or provides the scaffolding structure (which may be another agent). The underlying conceptions of interactions with others may be based on either an individualistic (methodological individualism) perspective, or a more relational or bi-directional conception that focuses on social scaffolds as part of the interaction, rather than on the individual.

Having conceptually investigated the theory/concept on the basis of the analysis road map, the issue of fruitfulness can be addressed.

In summary: What, if any, aspects of interest with regard to the phenomenon of social human-ITS interaction are focused on and/or can be adequately captured when applying this theory/concept as a framework?

If the ITS-agent's actions (including gestures, facial expressions, language) are conceptualised as a form of social scaffolding, then this theory may adequately capture interactive situations during tutoring, e.g. situations like the 'question-and-repetition' format described in part 2 (chapter 4.3.2.2). This would not relate to the content-related aspects of e.g. hints, prompts, or assistance by the tutor when collaboratively formulating a correct answer with the student, but rather pertain

to *how* this content is delivered. Theoretically, this may include gestures, facial expression, head movements like nodding, or supportive and affirmative utterances in the sense of supporting the learner through socially scaffolding not only his/her learning process but and his/her ‘motivation’ to find a correct answer and to continue the tutoring session.

Finally: In what manner, if any, can this theory/concept be fruitfully applied to the case of social interaction between human and ITS?

The concept of social scaffolding may inform research regarding clarification and specifications of (minimal) social scaffolding with regard to ITS, as well as its operationalisation for empirical investigations of human-ITS interaction. It may also fruitfully inform guidelines or design principles regarding the development of ITS and the implementation of specific tutor actions and moves as a form of social scaffolding.

5.3.2 Social Affordances

Summary:

Essentially, affordances are opportunities for action provided or offered by the environment (Gibson, 1986b; Rietveld & Kiverstein, 2014). As such, affordances are a perspective on the environmental dependency of cognition which is closely related to the idea that ‘perception is for action’ and Gibson’s (1986a) non-representational stance in his ecological theory of perception.

Social affordances are a subcategory of affordances that provide opportunities for social interaction (Baggs, 2021; Borghi, 2018; Rietveld, 2012; Rietveld & Kiverstein, 2014). Through a social affordance, an agent is drawn to a certain possibility of social actions and/or interactions. Social embodied actions (gestures or facial expressions), as well as social situations can carry a social affordance. But, although social affordances can be distinguished from object affordances, they should not be considered as being independent from each other because “object affordances make up an important part of the context of social affordances, and vice versa” (Rietveld, 2012, p. 208).

What is the paradigm?

The theory/concept under consideration is associated with embedded cognition, specifically, social variants of embedded cognition (socially embedded cognition).

What is the unit of analysis; what is (not) part of the cognitive system?

The unit of analysis – and the cognitive system – comprise the (human) agent, including his/her

brain and body, the environment, and others in this environment. Specifically: the agent and the particular object, artefact, or action and behaviour of ‘the other’ that the agents is ‘responsive’ or drawn to in the sense of affordances.

What is the underlying conception of the body and what ‘makes’ a body capable of cognition?

The conception of the body seems to be in line with the notions of body-enactivism, physical embodiment, historic embodiment, and organsimoid embodiment.

What is the underlying conception of the environment?

The (social) environment ‘offers’ or ‘provides’ opportunities for (social) actions and interactions. It is part of the cognitive system and the environment is (at least in part) perceived *directly* (in the sense of (Noë, 2005; O’Regan & Noë, 2001)).

What is the underlying conception of the relation between body and environment?

The relationship appears to be a kind of (strong) co-dependency. The environment is ‘accessed’ and actively perceived through bodily action (which is a form of perception) in the sense of direct perceptions and experiences.

What is the underlying conception concerning actions in and upon the environment?

Action is a form of perception. Through directly perceiving the environment, certain (social) affordances are perceived and acted upon in the sense of being drawn to a certain action over another.

What is the underlying conception concerning actions towards others in the environment?

Actions towards others too seem to be a form of perception. Through directly perceiving the others in the environment, certain (social) affordances are perceived and acted upon in the sense of being drawn to a certain opportunity for a social action over another.

What is the underlying conception concerning interactions with others in the environment?

It seems that the conceptualisation is rather similar to the one of actions towards others. During the interactions with others, embodied actions act as a social affordance which is directly perceived through the actions towards the other.

Having conceptually investigated the theory/concept on the basis of the analysis road map, the issue of fruitfulness can be addressed.

What, if any, aspects of interest with regard to the phenomenon of social human-ITS interaction are focused on and/or can be adequately captured when applying this theory/concept as a framework?

The concept of social affordances may be applied to the ITS-agent in terms its visually/animated embodied actions during tutoring. I.e. Gestures or facial expressions – if they are considered to be social (probably in an as-if mode) – may be interpreted as social affordances that provide specific opportunities and/or constraints for actions to the learner.

Finally: In what manner, if any, can this theory/concept be fruitfully applied to the case of social interaction between human and ITS?

The concept of social affordances can inform research regarding clarifications and specifications of (minimal) social affordances with regard to ITS, as well as its operationalisation for the empirical investigation of human-ITS interaction. It may also fruitfully inform guidelines or design principles regarding the development of ITS and the implementation of specific tutor actions and moves based on social (and object and affective) affordances.

5.4 Social Interaction with ITS from an Extended Cognition Perspective

As stated in chapter 3.6, conceptions in which

- cognition is *co-constituted by extra-bodily processes* (i.e. the natural, social, and/or technological environment) are associated with *extended cognition* (Newen et al., 2018; S. Walter, 2014b).

Thus, for the purpose of conceptual considerations and reasoning within the context of this master's thesis, one may postulate that:

A conception in which

- cognition occurs in a social environment in which cognition is *co-constituted by social extra-bodily processes* (through parity, integration, incorporation, or emergence) may be interpreted as socially extended cognitive processes.

5.4.1 Socially Extended Cognition

Broadly conceived, cognition that extends not only into the physical, but also the social environment, is a kind of social cognition. It is an ‘active’ form of externalism. Possible principles of cognitive extension are 1) functional parity/parity principle (Clark & Chalmers, 1998a, 2010a) 2) *integration* of external /extra-bodily resources based on the *complementarity principle* (Krueger, 2014a; Menary, 2010), and 3) extension via *incorporation* through *phenomenological transparency* (Colombetti, 2014, 2018a), including *mutual* incorporation (Fuchs & de Jaegher, 2009).

If a cognitive agent is linked with an external (i.e. extra-bodily) resource in a two-way interaction, then this interaction creates a kind of causal coupling. Temporarily, all relevant components in the system have a causal role and jointly organise cognition. If the external components were removed, the systems competence would decrease. With reference to the paradigmatic example of Otto and his notebook when trying to locate the museum (see chapter 3.9), Otto’s cognitive system consists of the causal ‘Otto-notebook-coupling’. If someone were to take away his notebook, Otto could not, or only poorly, perform the task of retrieving the museum’s location. This extension is based on “*functional isomorphism* [emphasis added] between internal and external cognitive resources” (Krueger, 2014a, p. 537). I.e. with respect to the aspects relevant to the given task, internal and external resources have to function in a similar way, fulfil the same purpose, and yield the same results as their internal ‘counterpart’ (Clark & Chalmers, 1998a, 2010a; Krueger, 2014a; Menary, 2010). Returning to the Otto-notebook-example: the way Otto uses his notebook is functionally isomorph to the way someone else uses his/her biological/internal memory. This is a cognitive extension via *functional parity*.

In comparison to the parity principle, cognitive *integration* has a stronger focus on body-based cognition, i.e. embodied, aspects of extended cognition (Krueger, 2014a; Menary, 2010; Varga, 2016). External entities (e.g. tools) become integrated into one cognitive system. That is, through bodily *manipulation* (manipulation thesis), cognitive neural processes and external, environmental resources become integrated into one cognitive system. Conceptually, what determines parts of the cognitive system in integrationist approaches is *complementarity* (ibid). That is, the way internal and external resources, each exhibiting different formats and functions, harmoniously integrate and in so doing collectively generate new cognitive capacities. In the case of the Otto-notebook-system or Otto-and-his-partner-system, it is precisely because Otto’s notebook or his partner’s memory is (currently) *not* subject to bio-effects like decay or Alzheimer’s that make it a reliable memory aid, i.e. it is complementary to Otto’s memory (ibid).

External objects or artefacts are *incorporated* into a cognitive system if they become *phenomeno-*

logically transparent (Colombetti, 2014, 2018a). In the special case of social environments, social extension may occur in the form of mutual incorporation, which too is based on phenomenological transparency. In a social interaction, an agent may not only incorporate a certain object or artefact, but also others (e.g. their movements and position in space) and may fluently switch from incorporating the body of the other to one's own embodiment and back again (cf. (Fuchs & de Jaegher, 2009)). Mutual incorporation as a “reciprocal interaction of two agents in which each lived body reaches out to embody the other” (Fuchs & de Jaegher, 2009, p. 474).

What is the paradigm?

The theory/concept under consideration is associated with extended cognition, specifically, social variants of extended cognition (socially extended cognition).

What is the unit of analysis; what is (not) part of the cognitive system?

The unit of analysis (and focus) is the (human) agent, including his/her brain and body, and the environment. Specifically: the agent and the particular object, artefact, or action and behaviour of ‘the other’ that the agent’s cognition ‘extends into’ in the sense of (temporarily) becoming a constitutive part of the agent’s cognitive system.

What is the underlying conception of the body and what ‘makes’ a body capable of cognition?

The conceptualisation of the (human) body seems comparable with embodiment as structural coupling and physical embodiment. Seeing that forms or principles of cognitive extension may include functionalist/computational, as well as dynamic and/or enactive interpretations, the underlying concept of the body may be conceptualised differently. With regard to extension via functional parity, it seems to be in line with body-functionalism/embodied functionalism. In the case of cognitive extension via integration, the conception appears to be more compatible with the notion of body-conservatism). Concerning extension via incorporation, i.e. through phenomenological transparency, the body is a ‘living’ and ‘experiencing’ body in the phenomenological sense and in the sense of body-enactivism/enactive embodiment.

What is the underlying conception of the environment?

In all three variants of cognitive extension, the environment is – essentially – an external, i.e. extra-bodily, resource that can be actively created, used, or manipulated so that it may become a

part of cognitive system. The difference between the variants mainly pertains to *how* this is done or how the ‘external’ relates to the agent (see also next question).

What is the underlying conception of the relation between body and environment?

It is co-constitutional relation. That is, the physical and/or social environment (or rather, parts of it) temporarily become a constitutive part of the agent’s cognitive system – which includes the agent’s body, but is based on different notions of embodiment – either through functional parity, integration, or incorporation.

What is the underlying conception concerning actions in and upon the environment?

Simply put, actions are a form of embodied perception and/or cognition. Depending on the underlying perspective (functionalist and/or computational vs. dynamic and/or enactive) actions in and upon the environment may be leaning towards body-conservatism, or towards a dynamic and/or enactive conception of embodied actions. In the case of functional parity and/or integration, the body contributes – to a different extent in terms of being body-based – to the processing of (social) information, i.e. extra-bodily aspects or processes, and actions are outcomes of the processing of (social) information and resulting motor instructions that produce motor functions in accordance with the respective environment. In the case of incorporation, actions are better construed as a kind of embodied cognition and/or (direct) perception in which actions occur in the context of dynamic body-environment interactions.

What is the underlying conception concerning actions towards others in the environment?

In short, actions upon the environment, including others, are a form of active externalism. (See also previous question concerning actions in and upon the environment).

What is the underlying conception concerning interactions with others in the environment?

Depending on the conception of the body, functionalistic/computational as well as dynamic interpretations may be applied to conceptualise interactions with regard to the relation to the external (social) resource and/or ‘the other’. The underlying conceptions of interactions with others may be based on either an individualistic (methodological individualism) perspective, or a more relational or bi-directional conception that focuses on social scaffolds as part of the interaction, rather than on the individual, respectively.

Having conceptually investigated the theory/concept on the basis of the analysis road map, the issue of fruitfulness can be addressed.

In summary: What, if any, aspects of interest with regard to the phenomenon of social human-ITS interaction are focused on and/or can be adequately captured when applying this theory/concept as a framework?

The focus is on the human learner.

In the case of social extension via functional parity and/or integration (stronger focus on the body), instances in which the human learner's cognitive – temporarily – extends to include the ITS may be captured. This may pertain to the ITS-agent's actions and/or the information it provides, how it provides it, as well as by which means or tools the human agent interacts with it. During situations like the question-repetition format described in chapter 4.3.2, i.e. when formulating an answer to tutor question the learner does not know, and, thus, receives help via hints, pumps, and the like, these actions may temporarily become a constitutive part of the human's cognitive system. However, the ITS would probably best be conceptualised as a kind of 'social tool' that can, temporarily, become a constitutive part of the human learner's cognitive system during the interaction with the ITS. Although the ITS agent produces actions (providing hints, motivational prompts, gestures, etc.) and the 'external information' can be retrieved from it in an interactive manner, it still 'holds' information more in the sense of the 'notebook' than 'Inga' from classical the 'Otto-example' (as in (Clark, 2010b; Clark & Chalmers, 1998a, 2010a).

In the case of social extension based on the concept of (mutual) incorporation, this is likely to not capture the social interaction and/or cognition during a tutoring situations as described in part 2. It is more likely that the concept of (mutual) incorporation adequately captures aspects of interest during tutoring or learning situations than involve coordinated movement, like dancing or two-person sports.

Finally: In what manner, if any, can this theory/concept be fruitfully applied to the case of social interaction between human and ITS?

In the case of social extension via functional parity and/or incorporation, the concept of (social) extension may be applied to the ITS-agent in terms its visually/animated embodied actions (including language and generally referring to how content is delivered by the ITS-agent) during tutoring. I.e. gestures or facial expressions – if they are considered to be social (probably in an as-if mode) – may be interpreted as a kind of external 'social tool' that temporarily extends the agents

cognitive system. However, particularly in case of the parity principle, the criteria for extension would likely need to be re-conceptualised in a (minimal) sense and also require further conceptual clarification with regard to temporal aspects (e.g. first or regular usage of ITS, experience with or disposition towards ITS in general). These criteria are: reliably available, typically invoked, retrieved more or less automatically endorsed, deemed about as trustworthy as something retrieved clearly from (someone's) biological memory, easily accessible as and when required (cf. (Clark, 2010b)). With regard to integrationist approaches, the requirements for 'collectivity' or 'collective action' would need to be addressed, i.e. cognitive integration in the sense of *collectively* generating new cognitive capacities. This requires a re-conceptualisation in the sense of a minimal approach (e.g. like the minimal approach to joint action in (Strasser, 2018a)) to be fruitful.

In the case of social extension based on the concept of (mutual) incorporation, an application of this concept does not appear to be fruitful. While the computer mouse the learner is using may become phenomenologically transparent, the ITS is unlikely to qualify as an entity that can become incorporated into the cognitive system.

5.4.2 Social Interaction as Emergent Extended Cognition

Summary:

The approach of emergent extended cognition (Hypothesis of Emergent Extended Cognition, HEEC, (Varga, 2016)) is a complementary conception to the work of Clark and Chalmers (1998a, 2010a). In essence, its main argument is that the prime example of Otto and his notebook is based on a weak notion of emergence that cannot account for 'real' (social) encounters and interactions. 'Real-world' social interaction require a stronger notion of emergence to adequately capture their complex dynamics.

The (emergent) properties of the 'Otto-notebook-system' are linear and "simply aggregate decomposition" (Varga, 2016, p. 2484); they are a direct and predictable result from Otto's belief that the location is in the notebook. The Otto-notebook-system is an 'aggregate decomposition' that could be predicted from a 'pre-emergent' state (cf. (ibid)). Such a conception does not suffice to capture many 'standard' or 'real-world' instances of dyadic social interaction, because they tend to involve non-linear elements of irreducibility and uncontrollability (ibid). Socially extended cognition may involve dyadic social interactions in which cognition is not about achieving a pre-defined task by manipulating external resources. Rather, it emerges as an unintended result of social interaction. Such dyadic social interactions are a distinctly different extended phenomenon

than the ones described by Clark and Chalmers (1998a, 2010a). There is a need to “take into account the asymmetry in the dyadic system, and thus the gradual difference between a prime cognizer and a ‘mere’ cognition supporter. [...] we need to bear in mind that our dyadic system arises spontaneously, without involving common goals and knowledge about domains of expertise” (Varga, 2016, p. 2479).

The HEEC (ibid) has two versions:

‘Emergence₁’ and the HEEC₁ are intended to capture irreducible systemic properties. That is, the systemic properties of a (cognitive) system are not reducible to properties of the system’s parts. Irreducibly emergent properties of a system arise autonomously as an uncontrollable effect of (social) interaction. Emergence₁ involves ‘diachronic novelty’. That is, (non-trivially) novel emergent properties.

‘Emergence₂’ and the HEEC₂ additionally capture top-down influences in extended cognitive systems. ‘The whole’, i.e. the emergent systemic property, directly influences the behaviour of the system’s parts.

What is the paradigm?

The theory/concept under consideration is associated with extended cognition, specifically, social variants of extended cognition (socially extended cognition) in the sense of emergent extended cognition.

What is the unit of analysis; what is (not) part of the cognitive system?

The unit of analysis (and focus) are the two agents, including their respective brain and body, and the environment, and the emergent interactions and/or aspects of it. Specifically: the agent and the particular ‘other’ who become a constitutive part of the cognitive system that emerges during dynamic and complex interactions between two agents.

What is the underlying conception of the body and what ‘makes’ a body capable of cognition?

This appears to be unclear. Likely, the conception is similar to body-enactivism/enactive embodiment, the historic notion of embodiment, and possibly organismic embodiment. The living and or enactive body in a specific (social) environment seems to be conceptualised as an autonomous entity in the sense of an autonomous agent or system.

What is the underlying conception of the environment?

This, too, appears unclear. The focus is on the cognitive system and its parts, as well as how these parts, i.e. the others in the environment and/or physical objects/artefacts they include in their interaction, interact and/or influence each other (in the sense of either Emergence₁ or Emergence₂).

What is the underlying conception of the relation between body and environment?

The relation between the body and environment (including others in the environment) is co-constitutive. It furthermore seems to be bi-directional and relational in an enactive sense, involving dynamic conceptions of autonomy.

What is the underlying conception concerning actions in and upon the environment?

Given that the concept of emergent extended cognition explicitly refers or pertains to dyadic interaction, see answer to the question concerning actions towards others in the environment (just below).

What is the underlying conception concerning actions towards others in the environment?

Action in and upon the environment (including others) are the means by which different parts or components of the (emergent) cognitive system relate and/or interact with each and may (unintentionally) yield irreducible systemic properties (emergence).

What is the underlying conception concerning interactions with others in the environment?

Interaction with others – in the case of dyadic interaction – may automatically or uncontrollably yield irreducible cognitive properties. These properties may be diachronically novel, i.e. novel emergent properties of the cognitive system (e.g. particular knowledge, skills, or abilities). The emergent cognitive system as a whole may also directly influence the systems parts (i.e. the two interacting agents, who may have an asymmetrical relation) in a top-down manner.

Having conceptually investigated the theory/concept on the basis of the analysis road map, the issue of fruitfulness can be addressed.

In summary: What, if any, aspects of interest with regard to the phenomenon of social human-ITS interaction are focused on and/or can be adequately captured when applying this theory/concept as a framework?

Generally, this framework may be adequate to capture emergent cognitive properties that arise

due to the interaction of two agents and may yield diachronically novel properties. Emergent extended cognition arises as an unintentional property of dyadic interaction, especially in situations in which no explicit goal is known or intended by (all) partaking agents. Initially, this seems to be of interest, because it may capture instances of social and/or cooperative learning, in which one agent may have different knowledge, skills, and abilities than the other. It may also capture instances in which there is no clear or explicit learning goal known to either or none of the two agents; e.g. instances of improvisational theatre or dance.

However, investigating the dyadic social interactions and actions during tutoring as a constitutive property of human-ITS-system that emerges during interaction, does not appear adequate. The tutoring situation does not appear to be a phenomenon HEEC is intended to capture. The human-ITS-system and or its properties likely do not produce ‘emergence’ as it is proposed in HEEC. Furthermore, underlying notion of the body of the agents constituting the system is unclear and may rely on a notion of embodiment as organismic, which would exclude non-living systems (the ITS).

Finally: In what manner, if any, can this theory/concept be fruitfully applied to the case of social interaction between human and ITS?

In a loose interpretation of fruitfulness in the sense of generating new research questions, HEEC may potentially be fruitful. Hypothetically, this framework *could* produce research questions concerning human-computer interaction (in the specific case of this master’s thesis: human/learner-ITS interaction during tutoring). For example, questions with regard to aspects of asymmetry in the interaction, as well as the modelling of dyadic interaction in tutoring situations as a complex system, or with regard to cooperative learning of novel skills or abilities. However, given the – likely to be limited – information on this particular approach that has been considered up to the time of writing this conceptual assessment, it appears to be more adequate to refrain from assessing its fruitfulness.

5.5 Social Interaction with ITS from a Distributed Cognition Perspective

As stated in chapter 3.6, conceptions in which

- cognition is *either co-dependent upon or co-constituted by extra-bodily processes* that comprise unit(s) of interacting agents who utilise natural and/or technological re-

source, are generally considered to be instances of *distributed cognition* (Newen et al., 2018; S. Walter, 2014b).

- Whereby the idea of distributed cognition – at least the ‘classical’ variant based on the work of Hutchins (1995a) – is the *computational/representational ‘counterpart’* of non- and/or anti-representational enactive approaches to cognition (Newen et al., 2018; S. Walter, 2014b).

5.5.1 Distributed Cognition (‘classical’ variant)

Summary:

Distributed cognition is based on the idea that cognitive processes are distributed or ‘spread’ over complex networks of interactions between a) different actors and b) technological devices or resources, as well as c) both actors and devices/resources in the environment. It is “a framework for thinking about cognition which seeks to understand how the cognitive properties of aggregates emerge from the interactions of component parts [...] from areas of an individual brain to communities of interacting persons.” (Hutchins, 2001a, para. 1). The ‘component parts’ Hutchins (ibid) writes about are multiple human actors and the material resources (e.g. artefacts or technological tools/devices) they use during a specific cognitive activity or task. Distributed cognition is concerned with cognitive processes that not only involve the interaction between multiple actors and/or material resources, but, in fact, are only possible *because* of these interactions (cf. (Hutchins, 2006; S. Walter, 2014a)). Simply put: distribution *is* interaction (Hutchins, 2001a, 2006).

What is the paradigm?

The theory/concept under consideration is distributed cognition, specifically, the ‘classical’ conception of distributed cognition after Hutchins.

What is the unit of analysis; what is (not) part of the cognitive system?

The unit of analysis – and, thus, the cognitive system – comprises the specific agents and the environmental resources (technologies, devices, tools, etc.) they utilise during a specify task or activity. Note that in comparison to extended approaches, there is no ‘centre’ or ‘locus’ of cognition in terms as a prime cogniser; cognition is ‘spread’ over all the interacting components.

What is the underlying conception of the body and what ‘makes’ a body capable of cognition?

The conceptualisation of the (human) body seems comparable with embodiment as structural coupling and physical embodiment. Seeing that the classic conception of distributed cognition retains functionalist/computational perspectives (as ‘wide computation’, cf. (Wilson, 1994)), the underlying concept of the body may be conceptualised in line with body-functionalism/embodied functionalism. It also seems compatible with the notion of body-conservatism, because the body is given a prominent role of relevance, since it is the means of actions and interaction; and interaction *is* distribution.

What is the underlying conception of the environment?

The physical and social environment is a ‘source’ of ‘component parts’ (including objects, artefacts, devices, or others) that can form an aggregate in which cognitive capacities are kind of ‘collective’ or ‘group’ cognition. This environment also includes social stimuli, which produce cognitive and bodily responses in the observer. The body of ‘the other’ is a part of this environment, i.e. it is an extra-bodily/environmental ‘component’.

What is the underlying conception of the relation between body and environment?

Depending on the specific form of interaction and or the situation given a specific cognitive task or activity, relations between others and their environment can be either co-dependent upon or co-constituted. Distributed cognition involves extra-bodily processes and comprise aggregates of interacting agents and/or artefacts.

What is the underlying conception concerning actions in and upon the environment?

Distributed cognition generally captures instances of cognition involving actions either towards physical, technological, or social components of the environment; the actions towards them and interactions with them is what constitutes, or at least enables, the distribution of cognition.

What is the underlying conception concerning actions towards others in the environment?

See above question.

What is the underlying conception concerning interactions with others in the environment?

Essentially, distribution *is* interaction (see also question concerning actions in or upon the environment).

Having conceptually investigated the theory/concept on the basis of the analysis road map, the issue of fruitfulness can be addressed.

In summary: What, if any, aspects of interest with regard to the phenomenon of social human-ITS interaction are focused on and/or can be adequately captured when applying this theory/concept as a framework?

All aspects of the interaction between the ITS and the human learner may be captured, as long as the primary focus lies on the interaction, e.g. during situations like the question-repetition mode (see part 2, chapter 4.3.2). Distributed cognition is not a specific kind of cognition that requires ‘capturing’, but is a perspective for thinking about or analysing cognition.

Finally: In what manner, if any, can this theory/concept be fruitfully applied to the case of social interaction between human and ITS?

As a framework, distributed cognition may be fruitfully applied to human-ITS interaction – conceptualised as an as-is or an as-if mode of social interaction or cognition – as long as the focus is on the interaction and not on a particular agent in the sense of a ‘centre’ of cognition. This framework also appears to be fruitful because the ITS may be (re-)conceptualised as either an agent, device, or a social tool without meeting requirements of non-representationalism and or enactivism (autonomy, ‘living’ body). Also, distributed cognition has already been employed in HCI research (e.g. (Perry, 2003a)

5.5.2 Distributed Cognition (Third-Wave Externalism)

Summary:

Another approach to distributed cognition is the ‘socially and culturally distributed cognition’ by Cash (2013). This approach belongs to the so-called third-wave externalist approaches. It has a broader understanding of cognitive extension that goes beyond the typical examples of notebooks or technological devices. In contrast to the distributed cognition approach of Hutchins – which is based on computationalism/representationism – third-wave externalists typically draw on enactive approaches to cognition. In this account to cognitive distribution, individual cognition takes place within a mutually co-constructed larger context of social and technological systems, and cultural practices and institutions, like for example the legal system (Cash, 2013; S. Gallagher, 2013).

Notice:

Socially and culturally distributed cognition (and similar third-wave approaches) is particularly concerned with the issue of social interaction. However, they generally argue that cognition does not only involve social interaction, but also complex cognitive-cultural collectives, social practices, and institutions that allow for socially extended cognition. This approach does not only take into account direct social interaction, but addresses cognitive extension as interaction with cognitive-cultural practices and institutions as a complex two-way interaction between individual and practices (Cash, 2013; S. Gallagher, 2013). Its focus does not lie on direct interaction. This distributed cognition perspective is only of interest if ‘more’ than the micro-level is focused on, e.g. the educational system and/or associated social practices. If the broader context of practices and institutions (in a ‘more-than’ micro-structural sense) *are* of interest with respect to the interacting individuals, then this form of distributed cognition may lead to fruitful research questions and/or insights. However, in the concrete case of this master’s thesis, it does not seem adequate to apply this paradigm, and doing so would not or only partially be fruitful. This is because the conception of ‘social interaction’ within third-wave externalism is differently conceived and can – in terms of a concrete analysis of the specific interaction of interest for this master’s thesis – not meaningfully be applied to ‘only’ the direct, moment-to-moment interaction and its situational structure. Hence, it will not be discussed along the analysis road map.

5.6 Social Interaction with ITS from an Enactive Cognition Perspective

As stated in chapter 3.6, conceptions in which

- cognition is ‘localised’ in neither the brain, nor the body, nor in the natural, social, and/or technological environment, i.e. cognition involves action and interaction, thus being a *relational* and temporally extended process of *sense-making between autonomous systems*, are generally considered to be approaches of *enactive cognition*, in which:
 - the question of the ‘where’ of cognition is replaced with a relational understanding of cognition with a constitutional, rather than a dependence-claim (cf. (S. Walter, 2014b)). Specifically:

- a cognitive processes that is *co-dependent* on the ability or disposition to act is *weakly enacted*, and
- a cognitive process that is *co-constituted* by the ability or disposition to act is considered to be *strongly (or radically) enacted* (Newen et al., 2018).

Thus, for the purpose of conceptual considerations and reasoning within the context of this master's thesis, one may postulate that:

A conception in which

- cognition occurs in a social environment in which cognition is 'localised' in neither the brain, nor the body, nor in the natural, social, and/or technological environment, i.e. cognition involves social action and interaction, may be interpreted in the sense of social contingencies or participatory sense-making (as in (McGann & De Jaegher, 2009) and (De Jaegher & Di Paolo, 2007, 2008), respectively).

5.6.1 Self-Other/Social Contingencies

Summary:

McGann and De Jaegher (2009) formulate an approach to social perception via the concept of self-other contingencies, which they also refer to in broader terms, e.g. as social contingencies or intersubjective contingencies. Contingencies of action are not only sensorimotor, but also social (REF). Perception is conceived of as skill in the sense of mastering the set of contingencies or regularities within a specific domain (ibid). Thus, social perception is a kind of skill within the social domain. This way, social perception – conceived of as a social skill – can be examined in terms of self-other contingencies that structure social interaction (ibid).

Notice:

Although the concept of social contingencies seems to be suited to capture aspects of direct social interaction, it does not appear meaningful to consider it with respect to the social interaction with ITS. That is because the ITS would need to qualify as 'the other' in an enactive sense (see discussion of participatory sense-making below). In short, its application appears to be problematic and the assessment of its fruitfulness appears to be more meaningful when considering it in relation to the concept of participatory sense-making. The concept of social/self-other contingencies seems to be strongly related to participatory sense-making and may even be considered as a kind of basis for it. For these reasons, it will not be discussed explicitly along the analysis road map. In-

stead, the discussion will focus on participatory sense-making (like it was done in part I of this master's thesis).

5.6.2 Participatory Sense-Making

Summary:

Participatory sense-making (De Jaegher & Di Paolo, 2007, 2008) is an enactive account to social cognition. Rather than focussing on the enactment of meaningful structures through an *individual* agent and its environment, participatory sense-making is concerned with the sense-making processes that emerge in the interaction of (at least) two agents. Specifically, participatory sense-making is defined as: “the coordination of intentional activity in interaction, whereby individual sense-making processes are affected and new domains of social sense-making can be generated that were not available to each individual on her own.” (De Jaegher & Di Paolo, 2007, p. 497).

Crucially, participatory sense-making of this kind involves embodied and expressive actions in the context of social interaction. This requires coordination, and autonomy. Coordination is defined as the non-accidental correlation between the behaviours of at least two systems that are either i) sustained in coupling, or ii) have been coupled in the past, or iii) have been coupled to another common system. Here, correlation is “a coherence in the behaviour of two or more systems over and above what is expected, given what those systems are capable of doing.” (De Jaegher & Di Paolo, 2007, p. 490). In biological systems, patterns of coordination can directly influence the interactors and their disposition to maintain or modify their encounters, e.g. via embodied actions such as gestures or utterances that continue or end the coordinated coupling.

For De Jaegher and colleagues (2010a; 2007, 2008) instances in which this kind of bi-directional influence between interactors (in the sense the autonomous systems sustained in coordinated coupling) and their emergent interaction (i.e. the dynamic pattern of coordination) occurs are instances of *social interaction*. They explicitly define an autonomous system as a system that is composed of several processes that *actively generate and sustain an identity* as a distinct, *operationally closed* entity under *precarious conditions*. ‘Precarious’ pertains to the fact that if the system was not organised in terms of a network of its conditional processes, isolated component processes of the system would tend to extinguish. Thus, an autonomous systems is a system that maintain its autonomous identity under precarious conditions (ibid). This aspect is pivotal for their specification of social interaction. Social interaction processes may take on a form of

autonomy themselves, and resulting relational dynamics are autonomous strictly in the sense of the definition provided above (ibid). Participatory sense-making is not restricted to humans; direct cross-species interactions or the direct interaction with robots (cf. (De Jaegher et al., 2010a)) may also qualify as participatory sense-making, as long as they fulfil the criteria of being autonomous.

What is the paradigm?

The theory/concept under consideration is associated with enactivism/enactive cognition, specifically, it appears to be compatible with 4EA interpretations of enactivism that draws on the enactive concepts of autonomy, sense-making, emergence, and experience.

What is the unit of analysis; what is (not) part of the cognitive system?

The interaction of autonomous systems is assumed to be the proper unit of analysis. Within cognitive systems, sense-making is enacted through the dynamic and reciprocal interactions between agents and their environment – cognition comprises all of these relational elements. Interaction is not independent from the individual agents and individual agents are partially dependent on it because it constitutes the process of participatory sense-making. i.e. participatory sense-making means that processes of sense-making are jointly enacted through the dynamic interactions of the participating agents and their environment.

What is the underlying conception of the body and what ‘makes’ a body capable of cognition?

The body is a human body (or a body with comparable characteristics and capacities regarding agency) as a ‘living’ body in the sense of body-enactivism or enactive embodiment. It is also the body of an *autonomous* agent.

What is the underlying conception of the environment?

The environment (including physical, social, and cultural dimensions) is involved in cognitive processes in dynamic and relational terms; cognition is not ‘located’ in (either) the environment nor e.g. the body or the brain. The agent exists in relation to its specific environment and ‘meaning’ is enacted through the agent’s capacity to act.

What is the underlying conception of the relation between body and environment?

The relationship between the body and the environment is co-constitutional (but may, in some in-

stances, be co-dependent), and is defined by an agent's (and its bodily capacities) ability to act in and upon this environment.

What is the underlying conception concerning actions in and upon the environment?

Action is cognition. The ability or disposition to act in and upon the specific environment an agent is situated in, is the basis for its ability to sustain itself and its identity under (potentially) precarious conditions. Being capable of actions (in the sense of agency) also seems to be a requirement for the autonomy of individual agents and/or their interactions (see also question concerning interactions towards others).

What is the underlying conception concerning actions towards others in the environment?

The ability or disposition to act towards others is the basis for meaning to be – in a participatory manner – enacted as an emergent property that the agents would not have generated if they acted on their own.

What is the underlying conception concerning interactions with others in the environment?

Interactions with others are social interactions if there is a bi-directional influence between interactors (i.e. autonomous systems sustained in coordinated coupling) and the emergent interaction (i.e. the dynamic pattern of coordination). Due to its dynamic and emergent characteristics, this interaction may also take on a form of autonomy. In this sense, (social) interaction may constitute (social) cognition.

Having conceptually investigated the theory/concept on the basis of the analysis road map, the issue of fruitfulness can be addressed.

In summary: What, if any, aspects of interest with regard to the phenomenon of social human-ITS interaction are focused on and/or can be adequately captured when applying this theory/concept as a framework?

When employing the framework of participatory sense-making, social interaction between a human and an ITS would likely qualify as *individual* sense-making; not as participatory sense-making. This is because the ITS-agent cannot be considered an 'agent' in the enactive sense.

Finally: In what manner, if any, can this theory/concept be fruitfully applied to the case of social interaction between human and ITS?

Ways to apply participatory sense-making fruitfully appear to be very limited and/or fruitful only in a – seemingly – paradoxical way. On the one hand, participatory sense-making very adequately describes and captures dyadic social interaction. On the other hand, the ITS is likely not to qualify as an agent capable of ‘participating’ in such a sense-making process (due to issues of autonomy, coordination). A possibly fruitful question would be why this is the case and whether – and if so, under what conditions – an ITS agent may (socially) contribute to or join in sense-making-equivalent process without being an autonomous agent in the strong enactive sense. Minimal approaches to agency, coordination, and autonomy – that are not based on enactive conceptions – may be able to capture instances that are similar or comparable to social interaction as it is described in this theoretical framework. In this sense, participatory sense-making would be a kind of conceptual starting point or a point of reference, and not an explanatory framework. Albeit seeming paradox, or even ‘far-fetched’, this may be a way for developing fruitful conceptions that can aid the empirical investigation and understanding of social interaction (as-if and as-is mode) with ITS or similar artificial systems/agents.

6 Conclusions across Part I and II

The present master’s thesis was concerned with the topic of social interaction in the context of social cognition, and aimed to investigate conceptualisations of social interaction within cognitive science. Additionally, a specific case of human-computer interaction as a social kind of social interaction was focused on; the interaction between a human learner and an ITS. For these purposes, three research questions (RQ) were formulated, each associated with one of the three parts of the master’s thesis.

Part I of this thesis was concerned with RQ1:

- RQ1: What conceptions of social interaction currently exist in classical/computational and situated/4E cognitive science paradigms?

After an introduction to the history of cognitive science, including major paradigm shifts and turn (cognitive revolution, embodied turn, pragmatic turn, social/intersubjective/interactive turn), it was demonstrated that the phenomenon of social interaction appears to be of increased relevance within cognitive science(s). However, when focusing on social instances of interaction and cognition, several conceptual and methodological issues are to be considered. With respect to the scope

of this thesis, those are: methodological individualism, body-social problem, and human-centredness. With reference to the latest paradigm shifts, as well as the associated conceptual and methodological issues, concepts that are closely related to social interaction were described across several disciplines (philosophy, psychology, neuroscience, AI and computer science). The main focus of part I was on the systematic review of conceptions of social interaction in the cognitive science paradigms, in which situated cognition approaches were highlighted. The level of paradigms was chosen as the main level of conceptual analysis in order to reveal the theoretical assumptions, metaphysical claims, and methodological perspectives with regard to conceptions of social interaction. The findings of the conceptual work addressing RQ1 are depicted in the following illustration (figure 7).

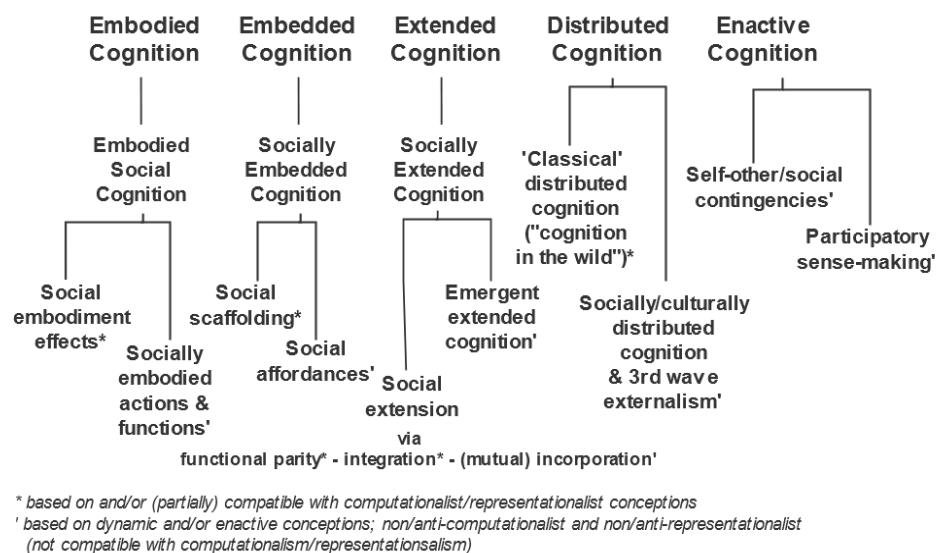


Figure 7: Identified 'social variants' of situated cognition paradigms and associated theories concerned with social interaction and/or cognition

Part II introduced the field of Intelligent Tutoring Systems (ITS), presented the theoretical research context with regard to RQ2, and described a concrete research design proposal that addresses RQ2.

- RQ2: How do power and related affect influence and shape the social interaction with ITS regarding: i) social roles; ii) interaction and the (situational) structures it is embedded in; and iii) social relations?

The proposed research setting comprised three consecutive tutoring sessions about the history of the University of Vienna, that have been implemented with the conversational ITS AutoTutor. Three to five native speakers, aged 20-35, were planned to participate in all three sessions,

which were to take place weekly. Special attention was given to the tutoring session design with respect to its structure, instructional modes, and tutor actions in order to create a session that would, potentially, allow for the occurrence of different power dynamics and associated affective behaviours and actions. The methods for data analysis comprise the qualitative methods of Grounded Theory, specifically Constructivist Grounded Theory (CGT, (Charmaz, 2006)) and Situational Analysis (SA, (Clarke, 2015, 2018)). CGT is a theory-generating inductive method of qualitative data collection and analysis. It is well-suited for studying social actions and processes because it places actions and how they come about social situations and social structures in the focus of its analysis (Charmaz, 2006). SA extends and complements CGT. In contrast to CGT, SA focuses on situational aspects (Clarke, 2018; Clarke et al., 2015). SA supplements the analysis of actions and basic social processes of CGT with an ecological focus on the situatedness of actions, i.e. it centres on the situation as a conceptual structure to focus its analysis (ibid). A crucial aspect in regard to the present thesis is that SA explicitly addresses the role of non-humans in social situations (ibid). Combined, CGT and SA appeared to be adequate means to address all three levels of RQ3.

Finally, part III was concerned with RQ3 and comprised a conceptual discussion of the conceptions of social interaction identified in Part I with regard to their applicability – specifically, their fruitfulness – when being applied to the phenomenon of social interaction between a human learner and an ITS. For this purposes, an ‘analysis road map’ (figure 6), that has been developed for this master’s thesis, has been used.

- RQ3: Which of the conceptions from RQ1 are relevant for and can be fruitfully applied to the case of social interaction with ITS?

The (preliminary) findings with regard to RQ3 are depicted in the following table (table 7). The table summarises the following information: 1) the respective paradigm, 2) the related theory or concept, 3) the focus regarding the phenomenon of social human-ITS interaction when applying this theory or concept as a conceptual framework, 4) a statement on whether this framework would adequately capture or depict the phenomenon with regard to the aspects of interest, as well as 5) a statement on whether the theory/concept can be applied in a fruitful manner.

(Sub-)Paradigm	Theory/ concept	Focus/aspects of interest	Adequate framework?	Fruitfully applicable?
Embodied Social Cognition	Social embodiment effects	Bodily actions, movement, e.g. gestures, facial expression during interactions	Partially (but limited: focus on human cognition & individual)	Yes
Embodied Social Cognition	Social actions and functions of the body	Bodily actions, movement, e.g. gestures, facial expression during interactions	Partially (but limited: focus on human body)	Partially (but limited: focus on human body)
Socially Embedded Cognition	Social scaffolding	Environment/environmental effects during agent-environment interaction, (incl. technologies, artefacts, humans, non-human agents)	Yes	Yes
Socially Embodied Cognition	Social affordances	Environment/environmental effects during agent-environment interaction, (incl. technologies, artefacts, humans, non-human agents)	Yes	Yes
Socially Extended Cognition	Social extension via functional parity	Environment, interaction of agent with environment (incl. technologies, artefacts, humans, non-human agents)	Partially (ITS as 'social tool')	Yes
Socially Extended Cognition	Social extension via integration	Environment, interaction of agent with environment (incl. technologies, artefacts, humans, non-human agents)	Partially (ITS as 'social tool')	Yes
Socially Extended Cognition	Social extension via (mutual) incorporation	Dynamically unfolding interaction particular focus on bodily interaction and bodily experiences	No	No
Socially Extended Cognition	Emergent extended cognition	Dynamically unfolding interaction and its emergent properties, particular focus on bodily interaction in self-other-system incl. their environment	Unclear (may rely on organismic notions of embodiment)	Unclear (may rely on organismic notions of embodiment)
Distributed Cognition	Distributed cognition ('classical')	Environment, interaction of agent with environment (incl. technologies, artefacts, humans, non-human agents)	Yes	Yes

(Sub-)Paradigm	Theory/ concept	Focus/aspects of interest	Adequate framework?	Fruitfully applicable?
Distributed Cognition	Distributed cognition (3 rd wave externalism)	Interaction of agent with environment (incl. technologies, artefacts, humans, non-human agents, but primarily social norms, structures, practices, institutions)	n/a (was not discussed via analysis road map)	Potentially (but: focus on practices, institutions; not direct dyadic interaction)
Enactive Cognition	Self-other/ social contingencies	Dynamically unfolding interaction, body-environment interactions (incl. others)	No (was not discussed via analysis road map)	Partially (but very limited: ITS ≠ autonomous ‘other’)
Enactive Cognition	Participatory sense-making	Dynamically unfolding interaction, body-environment interactions (incl. interaction with others)	No	Partially (but very limited; no ‘proper’ autonomous interaction)

Table 7: Summary regarding fruitfulness of paradigms and theories/concepts with regard to social human-ITS interaction

6.1 Limitations and Future Work

The first limitation with regard to the work presented in the context of this master’s thesis is the *selectivity regarding theoretical frameworks and paradigms*. For example, the paradigm of predictive processing or predictive coding has not been addressed within this thesis, and should be considered in future research. With regard to theoretical frameworks, it should also be noted that theories from the cognitive science disciplines of biology and linguistics have not been considered. Given the topic of social interaction and its relation to language, as well as the importance of natural language in the interaction with ITS, linguistic conceptions may fruitfully contribute to future research.

Although a selection of related concepts has been reviewed, the *focus on the level of paradigms instead of disciplinary conceptions* may be criticised. It should be noted, that the focus on the level of paradigms does not entail that disciplinary perspectives and conceptions should be ignored entirely. Disciplinary conceptions may provide additional explanatory value, but have, due to time and ECTS-related constraints, not been focussed on.

An issue that is generally a matter of debate (i.e. within current theoretical and conceptual discourse, see e.g. REF), is the *use of social interaction as a means to study social cognition*. By fo-

cussing on the concept of social interaction from the very start, potential other concepts – that may be (more) relevant in and for social cognition – have automatically been neglected. Although the debate of social interaction as an explanatory tool for social cognition has not yet been resolved, social interaction is considered to be a fruitful *exploratory* tool for investigating social interaction within the context of this master’s thesis.

Similarly, the matter of human-centredness, and the approach of *applying concepts or theories from human-human to human-computer social interaction and cognition* (which this thesis too has partially followed), affords further research regarding conceptual elaborations, clarifications, and investigations.

Crucially, *the state of empirical research and the proposed research design* needs to be addressed with regard to limitations. It should be noted that, while the curriculum and session structure for all three sessions has been finished and, the sessions were only partially implemented and debugged (session 3 is not finished). Neither have they been pre-tested with participants, nor have they been used for any empirical analysis of the social interaction between learner and ITS. Consequently, the data collection and analysis as it is described in the chapters 4.3.1 to 4.3.4 has not taken place. Furthermore, it is currently not possible to use those exact tutoring sessions for purposes of the proposed study. This is not to say, however, that the proposed methodology cannot be applied to other tutoring sessions and scenarios in the future. It might be possible (but has, so far, not been planned with certainty or been arranged) to investigate power relations and related affect between not just one human learner and one ITS-agent, but potentially also among a human learner and multiple agents that have been designed to perform different instructional roles. E.g. as it is the case in CSAL²⁶ (Graesser, 2016a; Shi et al., 2018), which targets the area of adult literacy (among others, reading, media, and computer literacy) and has been developed by the University of Memphis. Furthermore, the general research design and methodology may be applicable to human-human tutoring situations, as well as comparisons between human and artificial tutors (potentially in conjunction with quantitative and/or bio-physiological measurements). Such research may yield insights with regard to studying social interaction and its role in the context of (social) cognition during human-computer interaction, specifically, learning systems like ITS. It may also provide insights regarding guidelines for the design and development of such system in order to enable and support learning by means of social interaction.

With regard to the research design, *employing qualitative methods in the study of social interaction and cognition* may warrant critique, but also offers opportunities for future research. Based

26 For more information on this ITS, see also official CSAL website: <https://sites.gsu.edu/csdl/> (last accesses: 29.04.2022)

on feedback and inputs received at the ITS conference 2016, future research should focus on constructing standardised observation schemes based on insights from qualitative investigations, and/or a mixed-methods approach. This may include standardised measures such as the Attitude Toward Tutoring Agent Scale, ATTAS, which has been used with AutoTutor (Adcock and Van Eck 2005) and the Agent Persona Instrument, API (A. Baylor & Ryu, 2003). Particularly with regard to affect, considering bio-physiological measurements (e.g. heart rate, skin conductance, neural activity) may be of interest, as it could yield complementary results. Within the proposed research design, affect was planned to be investigated by observational means, including qualitative observations of the tutoring sessions, as well as first-person reporting on affect by means of open interviews. Note also that, although some versions of AutoTutor, i.e. Affective AutoTutor (D'mello & Graesser, 2012; D'Mello et al., 2011), are able to detect and classify affect, this is not implemented in the version that was used for the present master's thesis.

Additionally, the *investigating of social interaction in social cognition in general, as opposed to focussing on power and related affect in particular*, should be addressed. The initial focus of the present master's thesis was on the role of power and related affect in social interaction (part 2). During the course of writing the thesis, this interest changed and social interaction in general was focussed on (part I and part III). Preparatory work and literature review for this thesis did consider conceptions of power and affect in cognitive science(s). This could, however, due to time and ECTS-related constraints, neither be finished, nor be included in the master's thesis. Future work should explicitly address cognitive science conceptions and relate them to the social science conceptions applied within the present thesis. Research on affect and power within cognitive science may provide promising frameworks with regard to social interaction and cognition – in humans-human interaction, as well as in interaction between or with artificial agents. For example, work on situated affectivity, e.g. embodied, extended, and enactive affectivity (Candiotti, 2016; Colombetti, 2014, 2017, 2018a) and extended (Krueger 2014; Slaby 2014) and embodied emotions (Niedenthal et al., 2014; Prinz, 2004), as well as affective affordances (Rietveld & Kiverstein, 2014; van Vugt et al., 2006), social cognition of power (Guinote, 2013), and power and affordances (Guinote 2008), should be considered in future research.

Finally, with regard to the *assessment of fruitfulness* as it was demonstrated in this master's thesis, it should be noted that this assessment is limited and preliminary. The purpose of this thesis was to identify, review, and consider a large variety of conceptions of social interaction within cognitive science. As a result, the selected theories and concepts have not been described 'in-depth', as the master's thesis rather meant to demonstrate the range or 'quantity' of possible

concepts/theories. Hence, the information considered for the purpose of assessing the fruitfulness of a certain concept/theory (with regard to the phenomenon of social interaction between a human learner and an ITS) is limited as well. The ‘analysis road map’ that has been used to discuss concepts with regard to their fruitfulness should be seen as an initial attempt of making such assessments. Many of these assessments, as well as the concepts to be assessed, require further conceptual clarification, which presents an opportunity for future research.

7 References

- Aamodt, A., & Plaza, E. (1994). Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AI Commun.*, 7(1), 39–59.
- Abu-Lughod, L., & Lutz, Catherine, A. (1990). Introduction: Emotion, Discourse, and the Politics of Everyday Life. In *Language and the Politics of Emotion* (pp. 1–23). Cambridge University Press.
- Adam, C., Johal, W., Pellier, D., Fiorino, H., & Pesty, S. (2016). Social human-robot interaction: A new cognitive and affective interaction-oriented architecture. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9979 LNAI, 253–263. https://doi.org/10.1007/978-3-319-47437-3_25
- Adams, F., & Aizawa, K. (2010). Defending the Bounds of Cognition. In R. Menary (Ed.), *The Extended Mind* (pp. 67–80). The MIT Press. <https://doi.org/10.7551/mitpress/9780262014038.003.0004>
- Adcock, A. B., & Van Eck, R. N. (2005). Reliability and Factor Structure of the Attitude Toward Tutoring Agent Scale (ATTAS). *Journal of Interactive Learning Research*, 16(2), 195–217.
- Ahmed, S. (2004). Affective Economies. *Social Text*, 22(2), 117–139. https://doi.org/10.1215/01642472-22-2_79-117
- Ahrendt, H. (1958). *The Human Condition*. University of Chicago Press.
- Alessandroni, N. (2018). Varieties of embodiment in cognitive science. *Theory & Psychology*, 28(2), 227–248. <https://doi.org/10.1177/0959354317745589>
- Aleven, V. (2003). Using background knowledge in case-based legal reasoning: A computational model and an intelligent learning environment. *Artificial Intelligence*, 150(1–2), 183–237. [https://doi.org/10.1016/S0004-3702\(03\)00105-X](https://doi.org/10.1016/S0004-3702(03)00105-X)
- Aleven, V., McLaren, B. M., Sewall, J., & Koedinger, K. R. (2005). *Example-Tracing Tutors: A New Paradigm for Intelligent Tutoring Systems*.
- Aleven, V., McLaren, B. M., Sewall, J., & Koedinger, K. R. (2009). A New Paradigm for Intelligent Tutoring Systems: Example-Tracing Tutors. *Int. J. Artif. Intell. Ed.*, 19(2), 105–154.
- Allen, M. (2017). Ethnographic Interview. In *The SAGE Encyclopedia of Communication Research Methods*. SAGE Publications, Inc. <https://doi.org/10.4135/9781483381411.n168>
- Anderson, J. R. (1990). *The adaptive character of thought*. Erlbaum.
- Anderson, J. R. (1993). *Rules of the Mind*. Erlbaum.
- Anderson, J. R., Boyle, C. F., & Reiser, B. J. (2018). *Intelligent Tutoring Systems* (R. Nkambou, R. Azevedo, & J. Vassileva, Eds.). Springer International Publishing. <https://doi.org/10.1007/978-3-319-91464-0>
- Andreou, G., & Galantomos, I. (2009). The Native Speaker Ideal in Foreign Language Teaching. *Electronic Journal of Foreign Language Teaching*, 6(2), 200–208.
- Aneja, G. (2014). Disinventing and Reconstituting Native Speaker Ideologies through the Classroom Experiences of International TESOL Students. *Working Papers in Educational Linguistics*, 29(1), 23–39.
- Arendt, H. (1958). *The Human Condition*. University of Chicago Press.
- Auvray, M., & Rohde, M. (2012). Perceptual crossing: The simplest online paradigm. *Frontiers in Human Neuroscience*, 6(JUNE 2012), 181. <https://doi.org/10.3389/fnhum.2012.00181>
- Baggs, E. (2021). All Affordances Are Social: Foundations of a Gibsonian Social Ontology. *Ecological Psychology*, 33(3–4), 257–278. <https://doi.org/10.1080/10407413.2021.1965477>
- Baker, R. S. (2016). Stupid Tutoring Systems, Intelligent Humans. *International Journal of Artificial Intelligence in Education*, 26(2). <https://doi.org/10.1007/s40593-016-0105-0>

- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall, Inc.
- Bandura, A. (2001). Social Cognitive Theory: An Agentic Perspective. *Annual Review of Psychology*, 52(1), 1–26. <https://doi.org/10.1146/annurev.psych.52.1.1>
- Barnes, C. M., Ekart, A., & Lewis, P. R. (2019). Social Action in Socially Situated Agents. *2019 IEEE 13th International Conference on Self-Adaptive and Self-Organizing Systems (SASO)*, 97–106. <https://doi.org/10.1109/SASO.2019.00021>
- Barr, A., Beard, M., & Atkinson, R. C. (1976). The computer as a tutorial laboratory: The Stanford BIP project. *International Journal of Man-Machine Studies*, 8(5), 567–582. [https://doi.org/10.1016/S0020-7373\(76\)80021-1](https://doi.org/10.1016/S0020-7373(76)80021-1)
- Barsalou, L. W., Niedenthal, P. M., Barbey, A. K., & Ruppert, J. A. (2003). Social Embodiment. In *Psychology of Learning and Motivation* (Vol. 43, pp. 43–92). Academic Press. [https://doi.org/10.1016/S0079-7421\(03\)01011-9](https://doi.org/10.1016/S0079-7421(03)01011-9)
- Baylor, A., & Kim, Y. (2004). Pedagogical agent design: The impact of agent realism, gender, ethnicity, and instructional role. *Intelligent Tutoring Systems, 1997*, 592–603. https://doi.org/10.1007/978-3-540-30139-4_56
- Baylor, A., & Kim, Y. (2005). Simulating Instructional Roles through Pedagogical Agents. *International Journal of Artificial Intelligence in Education*, 15(2), 95–115.
- Baylor, A. L., & Plant, E. A. (2005). Pedagogical Agents as Social Models for Engineering: The Influence of Agent Appearance on Female Choice. *Proceedings of the 2005 Conference on Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology*, 65–72.
- Baylor, A., & Ryu, J. (2003). *The API (Agent Persona Instrument) for Assessing Pedagogical Agent Persona*. 448–451. <https://www.learntechlib.org/primary/p/13799/>
- Baylor, Amy L, Kim, Y. (2005a). Simulating Instructional Roles through Pedagogical Agents. *International Journal of Artificial Intelligence in Education*. <https://doi.org/10.1007/BF02504991>
- Beale, R., & Peter, C. (Eds.). (2008). *Affect and Emotion in Human-Computer Interaction. From Theory to Applications: Vol. LNCS 4868*. Springer-Verlag Berlin Heidelberg. https://doi.org/10.1007/978-3-540-85099-1_13
- Beer, R. D. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, 4(3), 91–99. [https://doi.org/10.1016/S1364-6613\(99\)01440-0](https://doi.org/10.1016/S1364-6613(99)01440-0)
- Ben-Ze'ev, A. (2009). The Thing Called Emotion. In P. Goldie (Ed.), *The Oxford Handbook of Philosophy of Emotion* (Issue February). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199235018.003.0003>
- Bericat, E. (2016). The sociology of emotions: Four decades of progress. *Current Sociology*, 63(3), 491–513. <https://doi.org/10.1177/0011392115588355>
- Bermúdez, J. L. (2019). *Cognitive Science. An Introduction to the Science of the Mind* (3rd ed.). Cambridge University Press. <https://doi.org/10.1017/9781108339216>
- Besold, T. R., & Kühnberg, K.-U. (2013). Kognition als Symbolverarbeitung: Das Computermodell des Geistes. In A. Stephan & S. Walter (Eds.), *Handbuch Kognitionswissenschaft2* (pp. 156–163). J.B. Metzler.
- Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction*, 12(2), 293–327. <https://doi.org/10.1145/1067860.1067867>
- Bierstedt, R. (1950). An Analysis of Social Power. *American Sociological Review*, 15(6), 730–738.
- Boiger, M., & Mesquita, B. (2012). The Construction of Emotion in Interactions, Relationships, and Cultures. *Emotion Review*, 4(3), 221–229. <https://doi.org/10.1177/1754073912439765>

- Bond, T. G. (2012). Piaget's Learning Theory. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 2634–2636). Springer US. https://doi.org/10.1007/978-1-4419-1428-6_39
- Borghini, A. M. (2018). Affordances, context and sociality. *Synthese*, 1–31. <https://doi.org/10.1007/s11229-018-02044-1>
- Bottaro, G. M., Roetzer, K., Schönberg, A., & Peschl, M. F. (2021). Lessons Learned from Covid-19: How to Design Virtual Enabling Spaces for Innovation Teams. In J. Fritz & N. Tomaschek (Eds.), *Konnektivität: Über die Bedeutung von Zusammenarbeit in der virtuellen Welt. University – Society – Industry, Band 10* (pp. 213–230). Waxmann.
- Bourdieu, P. (1977). *Outline of a Theory of Practice* (J. Goody, Ed.; Vol. 16). Cambridge University Press. <http://www.loc.gov/catdir/description/cam022/76011073.html>
- Breazeal, C. (2002). *Designing Sociable Robots*. A Bradford Book.
- Breazeal, C. L. (2000). *Sociable Machines: Expressive Social Exchange Between Humans and Robots*. Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science.
- Brooks, R. A. (1989). How to Build Complete Creatures Rather than Isolated Cognitive Simulators. In K. VanLehn (Ed.), *Architectures for Intelligence* (pp. 225–239). Erlbaum.
- Brooks, R. A. (1991a). Intelligence without Reason. *Proceedings of the 12th International Joint Conference on Artificial Intelligence - Volume 1*, 569–595.
- Brooks, R. A. (1991b). Intelligence Without Representation. *Artificial Intelligence*, 47(1–3), 139–159. [https://doi.org/10.1016/0004-3702\(91\)90053-M](https://doi.org/10.1016/0004-3702(91)90053-M)
- Brooks, R. A. (2014). *How to Build Complete Creatures Rather than Isolated Cognitive Simulators*. 239–254. <https://doi.org/10.4324/9781315807843-16>
- Bulmer, M. (1979). Concepts in the Analysis of Qualitative Data. *The Sociological Review*, 27(4), 651–677. <https://doi.org/10.1111/j.1467-954X.1979.tb00354.x>
- Burr, V. (2015). *Social Constructionism*. Routledge.
- Cade, W. L., Copeland, J. L., Person, N. K., & Mello, S. K. D. (2008). *Dialogue Modes in Expert Tutoring*. 470–479.
- Candiotto, L. (2016). Extended Affectivity as the Cognition of Primary Intersubjectivity. *Phenomenology and Mind*, 11, 232–241. https://doi.org/10.13128/Phe_Mi-20122
- Carbonell, J. R. (1970). AI in CAI: An Artificial-Intelligence Approach to Computer-Assisted Instruction. *IEEE Transactions on Man-Machine Systems*, 11(4), 190–202. <https://doi.org/10.1109/TMMS.1970.299942>
- Cash, M. (2013). Cognition without borders: “Third wave” socially distributed cognition and relational autonomy. *Cognitive Systems Research*, 25–26, 61–71. <https://doi.org/10.1016/j.cogsys.2013.03.007>
- Cerulo, K. A. (2009). Nonhumans in Social Interaction. *Annual Review of Sociology*, 35(1), 531–552. <https://doi.org/10.1146/annurev-soc-070308-120008>
- Cerulo, K. A. (2011). Social Interaction: Do Non-humans Count? *Sociology Compass*, 5(9), 775–791. <https://doi.org/10.1111/j.1751-9020.2011.00404.x>
- Chaminade, T., Rosset, D., Da Fonseca, D., Nazarian, B., Lutchter, E., Cheng, G., & Deruelle, C. (2012). How do we think machines think? An fMRI study of alleged competition with an artificial intelligence. *Frontiers in Human Neuroscience*, 6, 103. <https://doi.org/10.3389/fnhum.2012.00103>
- Charmaz, K. (2001a). Grounded Theory: Methodology and Theory Construction. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 6396–6399). Elsevier. <https://doi.org/10.1016/B0-08-043076-7/00775-0>
- Charmaz, K. (2006). *Constructing Grounded Theory. A Practical Guide through Qualitative Analysis*. SAGE Publications Ltd.

- Charmaz, K. (2008). Constructionism and the Grounded Theory Method. In J. A. Holstein & J. F. Gubrium (Eds.), *Handbook of constructionist research* (pp. 397–412). The Guilford Press.
- Chemero, A. (2013). Radical embodied cognitive science. *Review of General Psychology*, 17(2), 145–150.
- Chipman, S. E. F. (2016). An Introduction to Cognitive Science. In S. E. F. Chipman (Ed.), *The Oxford Handbook of Cognitive Science* (Vol. 1, Issue September). Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780199842193.013.21>
- Cho, A. (2015). Queer Reverb: Timblr, Affect, Time. In K. Hillis, S. Paasonen, & M. Petit (Eds.), *Networked Affect* (pp. 43–57). MIT Press.
- Clark, A. (1997a). *Being there. Putting brain, body, and world together again*. MIT Press.
- Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Sciences*, 3(9), 345–351.
- Clark, A. (2010a). Coupling, Constitution, and the Cognitive Kind: A Reply to Adams and Aizawa. In R. Menary (Ed.), *The Extended Mind* (pp. 81–99). The MIT Press.
<https://doi.org/10.7551/mitpress/9780262014038.003.0005>
- Clark, A. (2010b). Memento's Revenge: The Extended Mind, Extended. In *The Extended Mind* (pp. 43–66). The MIT Press. <https://doi.org/10.7551/mitpress/9780262014038.003.0003>
- Clark, A., & Chalmers, D. (1998a). The extended mind. *Analysis*, 58(1), 7–19.
- Clark, A., & Chalmers, D. J. (2010a). The Extended Mind. In R. Menary (Ed.), *The Extended Mind* (pp. 26–42). The MIT Press. <https://doi.org/10.7551/mitpress/9780262014038.003.0002>
- Clarke, A. E. (2003). Situational analyses: Grounded theory mapping after the postmodern turn. In *Symbolic Interaction* (Vol. 26, Issue 4, pp. 553–576). <https://doi.org/10.1525/si.2003.26.4.553>
- Clarke, A. E. (2005). *Situational Analysis: Grounded Theory after the Postmodern Turn*. SAGE Publications, Inc.
- Clarke, A. E. (2015). From Grounded Theory to Situational Analysis. What's new? Why? How? In A. E. Clarke, C. Friese, & R. S. Washburn (Eds.), *Situational Analysis in Practice. Mapping Research with Grounded Theory* (pp. 84–118). Left Coast Press.
- Clarke, A. E. (2018). *Situational analysis: Grounded theory after the interpretive turn* (Second edi). SAGE.
- Clarke, A. E., Friese, C., & Washburn, R. S. (2015). Introducing Situational Analysis. In A. E. Clarke, C. Friese, & R. S. Washburn (Eds.), *Situational Analysis in Practice. Mapping Research with Grounded Theory* (pp. 11–75). Left Coast Press.
- Colombetti, G. (2014). *The Feeling Body. Affective Science Meets the Enactive Mind*. The MIT Press.
<https://doi.org/10.7551/mitpress/9780262019958.001.0001>
- Colombetti, G. (2017). Enactive Affectivity, Extended. *Topoi*, 36(3), 445–455.
<https://doi.org/10.1007/s11245-015-9335-2>
- Colombetti, G. (2018a). Enacting affectivity. In A. Newen, L. De Bruin, & S. Gallagher (Eds.), *The Oxford Handbook of 4E Cognition*. Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780198735410.013.31>
- Corbett, A., McLaughlin, M., & Scarpinato, K. C. (2000). Modeling student knowledge: Cognitive tutors in high school and college. *User Modelling and User-Adapted Interaction*, 10(2–3), 81–108.
<https://doi.org/10.1023/A:1026505626690>
- Correll, S. J., & Ridgeway, C. L. (2006). Expectation States Theory. In J. DeLamater (Ed.), *Handbook of Social Psychology* (pp. 29–51). Springer US. https://doi.org/10.1007/0-387-36921-X_2
- Cottrell, G. W. (1991). Extracting features from faces using compression networks: Face, identity, emotion, and gender recognition using holons. *Connectionist Models*, 328–337.
<https://doi.org/10.1016/B978-1-4832-1448-1.50039-1>
- Cowie, R., Pelachaud, C., & Petta, P. (Eds.). (2011). *Emotion-Oriented Systems: The Humaine Handbook*. Springer. <https://doi.org/10.1007/978-3-642-15184-2>

- Cromby, J., & Willis, M. E. H. (2016). Affect—Or feeling (after Leys). *Theory & Psychology*, 26(4), 476–495. <https://doi.org/10.1177/0959354316651344>
- Cvetkovich, A. (2012). *Depression: A Public Feeling*. Duke University Press.
- Dautenhahn, K. (2011). THE ART OF DESIGNING SOCIALLY INTELLIGENT AGENTS: SCIENCE, FICTION, AND THE HUMAN IN THE LOOP. <https://doi.org/10.1080/088395198117550>, 12(7–8), 573–617. <https://doi.org/10.1080/088395198117550>
- Dautenhahn, K., Bond, A., Cañamero, L., & Edmonds, B. (2002). Socially Intelligent Agents. In K. Dautenhahn, A. Bond, L. Cañamero, & B. Edmonds (Eds.), *Socially Intelligent Agents: Creating Relationships with Computers and Robots* (pp. 1–20). Springer. https://doi.org/10.1007/0-306-47373-9_1
- Dautenhahn, K., Ogden, B., & Quick, T. (2002). From embodied to socially embedded agents – Implications for interaction-aware robots. *Cognitive Systems Research*, 3(3), 397–428. [https://doi.org/10.1016/S1389-0417\(02\)00050-5](https://doi.org/10.1016/S1389-0417(02)00050-5)
- de Bruin, L., van Elk, M., & Newen, A. (2012). Reconceptualizing second-person interaction. *Frontiers in Human Neuroscience*, 6, 151. <https://doi.org/10.3389/fnhum.2012.00151>
- De Jaegher, H. (2013). Rigid and fluid interactions with institutions. *Cognitive Systems Research*, 25–26, 19–25. <https://doi.org/10.1016/j.cogsys.2013.03.002>
- De Jaegher, H. (2018). The intersubjective turn. In A. Newen, L. De Bruin, & S. Gallagher (Eds.), *The Oxford Handbook of 4E Cognition* (pp. 1–19). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198735410.013.24>
- De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making. An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, 6(4), 485–507. <https://doi.org/10.1007/s11097-007-9076-9>
- De Jaegher, H., & Di Paolo, E. (2008). Making sense in participation: An enactive approach to social cognition. In F. Morganti, A. Carassa, & G. Riva (Eds.), *Enacting Intersubjectivity. A Cognitive and Social Perspective on the Study of Interaction* (pp. 33–47). IOS Press.
- De Jaegher, H., & Di Paolo, E. (2013). Enactivism is not interactionism. *Frontiers in Human Neuroscience*, 6, 345. <https://doi.org/10.3389/fnhum.2012.00345>
- De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010a). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, 14(10), 441–447. <https://doi.org/10.1016/j.tics.2010.06.009>
- De Jesus, P. (2016). Autopoietic enactivism, phenomenology and the deep continuity between life and mind. *Phenomenology and the Cognitive Sciences*, 15(2), 265–289. <https://doi.org/10.1007/s11097-015-9414-2>
- De Jesus, P. (2018). Thinking through enactive agency: Sense-making, bio-semiosis and the ontologies of organismic worlds. *Phenomenology and the Cognitive Sciences*, 17(5), 861–887. <https://doi.org/10.1007/s11097-018-9562-2>
- Dennett, D. C. (1988). Précis of The Intentional Stance. *Behavioral and Brain Sciences*, 11(03), 495. <https://doi.org/10.1017/S0140525X00058611>
- Di Paolo, E. (2005). Autopoiesis, Adaptivity, Teleology, Agency. *Phenomenology and the Cognitive Sciences*, 4(4), 429–452. <https://doi.org/10.1007/s11097-005-9002-y>
- Di Paolo, E. A., Rohde, M., & De Jaegher, H. (2013). Horizons for the Enactive Mind: Values, Social Interaction, and Play. In J. Stewart, O. Gapenne, & E. A. Di Paolo (Eds.), *Enaction* (pp. 32–87). MIT Press Scholarship Online. <https://doi.org/10.7551/mitpress/9780262014601.003.0003>
- Di Paolo, E. A., Rohde, M., & Iizuka, H. (2008). Sensitivity to social contingency or stability of interaction? Modelling the dynamics of perceptual crossing. *New Ideas in Psychology*, 26(2), 278–294. <https://doi.org/10.1016/j.newideapsych.2007.07.006>

- Di Paolo, E., & De Jaegher, H. (2017). Neither Individualistic nor Interactionist. In C. Durt, T. Fuchs, & C. Tewes (Eds.), *Embodiment, Enaction, and Culture* (pp. 87–105). The MIT Press.
<https://doi.org/10.7551/mitpress/9780262035552.003.0005>
- Dillenbourg, P., & Self, J. a. (1992). A computational approach to socially distributed cognition. *European Journal of Psychology of Education*, 7(4), 353–372. <https://doi.org/10.1007/BF03172899>
- D’mello, S., & Graesser, A. (2012). AutoTutor and Affective AutoTutor: Learning by Talking with Cognitively and Emotionally Intelligent Computers that Talk Back. *ACM Transactions on Interactive Intelligent Systems*, 2(4), 1–39. <https://doi.org/10.1145/2395123.2395128>
- D’Mello, S., Hays, P., Williams, C., Cade, W., Brown, J., & Olney, A. (2010). Collaborative Lecturing by Human and Computer Tutors. *Lecture Notes in Computer Science*, 6095, 178–187.
- D’Mello, S. K., Lehman, B., & Graesser, A. (2011). A Motivationally Supportive Affect-Sensitive AutoTutor. In *New Perspectives on Affect and Learning Technologies*.
https://doi.org/10.1007/978-1-4419-9625-1_9
- Dominey, P. F., Prescott, T. J., Bohg, J., Engel, A. K., Gallagher, S., Heed, T., Hoffmann, M., Knoblich, G., Prinz, W., & Schwartz, A. (2016a). Implications of Action-Oriented Paradigm Shifts in Cognitive Science. In A. K. Engel, K. J. Friston, & D. Kragic (Eds.), *The Pragmatic Turn* (pp. 333–356). The MIT Press. <https://doi.org/10.7551/mitpress/9780262034326.003.0020>
- Dreyfus, H. (2002a). Intelligence without representation – Merleau-Ponty’s critique of mental representation The relevance of phenomenology to scientific explanation. *Phenomenology and the Cognitive Sciences*, 1(4), 367–383. <https://doi.org/10.1023/A:1021351606209>
- Dreyfus, H. (2002b). Refocusing the question: Can there be skillful coping without propositional representations or brain representations? *Phenomenology and the Cognitive Sciences*, 1(4), 413–425. <https://doi.org/10.1023/A:1021303723047>
- Dreyfus, H. L. (1992). *What computers still can’t do: A critique of artificial reason*. MIT Press.
- Dreyfus, W. (1995). *Informatica*. 19.
- Dumas, G., Lachat, F., Martinerie, J., Nadel, J., & George, N. (2011). From social behaviour to brain synchronization: Review and perspectives in hyperscanning. *IRBM*, 32(1), 48–53.
<https://doi.org/10.1016/j.irbm.2011.01.002>
- Edmonds, B. (1999). Capturing Social Embeddedness: A Constructivist Approach. *Adaptive Behavior*, 7(3–4), 323–347. <https://doi.org/10.1177/105971239900700307>
- Ekman, P. (2005). Basic Emotions. In *Handbook of Cognition and Emotion* (Vol. 15, pp. 45–60). John Wiley & Sons, Ltd. <https://doi.org/10.1002/0470013494.ch3>
- Engel, A. K., Friston, K. J., & Kragic, D. (2016). Introduction: Where’s the Action? In A. K. Engel, K. J. Friston, & D. Kragic (Eds.), *The Pragmatic Turn* (pp. 1–16). The MIT Press.
<https://doi.org/10.7551/mitpress/9780262034326.003.0001>
- Eugenio, B. Di, Kershaw, T. C., Lu, X., Corrigan-Halpern, A., & Ohlsson, S. (2006). Toward a Computational Model of Expert Tutoring: A First Report. *FLAIRS Conference*, 503–508.
- Flemmen, A. B. (2017). Sensitizing Concepts in Action: Expanding the Framework. In H. Leiufrud & P. Sohlberg (Eds.), *Concepts in Action: Conceptual Constructivism. Studies in Critical Social Sciences, vol. 118* (pp. 79–94). Brill. https://doi.org/10.1163/9789004314207_006
- Fodor, J. A. (1983). *The Modularity of Mind*. MIT Press.
- Fogg, B. J., & Nass, C. (1997). Silicon sycophants: The effects of computers that flatter. In *International Journal of Human-Computer Studies* (Vol. 46, Issue 5, pp. 551–561). Elsevier Science.
<https://doi.org/10.1006/ijhc.1996.0104>
- Formosa, P. (2021). Robot Autonomy vs. Human Autonomy: Social Robots, Artificial Intelligence (AI), and the Nature of Autonomy. *Minds and Machines*, 0123456789. <https://doi.org/10.1007/s11023-021-09579-2>

- Forstmann, B. U., Jahfari, S., Scholte, H. S., Wolfensteller, U., Wildenberg, W. P. M. van den, & Ridderinkhof, K. R. (2008). Function and Structure of the Right Inferior Frontal Cortex Predict Individual Differences in Response Inhibition: A Model-Based Approach. *Journal of Neuroscience*, 28(39), 9790–9796. <https://doi.org/10.1523/JNEUROSCI.1465-08.2008>
- Friese, S. (2016). Computer-Assisted Grounded Theory Analysis With ATLAS.ti. *ATLAS.Ti User Conference 2015 – Qualitative Data Analysis and Beyond*, 1–22. <https://doi.org/10.14279/depositonce-5155>
- Friese, S. (2019a). Grounded Theory Analysis and CAQDAS: A Happy Pairing or Remodeling GT to QDA? In A. Bryant & K. Charmaz (Eds.), *The SAGE Handbook of Current Developments in Grounded Theory* (pp. 282–313). SAGE Publications Ltd. <https://doi.org/10.4135/9781526485656.n16>
- Friese, S. (2019b). *Qualitative Data Analysis with ATLAS.ti*. Sage.
- Frijda, N. H. (2008). The psychologists' point of view. In M. Lewis, J. M. Haviland-Jones, & L. Feldman Barrett (Eds.), *Handbook of emotions, 3rd ed.* (pp. 68–87). The Guilford Press.
- Frith, U., & Blakemore, S.-J. (2006). *Social Cognition* (Foresight Cognitive Systems Project Research Review). Foresight Cognitive Systems Project, Foresight Directorate.
- Froese, T., & Gallagher, S. (2012). Getting interaction theory (IT) together: Integrating developmental, phenomenological, enactive, and dynamical approaches to social interaction. *Interaction Studies*, 13(3), 436–468. <https://doi.org/10.1075/is.13.3.06fro>
- Froese, T., Iizuka, H., & Ikegami, T. (2015). Embodied social interaction constitutes social cognition in pairs of humans: A minimalist virtual reality experiment. *Scientific Reports*, 4(1), 3672. <https://doi.org/10.1038/srep03672>
- Froese, T., Lenay, C., & Ikegami, T. (2012). Imitation by social interaction? Analysis of a minimal agent-based model of the correspondence problem. *Frontiers in Human Neuroscience*, 6(JUNE 2012), 202. <https://doi.org/10.3389/fnhum.2012.00202>
- Froese, T., & Ziemke, T. (2009). Enactive artificial intelligence: Investigating the systemic organization of life and mind. *Artificial Intelligence*, 173(3–4), 466–500. <https://doi.org/10.1016/j.artint.2008.12.001>
- Fuchs, T., & de Jaegher, H. (2009). Enactive intersubjectivity: Participatory sense-making and mutual incorporation. *Phenomenology and the Cognitive Sciences* 8:4, 8(4), 465–486. <https://doi.org/10.1007/S11097-009-9136-4>
- Fusella, P. V. (2013). Dynamic systems theory in cognitive science: Major elements, applications, and debates surrounding a revolutionary meta-theory. *Dynamical Psychology*.
- Gallagher, H. L., Jack, A. I., Roepstorff, A., & Frith, C. D. (2002). Imaging the intentional stance in a competitive game. *NeuroImage*, 16(3 Pt 1), 814–821. <https://doi.org/10.1006/nimg.2002.1117>
- Gallagher, S. (2001). The practice of mind. Theory, simulation or primary interaction? *Journal of Consciousness Studies*, 8(5–7), 83–108.
- Gallagher, S. (2005). *How the Body Shapes the Mind*. Oxford University Press. <https://doi.org/10.1093/0199271941.001.0001>
- Gallagher, S. (2006). The narrative alternative to theory of mind. In R. Menary (Ed.), *Radical Enactivism. Intentionality, Phenomenology and Narrative. Focus on the philosophy of Daniel D. Hutto* (pp. 223–229). John Benjamins Publishing Co. <https://doi.org/10.1075/CEB.2.15GAL>
- Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition*, 17(2), 535–543. <https://doi.org/10.1016/j.concog.2008.03.003>
- Gallagher, S. (2011a). Interpretations of embodied cognition. In W. Tschacher & C. Bergomi (Eds.), *The implications of embodiment: Cognition and communication*. (Issue Id, pp. 59–71). Imprint Academi (print), Andrews UK (digital).

- Gallagher, S. (2011b). Strong interaction and self-agency. *Humana Mente*, 15, 55–76.
- Gallagher, S. (2013). The socially extended mind. *Cognitive Systems Research*, 25–26, 4–12.
<https://doi.org/10.1016/j.cogsys.2013.03.008>
- Gallagher, S. (2017). *Enactivist Interventions: Rethinking the Mind* (Intergovernmental Panel on Climate Change, Ed.; Vol. 1). Oxford University Press.
<https://doi.org/10.1093/oso/9780198794325.001.0001>
- Gallagher, S. (2018a). Decentering the brain: Embodied cognition and the critique of neurocentrism and narrow-minded philosophy of mind. *Constructivist Foundations*, 14(1), 8–21.
- Gallagher, S. (2018b). New Mechanisms and the Enactivist Concept of Constitution. In M. P. Guta (Ed.), *Consciousness and the Ontology of Properties* (pp. 207–220). Routledge.
<https://doi.org/10.4324/9781315104706-13>
- Gallagher, S. (2019). Precis: Enactivist Interventions. In *Philosophical Studies* (Vol. 176, Issue 3, pp. 803–806). Springer Netherlands. <https://doi.org/10.1007/s11098-018-01230-8>
- Gallagher, S. (2020). *Action and Interaction*. Oxford University Press.
<https://doi.org/10.1093/oso/9780198846345.001.0001>
- Gallagher, S., & Hutto, D. D. (2008). Understanding others through primary interaction and narrative practice. In J. Zlatev, T. P. Racine, C. Sinha, & E. Itkonen (Eds.), *Converging evidence in language and communication research (CELCR) Vol. 12. The shared mind: Perspectives on intersubjectivity* (pp. 17–38). John Benjamins Publishing Company.
<https://doi.org/10.1075/celcr.12.04gal>
- Gallagher, S., & Zahavi, D. (2008). *The Phenomenological Mind* (1st ed.). Routledge.
- Gambino, A., Fox, J., & Ratan, R. (2020). Building a Stronger CASA: Extending the Computers Are Social Actors Paradigm. *Human-Machine Communication*, 1, 71–86.
<https://doi.org/10.30658/hmc.1.5>
- Gardner, H. (1988). *The mind's new science. A history of the cognitive revolution*. Basic Books.
- Gibson, J. J. (1986a). *The ecological approach to visual perception*. Psychology Press. Taylor and Francis Group.
- Gibson, J. J. (1986b). The Theory of Affordances. In *The ecological approach to visual perception* (pp. 127–141). Lawrence Erlbaum Associates.
- Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. University of California Press.
- Giddens, A. (2006). *Sociology*. Polity Press.
- Giere, R. (2002). Scientific cognition as distributed cognition. In *The Cognitive Basis of Science* (pp. 285–299). Cambridge University Press. <https://doi.org/10.1017/CBO9780511613517.016>
- Gilson, L. L., & Goldberg, C. B. (2015). Editors' Comment: So, What Is a Conceptual Paper? *Group & Organization Management*, 40(2), 127–130. <https://doi.org/10.1177/1059601115576425>
- Glaser, B. G. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory*. The Sociology Press.
- Glaser, B. G. (1998). *Doing grounded theory: Issues and discussions*. Sociology Press.
- Glaser, B. G., & Strauss, A. L. (1968). *The Discovery of grounded theory: Strategies for qualitative research* (1. publ. i). Weidenfeld & Nicolson.
- Goffman, E. (1956). *The presentation of Self in Everyday Life*. University of Edinburgh Social Science Research Centre.
- Göhler, G. (2009). 'Power to' and 'Power over.' In S. R. Clegg & M. Haugaard (Eds.), *The SAGE Handbook of Power* (pp. 27–39). SAGE Publications Ltd.
<https://doi.org/10.4135/9780857021014.n1>

- Goldstein, M., Alsiö, G., & Werdenhoff, J. (2002). The media equation does not always apply: People are not polite towards small computers. *Personal and Ubiquitous Computing*, 6(2), 87–96. <https://doi.org/10.1007/s007790200008>
- Gorra, A. (2019). Keep your Data Moving: Operationalization of Abduction with Technology. In A. Bryant & K. Charmaz (Eds.), *The SAGE Handbook of Current Developments in Grounded Theory* (pp. 314–331). SAGE Publications Ltd. <https://doi.org/10.4135/9781526485656.n17>
- Graesser, A. C. (2016a). Conversations with AutoTutor Help Students Learn. *International Journal of Artificial Intelligence in Education*, 26(1), 124–132. <https://doi.org/10.1007/s40593-015-0086-4>
- Graesser, A. C., Baer, W., Feng, S., Walker, B., Clewley, D., Hays, D. P., & Greenberg, D. (2016a). Emotions in Adaptive Computer Technologies for Adults Improving Reading. In S. Y. Tettegah & M. Gartmeier (Eds.), *Emotions, Technology, Design, and Learning* (pp. 3–25). Elsevier. <https://doi.org/10.1016/B978-0-12-801856-9.00001-3>
- Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A. (2005). AutoTutor: An Intelligent Tutoring System With Mixed-Initiative Dialogue. *IEEE Transactions on Education*, 48(4), 612–618. <https://doi.org/10.1109/TE.2005.856149>
- Graesser, A. C., Conley, M. W., & Olney, A. (2012a). Intelligent tutoring systems. In *APA educational psychology handbook, Vol 3: Application to learning and teaching*. (Vol. 3, pp. 451–473). American Psychological Association. <https://doi.org/10.1037/13275-018>
- Graesser, A. C., D’Mello, S., Hu, X., Cai, Z., Olney, A., & Morgan, B. (2012). AutoTutor. In *Applied Natural Language Processing* (Issue September 2019, pp. 169–187). IGI Global. <https://doi.org/10.4018/978-1-60960-741-8.ch010>
- Graesser, A. C., VanLehn, K., Rosé, C. P., Jordan, P. W., & Harter, D. (2001). Intelligent Tutoring Systems with Conversational Dialogue. *AI Magazine*, 22(4), 39–52. <http://dx.doi.org/10.1609/aimag.v22i4.1591>
- Graesser, Arthur. C., Li, H., & Forsyth, C. (2014). Learning by Communicating in Natural Language With Conversational Agents. *Current Directions in Psychological Science*, 23(5), 374–380. <https://doi.org/10.1177/0963721414540680>
- Guinote, A. (2008). Power and affordances: When the situation has more power over powerful than powerless individuals. *Journal of Personality and Social Psychology*, 95(2), 237–252. <https://doi.org/10.1037/a0012518>
- Guinote, A. (2013). *Social Power and Cognition* (D. E. Carlston, Ed.). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199730018.013.0028>
- Gulz, A., & Haake, M. (2006). Design of animated pedagogical agents—A look at their look. *International Journal of Human Computer Studies*, 64(4), 322–339. <https://doi.org/10.1016/j.ijhcs.2005.08.006>
- Gulz, A., Haake, M., Silvervarg, A., Sjoden, B., & Veletsianos, G. (2011). Building a Social Conversational Agent: Design Challenges and Methodological Approaches. In D. Perez-Martin & I. Pascua-Nieto (Eds.), *Conversational Agents and Natural Language Interaction: Techniques and Effective* (pp. 128–155). IGI Global. <https://doi.org/10.4018/978-1-60960-617-6.ch006>
- Haake, M. (2009). *Embodied Pedagogical Agents. From Visual Impact to Pedagogical Implications*. Lund University, Sweden.
- Haake, M., & Gulz, A. (2006). Visual Stereotypes and Virtual Pedagogical Agents. *Educational Technology & Society*, 11(4), 1–15.
- Hagen, R. (2017). Abduction – Assessing Fruitfulness and the Construction of Scientific Concepts. In P. Sohlberg & H. Leilufsrud (Eds.), *Theory in Action* (pp. 126–144). BRILL. https://doi.org/10.1163/9789004314184_009
- Hamilton, K. A., & Yao, M. Z. (2018). *Cognitive Offloading and the Extended Digital Self* (pp. 257–268). https://doi.org/10.1007/978-3-319-91238-7_22

- Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1–3), 335–346. [https://doi.org/10.1016/0167-2789\(90\)90087-6](https://doi.org/10.1016/0167-2789(90)90087-6)
- Hartner-Tiefenthaler, M., Roetzer, K., Bottaro, G., & Peschl, F.-M. (2018). When relational and epistemological uncertainty act as driving forces in collaborative knowledge creation processes among university students When relational and epistemological uncertainty act as driving forces in collaborative knowledge creation process. *Thinking Skills and Creativity*, 28(February), 21–40. <https://doi.org/10.1016/j.tsc.2018.02.013>
- Hashemian, M., Prada, R., Santos, P. A., & Mascarenhas, S. (2018). Enhancing Social Believability of Virtual Agents using Social Power Dynamics. *Proceedings of the 18th International Conference on Intelligent Virtual Agents - IVA '18*, 147–152. <https://doi.org/10.1145/3267851.3267902>
- Haugaard, M. (2003). Reflections on Seven Ways of Creating Power. *European Journal of Social Theory*, 6(1), 87–113. <https://doi.org/10.1177/1368431003006001562>
- Haugaard, M. (2014). Power. In M. T. Gibbons (Ed.), *The Encyclopedia of Political Thought* (pp. 2965–2978). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118474396.wbept0820>
- Haugaard, M., & Clegg, S. R. (2009). Introduction: Why Power is the Central Concept of the Social Sciences. In S. R. Clegg & M. Haugaard (Eds.), *The SAGE Handbook of Power* (pp. 1–24). SAGE Publications Ltd. <https://doi.org/10.4135/9780857021014>
- Heaney, J. G. (2011). Emotions and power: Reconciling conceptual twins. *Journal of Political Power*, 4(2), 259–277. <https://doi.org/10.1080/2158379X.2011.591171>
- Heaney, J. G. (2013). Emotions and power: A bifocal prescription to cure theoretical myopia. *Journal of Political Power*, 6(3), 355–362. <https://doi.org/10.1080/2158379x.2013.849367>
- Heaney, J. G. (2019). Emotion as power: Capital and strategy in the field of politics. *Journal of Political Power*, 00(00), 1–21. <https://doi.org/10.1080/2158379X.2019.1618485>
- Hemmings, C. (2005). INVOKING AFFECT. Cultural theory and the ontological turn. *Cultural Studies*, 19(5), 548–567. <https://doi.org/10.1080/09502380500365473>
- Hernández, Y., Arroyo-Figueroa, G., & Sucar, L. E. (2015). A model of affect and learning for intelligent tutors. *Journal of Universal Computer Science*, 21(7), 912–934.
- Herschbach, M. (2012). On the role of social interaction in social cognition: A mechanistic alternative to enactivism. *Phenomenology and the Cognitive Sciences*, 11(4), 467–486. <https://doi.org/10.1007/s11097-011-9209-z>
- Higgins, J. (2018). Biosocial selfhood: Overcoming the ‘body-social problem’ within the individuation of the human self. *Phenomenology and the Cognitive Sciences*, 17(3), 433–454. <https://doi.org/10.1007/s11097-017-9514-2>
- Hoffmann, L., Krämer, N. C., Lam-Chi, A., & Kopp, S. (2009). Media equation revisited: Do users show polite reactions towards an embodied agent? *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5773 LNAI, 159–165. https://doi.org/10.1007/978-3-642-04380-2_19
- Holdnack, J. A. (2019). 4—The development, expansion, and future of the WAIS-IV as a cornerstone in comprehensive cognitive assessments. In G. Goldstein, D. N. Allen, & J. DeLuca (Eds.), *Handbook of Psychological Assessment (Fourth Edition)* (pp. 103–139). Academic Press. <https://doi.org/10.1016/B978-0-12-802203-0.00004-3>
- Hutchins, E. (1995a). *Cognition in the Wild*. The MIT Press.
- Hutchins, E. (1995b). How a Cockpit Remembers Its Speeds . *Cognitive Science*, 19(3), 265–288.
- Hutchins, E. (2000). *Distributed Cognition*. IESBS University of California.
- Hutchins, E. (2001a). Cognition, Distributed. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 2068–2072). Elsevier. <https://doi.org/10.1016/B0-08-043076-7/01636-3>

- Hutchins, E. (2006). The distributed cognition perspective on human interaction. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of Human Sociality: Culture, Cognition and Interaction* (pp. 375–398). Berg Publishers.
- Hutto, D. D. (2007). The Narrative Practice Hypothesis: Origins and Applications of Folk Psychology. *Royal Institute of Philosophy Supplement*, 60, 43–68.
<https://doi.org/10.1017/S1358246100009607>
- Hutto, D. D., & Myin, E. (2012). *Radicalizing Enactivism: Basic Minds without Content*. The MIT Press.
<https://doi.org/10.7551/mitpress/9780262018548.001.0001>
- Johnson, W. L., & Lester, J. C. (2016). Face-to-Face Interaction with Pedagogical Agents, Twenty Years Later. *International Journal of Artificial Intelligence in Education*, 26(1), 25–36.
<https://doi.org/10.1007/s40593-015-0065-9>
- Jonas, H. (1966). *The phenomenon of life: Toward a philosophical biology*. Harper & Row.
<http://catalog.hathitrust.org/api/volumes/oclc/1422747.html>
- Kemper, T. D. (1991a). Predicting Emotions from Social Relations. *Social Psychology Quarterly*, 54(4), 330–342.
- Kemper, T. D. (2006a). Power and Status in the Power-Status Theory of Emotions. In J. E. Stets & J. H. Turner (Eds.), *Handbook of the sociology of emotions* (pp. 87–113). Springer.
- Kemper, T. D. (2011). *Status, Power and Ritual Interaction. A Relational Reading of Durkheim. Goffman and Collins*. Ashgate Publishing Limited. <http://www.ashgate.com/isbn/9781409427360>
- Kemper, T. D., & Collins, R. (1990). Dimensions of Microinteraction. *American Journal of Sociology*, 96(1), 32. <https://doi.org/10.1086/229492>
- Kim, D., & Lim, Y. (2019). Co-Performing Agent. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–14. <https://doi.org/10.1145/3290605.3300714>
- Kim, Y. (2007). Desirable characteristics of learning companions. *International Journal of Artificial Intelligence in Education*, 17(4), 371–388.
- Kim, Y., Baylor, a. L., & Shen, E. (2007). Pedagogical agents as learning companions: The impact of agent emotion and gender. *Journal of Computer Assisted Learning*, 23(3), 220–234.
<https://doi.org/10.1111/j.1365-2729.2006.00210.x>
- Kim, Y., & Baylor, A. L. (2006a). A Social Cognitive Framework for Designing Pedagogical Agents as Learning Companions. *Educational Technology Research and Development*, 54(6), 569–596.
- Kim, Y., & Baylor, A. L. (2006b). Pedagogical agents as learning companions: The role of agent competency and type of interaction. *Educational Technology Research and Development*, 54(3), 223–243. <https://doi.org/10.1007/s11423-006-8805-z>
- Kim, Y., & Baylor, A. L. (2016). Research-Based Design of Pedagogical Agent Roles: A Review, Progress, and Recommendations. *International Journal of Artificial Intelligence in Education*, 26(1), 160–169. <https://doi.org/10.1007/s40593-015-0055-y>
- Kim, Y., Smith, D., & Thayne, J. (2016). Designing Tools that Care: The Affective Qualities of Virtual Peers, Robots, and Videos. In M. Tettegah, Sharon Y. Gartmeier (Ed.), *Emotions, Technology, Design, and Learning* (pp. 115–129). Elsevier. <https://doi.org/10.1016/B978-0-12-801856-9.00006-2>
- Kiverstein, J. (2012). The Meaning of Embodiment. *Topics in Cognitive Science*, 4(4), 740–758.
<https://doi.org/10.1111/j.1756-8765.2012.01219.x>
- Konvalinka, I., & Roepstorff, A. (2012). The two-brain approach: How can mutually interacting brains teach us something about social interaction? *Frontiers in Human Neuroscience*, 6, 215.
<https://doi.org/10.3389/fnhum.2012.00215>
- Kopp, S., Gesellensetter, L., Krämer, N. C., & Wachsmuth, I. (2005). A Conversational Agent as Museum Guide – Design and Evaluation of a Real-World Application. In T. Panayiotopoulos, J. Gratch, R.

- Aylett, D. Ballin, P. Olivier, & T. Rist (Eds.), *Intelligent Virtual Agents* (Vol. 3661, pp. 329–343). Springer Berlin Heidelberg. https://doi.org/10.1007/11550617_28
- Kotseruba, I., & Tsotsos, J. K. (2020). 40 years of cognitive architectures: Core cognitive abilities and practical applications. *Artificial Intelligence Review*, *53*(1), 17–94. <https://doi.org/10.1007/s10462-018-9646-y>
- Krämer, N., Kopp, S., Becker-Asano, C., & Sommer, N. (2013). Smile and the world will smile with you —The effects of a virtual agent’s smile on users’ evaluation and behavior. *International Journal of Human Computer Studies*, *71*(3). <https://doi.org/10.1016/j.ijhcs.2012.09.006>
- Krichmar, J. L., & Edelman, G. M. (2005). Brain-based devices for the study of nervous systems and the development of intelligent machines. *Artificial Life*, *11*(1–2), 63–77. <https://doi.org/10.1162/1064546053278946>
- Krueger, J. (2014a). Varieties of extended emotions. *Phenomenology and the Cognitive Sciences*, *13*(4), 533–555. <https://doi.org/10.1007/s11097-014-9363-1>
- Kulms, P., & Kopp, S. (2018). A Social Cognition Perspective on Human–Computer Trust: The Effect of Perceived Warmth and Competence on Trust in Decision-Making With Computers. *Frontiers in Digital Humanities*, *5*. <https://www.frontiersin.org/article/10.3389/fdigh.2018.00014>
- Kyselo, M. (2013a). Enaktivismus. In A. Stephan & S. Walter (Eds.), *Handbuch Kognitionswissenschaft* (pp. 197–201). J.B. Metzler.
- Kyselo, M. (2014). The body social: An enactive approach to the self. *Frontiers in Psychology*, *5*(SEP). <https://doi.org/10.3389/fpsyg.2014.00986>
- Lamb, M., & Chemero, A. (2018). Interacting in the Open: Where Dynamical Systems Become Extended and Embodied. In A. Newen, L. De Bruin, & S. Gallagher (Eds.), *The Oxford Handbook of 4E Cognition*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198735410.013.8>
- Lawrence, S. (2010). James bond and the barking dog: Evolution and extended cognition. *Philosophy of Science*, *77*(3), 400–418. <https://doi.org/10.1086/652963>
- Lee, J. E. R., & Nass, C. I. (2010). Trust in computers: The computers-are-social-actors (CASA) paradigm and trustworthiness perception in human-computer communication. In *Trust and Technology in a Ubiquitous Modern Environment: Theoretical and Methodological Perspectives* (pp. 1–15). IGI Global. <https://doi.org/10.4018/978-1-61520-901-9.ch001>
- Lehman, B., D’Mello, S., Cade, W., & Person, N. (2012). How do they do it? Investigating dialogue moves within dialogue modes in expert human tutoring. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7315 LNCS*, 557–562. https://doi.org/10.1007/978-3-642-30950-2_72
- Lepper, M. R., & Woolverton, M. (2002). The Wisdom of Practice: Lessons Learned from the Study of Highly effective tutors. In J. Aronson (Ed.), *Improving Academic Achievements, Impact on Psychological Factors in Education* (pp. 135–158). Academic Press.
- Leys, R. (2011). The Turn to Affect: A Critique. *Critical Inquiry*, *37*(3), 434–472.
- Li, E. C.-Y. (2015). Affect and Sociology: Reflection and Exploration through a Study of Media and Gender in Urban China. *Graduate Journal of Social Science*, *11*(1), 15–37.
- Liew, T. W., Tan, S. M., & Jayothisa, C. (2013a). The effects of peer-like and expert-like pedagogical agents on learners’ agent perceptions, task-related attitudes, and learning achievement. *Educational Technology and Society*, *16*(4).
- Limerick, H., Coyle, D., & Moore, J. W. (2014). The experience of agency in human-computer interactions: A review. *Frontiers in Human Neuroscience*, *8*, 643. <https://doi.org/10.3389/fnhum.2014.00643>
- Lindblom, J. (2015a). *Embodied Social Cognition* (Vol. 26). Springer International Publishing. <https://doi.org/10.1007/978-3-319-20315-7>

- Lindblom, J. (2015b). Meaning-Making as a Socially Distributed and Embodied Practice. In A. Scarinzi (Ed.), *Aesthetics and the Embodied Mind: Beyond Art Theory and the Cartesian Mind-Body Dichotomy* (pp. 3–19). Springer, Dordrecht. https://doi.org/10.1007/978-94-017-9379-7_1
- Lindblom, J. (2020). A Radical Reassessment of the Body in Social Cognition. *Frontiers in Psychology, 11*, 987. <https://doi.org/10.3389/fpsyg.2020.00987>
- Lindblom, J., & Ziemke, T. (2003). Social Situatedness of Natural and Artificial Intelligence: Vygotsky and Beyond. *Adaptive Behavior, 11*(2), 79–96. <https://doi.org/10.1177/10597123030112002>
- Lindblom, J., & Ziemke, T. (2005). The “body-in-motion” and social scaffolding: Implications for human and android cognitive development. *Toward Social Mechanisms of Android Science: A COGSCI 2005 Workshop, July 25-26*, 87–95.
- Lindblom, J., & Ziemke, T. (2008). Interacting socially through embodied action. In F. Morganti, A. Carassa, & G. Riva (Eds.), *Enacting Intersubjectivity. A Cognitive and Social Perspective on the Study of Interaction* (Vol. 10, pp. 49–63). IOS Press.
- Lindgren, S. (2007). *Digital Media & Society*. SAGE.
- Lu, X., Eugenio, B., Kershaw, T. C., Ohlsson, S., & Corrigan-Halpern, A. (2007). Expert vs. Non-expert Tutoring: Dialogue Moves, Interaction Patterns and Multi-utterance Turns. *Proceedings of the 8th International Conference on Computational Linguistics and Intelligent Text Processing*, 456–467. https://doi.org/10.1007/978-3-540-70939-8_40
- Lynch, M. F., Sun, R., & Wilson, N. (2011). CLARION as a Cognitive Framework for Intelligent Virtual Agents. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 6895 LNAI* (pp. 460–461). https://doi.org/10.1007/978-3-642-23974-8_62
- Lyre, H., & Walter, S. (2013). Situierte Kognition (situated cognition). In A. Stephan & S. Walter (Eds.), *Handbuch Kognitionswissenschaft* (pp. 184–185). J.B. Metzler.
- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology, 106*(4), 901–918. <https://doi.org/10.1037/a0037123>
- Marchant, J. L., & Frith, C. D. (2009). Social Cognition. In L. R. Squire (Ed.), *Encyclopedia of Neuroscience* (pp. 27–30). Elsevier. <https://doi.org/10.1016/B978-008045046-9.01540-0>
- Marchesi, S., Ghiglini, D., Ciardo, F., Perez-Osorio, J., Baykara, E., & Wykowska, A. (2019). Do We Adopt the Intentional Stance Toward Humanoid Robots? *Frontiers in Psychology, 10*, 450. <https://doi.org/10.3389/fpsyg.2019.00450>
- Mascarenhas, S., Prada, R., Paiva, A., & Hofstede, G. J. (2013). Social importance dynamics: A model for culturally-adaptive agents. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8108 LNAI*, 325–338. https://doi.org/10.1007/978-3-642-40415-3_29
- Massumi, B. (2002). *Parables for the Virtual: Movement, Affect, Sensation*. Durham University Press.
- Matsumoto, D., Keltner, D., Shiota, M. N., O’Sullivan, M., & Frank, M. (2008). Facial expressions of emotion. In M. Lewis, J. M. Haviland-Jones, & L. Feldman Barrett (Eds.), *Handbook of emotions, 3rd ed.* (pp. 211–234). The Guilford Press.
- Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and Cognition. The Realization of the Living*. <https://doi.org/EB N MATU>
- Maturana, H. R., & Varela, F. J. (1987). *The Tree of Knowledge: The Biological Roots of Human Understanding*. Shambhala Publication.
- McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences, 98*(20), 11832–11835. <https://doi.org/10.1073/PNAS.211415698>

- McCarthy, J., & Hayes, P. J. (1969). Some Philosophical Problems from the Standpoint of Artificial Intelligence. In B. Meltzer & D. Michie (Eds.), *Machine Intelligence 4* (pp. 463–502). Edinburgh University Press.
- McGann, M., & De Jaeger, H. (2009). Self–other contingencies: Enacting social perception. *Phenomenology and the Cognitive Sciences*, 8(4), 417–437. <https://doi.org/10.1007/s11097-009-9141-7>
- Mead, H. G. (1934). *Mind, Self, and Society*. University of Chicago Press.
- Menary, R. (2010). Cognitive Integration and the Extended Mind. In *The Extended Mind* (pp. 226–243). The MIT Press. <https://doi.org/10.7551/mitpress/9780262014038.003.0010>
- Menary, R. (2016). Pragmatism and the Pragmatic Turn in Cognitive Science. In A. K. Engel, K. J. Friston, & D. Kragic (Eds.), *The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science* (pp. 215–234). The MIT Press. <https://doi.org/10.7551/mitpress/9780262034326.003.0013>
- Merleau-Ponty, M. (1963). *The Structure of Behaviour* (A. L. Fisher, Trans.). Beacon Press.
- Merleau-Ponty, M. (2012). *Phenomenology of Perception* (D. A. Landes, Trans.). Taylor & Francis Group.
- Merleau-Ponty, M. (2013). *Phenomenology of Perception*. Routledge. <https://doi.org/10.4324/9780203720714>
- Merton, R. K. (1968). Social Theory and Social Structure. In *New York*. The Free Press. [https://edisciplinas.usp.br/pluginfile.php/4250035/mod_folder/content/0/Textos/Merton%2C Social Theory and Social Structure.pdf?forcedownload=1](https://edisciplinas.usp.br/pluginfile.php/4250035/mod_folder/content/0/Textos/Merton%2C%20Social%20Theory%20and%20Social%20Structure.pdf?forcedownload=1)
- Mesquita, B., & Frijda, N. H. (1992). Cultural variations in emotions: A review. *Psychological Bulletin*, 112(2), 179–204. <https://doi.org/10.1037/0033-2909.112.2.179>
- Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences*, 7(3), 141–144. [https://doi.org/10.1016/S1364-6613\(03\)00029-9](https://doi.org/10.1016/S1364-6613(03)00029-9)
- Mohammad, Y., & Nishida, T. (2015). *Data Mining for Social Robotics. Toward Autonomously Social Robots*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-25232-2>
- Montague, P. R., Berns, G. S., Cohen, J. D., McClure, S. M., Pagnoni, G., Dhamala, M., Wiest, M. C., Karpov, I., King, R. D., Apple, N., & Fisher, R. E. (2002). Hyperscanning: Simultaneous fMRI during Linked Social Interactions. *NeuroImage*, 16(4), 1159–1164. <https://doi.org/10.1006/NIMG.2002.1150>
- Morganti, F. (2008). What intersubjectivity affords: Paving the way for a dialogue between cognitive science, social cognition and neuroscience. In F. Morganti, A. Carassa, & G. Riva (Eds.), *Enacting Intersubjectivity. A Cognitive and Social Perspective on the Study of Interaction* (Vol. 10, pp. 3–14). IOS Press.
- Morishima, Y., Nakajima, H., Brave, S., Yamada, R., Maldonado, H., Nass, C., & Kawaji, S. (2004). The Role of Affect and Sociality in the Agent-Based Collaborative Learning System. In E. André, L. Dybkjær, W. Minker, & P. Heisterkamp (Eds.), *Affective Dialogue Systems. Tutorial and Research Workshop, ADS 2004, Kloster Irsee, Germany, June 14-16, 2004. Proceedings* (pp. 265–275). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-24842-2_28
- Müller, M. (2015). Assemblages and Actor-networks: Rethinking Socio-material Power, Politics and Space. *Geography Compass*, 9(1), 27–41. <https://doi.org/10.1111/gec3.12192>
- Najafi, M. (2018). *Speech Emotion Recognition using Connectionist Models in a Tandem System* [Florida Institute of Technology]. <https://repository.lib.fit.edu/handle/11141/2514>
- Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, 56(1), 81–103. <https://doi.org/10.1111/0022-4537.00153>

- Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, C. (1995). Can Computer Personalities Be Human Personalities? *Conference Companion on Human Factors in Computing Systems*, 228–229. <https://doi.org/10.1145/223355.223538>
- Nass, C., Steuer, J., & Tauber, E. R. (1994a). Computers are social actors. *Conference Companion on Human Factors in Computing Systems - CHI '94*, 204. <https://doi.org/10.1145/259963.260288>
- Newen, A., Gallagher, S., & De Bruin, L. (2018). 4E Cognition: Historical Roots, Key Concepts, and Central Issues. In A. Newen, L. De Bruin, & S. Gallagher (Eds.), *The Oxford Handbook of 4E Cognition* (Vol. 1). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198735410.013.1>
- Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in Attitudes, Social Perception, and Emotion. *Personality and Social Psychology Review*, 9(3), 184–211. https://doi.org/10.1207/s15327957pspr0903_1
- Niedenthal, P. M., Wood, A., & Rychlowska, M. (2014). Embodied Emotion Concepts. In L. Shapiro (Ed.), *The Routledge Handbook of Embodied Cognition* (pp. 240–249). Routledge.
- Nijholt, A. (2003). Disappearing computers, social actors and embodied agents. *Proceedings. 2003 International Conference on Cyberworlds*, 128–134. <https://doi.org/10.1109/CYBER.2003.1253445>
- Nkambou, R., Bourdeau, J., & Mizoguchi, R. (2010). Introduction: What Are Intelligent Tutoring Systems, and Why This Book? In *Studies in Computational Intelligence* (Vol. 308, Issue January 2015, pp. 1–12). https://doi.org/10.1007/978-3-642-14363-2_1
- Noë, A. (2004). *Action in Perception*. MIT Press.
- Noë, A. (2005). *Action in Perception*. MIT Press.
- Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and family: A review of 17 years of natural language tutoring. In *International Journal of Artificial Intelligence in Education* (Vol. 24, Issue 4, pp. 427–469). Springer New York LLC. <https://doi.org/10.1007/s40593-014-0029-5>
- Ohlsson, S. (1994). Constraint-Based Student Modeling. In *Student Modelling: The Key to Individualized Knowledge-Based Instruction* (pp. 167–189). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-03037-0_7
- Ohlsson, S. (2016). Constraint-Based Modeling: From Cognitive Theory to Computer Tutoring—And Back Again. *International Journal of Artificial Intelligence in Education*, 26(1), 457–473. <https://doi.org/10.1007/s40593-015-0075-7>
- Ohlsson, S., & Mitrovic, A. (2007). Fidelity and Efficiency of Knowledge Representations for Intelligent Tutoring Systems. *Technology, Instruction, Cognition and Learning (TICL)*, 5(2-3-4), 101–132.
- Olney, A. M., D’Mello, S., Person, N., Cade, W., Hays, P., Williams, C., Lehman, B., & Graesser, A. (2012). Guru: A computer tutor that models expert human tutors. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7315 LNCS, 256–261. https://doi.org/10.1007/978-3-642-30950-2_32
- O’Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5), 939–973. <https://doi.org/10.1017/S0140525X01000115>
- Ozogul, G., Johnson, A. M., Atkinson, R. K., & Reisslein, M. (2013). Investigating the impact of pedagogical agent gender matching and learner choice on learning outcomes and perceptions. *Computers and Education*, 67, 36–50. <https://doi.org/10.1016/j.compedu.2013.02.006>
- Paasonen, S., Hillis, K., & Petit, M. (2015). Introduction: Networks of Transmission: Intensity, Sensation, Value. In K. Hillis, S. Paasonen, & M. Petit (Eds.), *Networked Affect* (pp. 1–24). MIT Press.
- Padrón-Rivera, G., Rebolledo-Mendez, G., Parra, P. P., & Huerta-Pacheco, N. S. (2016). Identification of action units related to affective states in a tutoring system for mathematics. *Educational Technology and Society*, 19(2), 77–86.

- Paiva, A. N. A., Leite, I., Boukricha, H., & Wachsmuth, I. (2017). Empathy in Virtual Agents and Robots: A Survey. *ACM Transactions on Interactive Intelligent Systems*, 7(3).
<https://doi.org/10.1145/2912150>
- Palermos, S. O. (2014). Loops, constitution, and cognitive extension. *Cognitive Systems Research*, 27, 25–41. <https://doi.org/10.1016/j.cogsys.2013.04.002>
- Palermos, S. O. (2016). The Dynamics of Group Cognition. *Minds and Machines*, 26(4), 409–440.
<https://doi.org/10.1007/s11023-016-9402-5>
- Pavlik Jr., P. I., Brawner, K. W., Olney, A., & Mitrovic, A. (2013). A Review of Learner Models Used in Intelligent Tutoring Systems. In R. A. Sottolare, A. C. Graesser, X. Hu, & H. Holden (Eds.), *Design Recommendations for Intelligent Tutoring Systems—Volume 1: Learner Modeling* (Vol. 1, pp. 39–68).
- Pereira, G., Prada, R., & Santos, P. A. (2015). Towards Social Power Intelligent Agents. *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015)*, 1857–1858.
- Pereira, G., Prada, R., & Santos, P. A. (2016). Integrating social power into the decision-making of cognitive agents. *Artificial Intelligence*, 241, 1–44. <https://doi.org/10.1016/j.artint.2016.08.003>
- Pérez, D. I., & Gomila, A. (2021). Introduction. In *Social Cognition and the Second Person in Human Interaction* (1st ed., pp. 1–15). Routledge. <https://doi.org/10.4324/9781003133155-101>
- Perez-Osorio, J., & Wykowska, A. (2019). Adopting the intentional stance toward natural and artificial agents. <https://doi.org/10.1080/09515089.2019.1688778>, 33(3), 369–395.
<https://doi.org/10.1080/09515089.2019.1688778>
- Perry, M. (1999). The application of individually and socially distributed cognition in workplace studies: Two peas in a pod? *Proceedings of European Conference on Cognitive Science, 1972*, 87–92.
- Perry, M. (2003a). Distributed Cognition. In J. M. Carroll (Ed.), *HCI Models, Theories, and Frameworks* (pp. 193–223). Elsevier. <https://doi.org/10.1016/B978-155860808-5/50008-3>
- Person, N. K., Graesser, a C., Kreuz, R. J., & Pomeroy, V. (2001). Simulating human tutor dialog moves in AutoTutor. *International Journal of Artificial Intelligence in Education*, 12, 23-39 (16).
- Peschl, M. F., Bottaro, G., Hartner-Tiefenthaler, M., & Rötzer, K. (2014). Learning how to innovate as a socio-epistemological process of co-creation: Towards a constructivist teaching strategy for innovation. *Constructivist Foundations*, 9(3), 421–433.
- Pfeifer, L. M., & Bickmore, T. (2011). Is the media equation a flash in the pan?: The durability and longevity of social responses to computers. *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems*, 777–780. <http://doi.acm.org/10.1145/1978942.1979055>
- Pfeifer, R., & Scheier, C. (1999). *Understanding intelligence*. MIT Press.
- Pfeiffer, U. J., Timmermans, B., Vogeley, K., Frith, C. D., & Schilbach, L. (2013). Towards a neuroscience of social interaction. *Frontiers in Human Neuroscience*, 7, 22.
<https://doi.org/10.3389/fnhum.2013.00022>
- Piaget, J. (1954). *The construction of reality in the child*. Routledge.
- Picard, R. W. (2002). What does it mean for a computer to “have” emotions? In R. Trappl, P. Petta, & S. Payr (Eds.), *Emotions in Humans and Artifacts* (pp. 213–235). A Bradford Book, The MIT Press.
http://search-ebshost-com.uaccess.univie.ac.at/login.aspx?direct=true&db=nlebk&AN=100108&site=ehost-live&ebv=EB&ppid=pp_213
- Pitkin, H. F. (1972). *Wittgenstein and Justice*. University of California Press.
<https://doi.org/10.2307/2218290>
- Prendinger, H., & Ishizuka, M. (2001). Carrying the Role-Playing Metaphor to Interactive Learning Environments. *Proceedings International Conference on Intelligent User Interfaces (IUI-2001)*.
<https://doi.org/>

- Prinz, J. (2004). Embodied emotions. In *Thinking about feeling: Contemporary philosophers on emotions*. Oxford University Press.
- Przyrembel, M., Smallwood, J., Pauen, M., & Singer, T. (2012). Illuminating the dark matter of social neuroscience: Considering the problem of social interaction from philosophical, psychological, and neuroscientific perspectives. *Frontiers in Human Neuroscience*, 6, 190.
<https://doi.org/10.3389/fnhum.2012.00190>
- Qu, C., Brinkman, W. P., Ling, Y., Wiggers, P., & Heynderickx, I. (2014). Conversations with a virtual human: Synthetic emotions and human responses. *Computers in Human Behavior*, 34.
<https://doi.org/10.1016/j.chb.2014.01.033>
- Reeke, G., & Krichmar, J. (2005). The Darwin Brain-Based Automata: Synthetic neural models and real-world devices. In G. N. Reeke, R. R. Poznanski, K. A. Lindsay, J. R. Rosenberg, & O. Sporns (Eds.), *Modeling in the Neurosciences: From biological systems to neuromimetic robotics* (pp. 613–638). CRC Press. <https://doi.org/10.1201/9780203390979.ch23>
- Reeves, B., & Nass, C. I. (1996). The media equation: How people treat computers, television, and new media like real people and places. In *The media equation: How people treat computers, television, and new media like real people and places*. (pp. xiv, 305–xiv, 305). Cambridge University Press.
- Richardson, M. J., & Chemero, A. (2014). Complex Dynamical Systems and Embodiment. In L. Shapiro (Ed.), *The Routledge Handbook of Embodied Cognition* (pp. 39–50). Routledge.
<https://doi.org/10.4324/9781315775845.ch4>
- Riegler, A. (2002). When is a cognitive system embodied? *Cognitive Systems Research*, 3(3), 339–348.
[https://doi.org/10.1016/S1389-0417\(02\)00046-3](https://doi.org/10.1016/S1389-0417(02)00046-3)
- Rietveld, E. (2012). Bodily intentionality and social affordances in context. In F. Paglieri (Ed.), *Consciousness in Interaction. The role of the natural and social context in shaping consciousness* (pp. 207–226). John Benjamins Publishing. <https://doi.org/10.1075/aicr.86.11rie>
- Rietveld, E., de Haan, S., & Denys, D. (2013). Social affordances in context: What is it that we are bodily responsive to? *Behavioral and Brain Sciences*, 36(4), 436–436.
<https://doi.org/10.1017/S0140525X12002038>
- Rietveld, E., & Kiverstein, J. (2014). A Rich Landscape of Affordances. *Ecological Psychology*, 26(4), 325–352. <https://doi.org/10.1080/10407413.2014.958035>
- Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2004a). The neural correlates of theory of mind within interpersonal interactions. *NeuroImage*, 22(4), 1694–1703.
<https://doi.org/10.1016/j.neuroimage.2004.04.015>
- Ritter, F. E., Tehranchi, F., & Oury, J. D. (2019). ACT-R: A cognitive architecture for modeling cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 10(3), e1488.
<https://doi.org/10.1002/wcs.1488>
- Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007). Cognitive tutor: Applied research in mathematics education. *Psychonomic Bulletin and Review*, 14(2), 249–255.
<https://doi.org/10.3758/BF03194060>
- Roetzer, K. (2016). Social interaction with intelligent tutoring systems: An investigation of power and related affect. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 9684).
<https://doi.org/10.1007/978-3-319-39583-8>
- Ross, D., & Ladyman, J. (2010). The Alleged Coupling-Constitution Fallacy and the Mature Sciences. In R. Menary (Ed.), *The Extended Mind* (pp. 154–166). The MIT Press.
<https://doi.org/10.7551/mitpress/9780262014038.003.0007>
- Rottschaefer, W. A. (2017). How Otto did not extend his mind, but might have: Dynamic systems theory and social-cultural group selection. *Cognitive Systems Research*, 45, 124–144.
<https://doi.org/10.1016/j.cogsys.2017.06.001>

- Roulston, K. J. (2008). Conversational Interviewing. In L. M. Given (Ed.), *The SAGE Encyclopedia of Qualitative Research Methods* (pp. 127–129). SAGE Publications, Inc.
<https://doi.org/10.4135/9781412963909.n70>
- Rupert, R. D. (2009). *Cognitive Systems and the Extended Mind*. Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780195379457.001.0001>
- Rus, V., D’Mello, S., Hu, X., & Graesser, A. C. (2013). Recent advances in conversational intelligent tutoring systems. *AI Magazine*, 34(3), 42–54. <https://doi.org/10.1609/aimag.v34i3.2485>
- Sabourin, J., Mott, B., & Lester, J. C. (2011). Modeling Learner Affect with Theoretically Grounded Dynamic Bayesian Networks. In S. D’Mello, A. Graesser, B. Schuller, & J.-C. Martin (Eds.), *Affective Computing and Intelligent Interaction* (pp. 286–295). Springer.
https://doi.org/10.1007/978-3-642-24600-5_32
- Samuels, R. (2019). Classical computational models. In M. Sprevak & M. Colombo (Eds.), *The {Routledge} Handbook of the Computational Mind* (pp. 103–119). Routledge, Taylor & Francis Group.
- Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The Neural Basis of Economic Decision-Making in the Ultimatum Game. *Science*, 300(5626), 1755–1758.
<https://doi.org/10.1126/science.1082976>
- Sayes, E. (2014). Actor – Network Theory and methodology: Just what does it mean to say that nonhumans have agency? *Social Studies of Science*, 44(1).
<https://doi.org/10.1177/0306312713511867>
- Schaumburg, H. (2001). Computers as tools or as social actors? - The users’ perspective on anthropomorphic agents. In *International Journal of Cooperative Information Systems* (Vol. 10, Issues 1–2, pp. 217–234). World Scientific Publishing Co. Pte Ltd.
<https://doi.org/10.1142/s0218843001000321>
- Schilbach, L. (2010). A second-person approach to other minds. *Nature Reviews Neuroscience*, 11(6), 449–449. <https://doi.org/10.1038/nrn2805-c1>
- Schilbach, L. (2014). On the relationship of online and offline social cognition. *Frontiers in Human Neuroscience*, 8, 278. <https://doi.org/10.3389/fnhum.2014.00278>
- Schilbach, L. (2015). Eye to eye, face to face and brain to brain: Novel approaches to study the behavioral dynamics and neural mechanisms of social interactions. In *Current Opinion in Behavioral Sciences* (Vol. 3). <https://doi.org/10.1016/j.cobeha.2015.03.006>
- Schilbach, L., Eickhoff, S. B., Cieslik, E., Shah, N. J., Fink, G. R., & Vogeley, K. (2011). Eyes on me: An fMRI study of the effects of social gaze on action control. *Social Cognitive and Affective Neuroscience*, 6(4), 393–403. <https://doi.org/10.1093/scan/nsq067>
- Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013a). Authors Response: A second-person neuroscience in interaction. *Behavioral and Brain Sciences*, 36(4), 441–462. <https://doi.org/10.1017/S0140525X12002452>
- Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013b). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, 36(4), 393–414.
<https://doi.org/10.1017/S0140525X12000660>
- Schilbach, L., Wohlschlaeger, A. M., Kraemer, N. C., Newen, A., Shah, N. J., Fink, G. R., & Vogeley, K. (2006). Being with virtual others: Neural correlates of social interaction. *Neuropsychologia*, 44(5), 718–730. <https://doi.org/10.1016/j.neuropsychologia.2005.07.017>
- Scholz, O. R. (2013). Soziale und verteilte Kognition (social/distributed cognition). In A. Stephan & S. Walter (Eds.), *Handbuch Kognitionswissenschaft* (pp. 202–206). J.B. Metzler.
- Schöner, G. (2013). Theorie dynamischer Systeme. In A. Stephan & S. Walter (Eds.), *Handbuch Kognitionswissenschaft* (pp. 175–179). J.B. Metzler.

- Schubert, L. (2020). Computational Linguistics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 202). Metaphysics Research Lab, Stanford University.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3.
- Searle, J. R. (1990). Is the brain's mind a computer program? *Scientific American*, 262(1), 26–31.
- Seel, N. M. (Ed.). (2012). Epigenetic Robotics. In *Encyclopedia of the Sciences of Learning* (p. 1157). Springer US. https://doi.org/10.1007/978-1-4419-1428-6_3987
- Self, J. (1990). Theoretical Foundations for Intelligent Tutoring Systems. *J. Artif. Intell. Educ.*, 1(4), 3–14. <https://doi.org/10.5555/95885.95888>
- Serholt, S., & Barendregt, W. (2016). Robots Tutoring Children. *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, 1–10. <https://doi.org/10.1145/2971485.2971536>
- Shapiro, L. (2014). When is Cognition Embodied? In U. Kriegel (Ed.), *Current Controversies in Philosophy of Mind* (pp. 73–90). Routledge.
- Shi, G., Lippert, A. M., Shubeck, K., Fang, Y., Chen, S., Pavlik, P., Greenberg, D., & Graesser, A. C. (2018). Exploring an intelligent tutoring system as a conversation-based assessment tool for reading comprehension. *Behaviormetrika*, 45(2), 615–633. <https://doi.org/10.1007/s41237-018-0065-9>
- Shiri A, M. E., Aïmeur, E., & Frasson, C. (1998). Student modelling by case based reasoning. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1452, 394–403. https://doi.org/10.1007/3-540-68716-5_45
- Shouse, E. (2005). Feeling, Emotion, Affect. *Media/Culture Journal*, 8(6). <http://journal.media-culture.org.au/0512/03-shouse.php>
- Slaby, J. (2014). Emotions and the extended mind. In C. von Scheve & M. Salmela (Eds.), *Collective Emotions* (pp. 32–46). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199659180.003.0003>
- Sleeman, D. H., & Brown, J. S. (1979). Editorial: Intelligent tutoring systems. *International Journal of Man-Machine Studies*, 11(1), 1–3. [https://doi.org/10.1016/S0020-7373\(79\)80002-4](https://doi.org/10.1016/S0020-7373(79)80002-4)
- Slors, M. (2019). Symbiotic cognition as an alternative for socially extended cognition. *Philosophical Psychology*, 32(8), 1179–1203. <https://doi.org/10.1080/09515089.2019.1679591>
- Smith, L., & Gasser, M. (2005). The Development of Embodied Cognition: Six Lessons from Babies. *Artificial Life*, 11(1–2), 13–29. <https://doi.org/10.1162/1064546053278973>
- Sohlberg, P., & Leiulfstrud, H. (2018). Conceptual Constructionism: An Introduction. In P. Sohlberg & H. Leiulfstrud (Eds.), *Concepts in Action: Conceptual Constructivism*. (pp. 1–22). Brill. https://doi.org/10.1163/9789004314207_002
- Spradley, J. P. (1979). *The ethnographic interview*. Harcourt Brace College Publishers.
- Spradley, J. P. (1980). *Participant observation*. Harcourt Brace College Publishers.
- Stapleton, M., & Froese, T. (2015). Is Collective Agency a Coherent Idea? Considerations from the Enactive Theory of Agency. In C. Misselhorn (Ed.), *Collective Agency and Cooperation in Natural and Artificial Systems* (pp. 219–236). Springer International Publishing. https://doi.org/10.1007/978-3-319-15515-9_12
- Stephan, A. (2018). Critical Note: 3E's Are Sufficient, But Don't Forget the D. In A. Newen, L. De Bruin, & S. Gallagher (Eds.), *The Oxford Handbook of 4E Cognition* (Vol. 1). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198735410.013.32>
- Stevanovic, M., & Koski, S. E. (2018). Intersubjectivity and the domains of social interaction: Proposal of a cross-sectional approach. *Psychology of Language and Communication*, 22(1), 39–70. <https://doi.org/10.2478/plc-2018-0003>

- Strasser, A. (2015). Can Artificial Systems Be Part of a Collective Action? In C. Misselhorn (Ed.), *Collective Agency and Cooperation in Natural and Artificial Systems* (pp. 205–218). Springer International Publishing. https://doi.org/10.1007/978-3-319-15515-9_11
- Strasser, A. (2018a). Social Cognition and Artificial Agents. In V. C. Müller (Ed.), *Philosophy and Theory of Artificial Intelligence 2017. PT-AI 2017. Studies in Applied Philosophy, Epistemology and Rational Ethics* (Vol. 44, pp. 106–114). Springer International Publishing. https://doi.org/10.1007/978-3-319-96448-5_12
- Strauss, A. L. (1987). *Qualitative analysis for social scientists*. Cambridge University Press.
- Strauss, A. L., & Corbin, J. (1991). *Basics of qualitative research: Grounded theory procedures and techniques* (5. print..). Sage.
- Sturm, T., & Gundlach, H. (2013). Ursprünge und Anfänge der Kognitionswissenschaft. Zur Geschichte und Geschichtsschreibung der ›kognitiven Revolution‹ Zur Geschichte und Geschichtsschreibung der ›kognitiven Revolution‹ – eine Reflexion schreibung der ›kognitiven Revolution‹ – eine Reflexion. In A. Stephan & S. Walter (Eds.), *Handbuch Kognitionswissenschaft* (pp. 7–21). J.B. Metzler. https://doi.org/10.1007/978-3-476-05288-9_2
- Suchman, L. (2006). *Human–Machine Reconfigurations*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511808418>
- Suchman, L. A. (1987). *Plans and situated actions: {The} problem of human-machine communication*. Cambridge University Press.
- Suchman, L., Trigg, R., & Blomberg, J. (2002). Working artefacts: Ethnomethods of the prototype. *The British Journal of Sociology*, 53(2), 163–179. <https://doi.org/10.1080/00071310220133287>
- Suraweera, P., & Mitrovic, A. (2004). An Intelligent Tutoring System for Entity Relationship Modelling. *Int. J. Artif. Intell. Ed.*, 14(3,4), 375–417.
- Tanevska, A., Rea, F., Sandini, G., Cañamero, L., & Sciutti, A. (2020). A Socially Adaptable Framework for Human-Robot Interaction. *Frontiers in Robotics and AI*, 7, 121. <https://doi.org/10.3389/FROBT.2020.00121/BIBTEX>
- Thagard, P. (2005). *Mind. {Introduction} to cognitive science*. MIT Press.
- Thamm, R. (2004). Towards A Universal Power And Status Theory Of Emotion. In J. H. Turner (Ed.), *Theory and Research on Human Emotions* (Vol. 21, pp. 189–222). Emerald Group Publishing Limited. [https://doi.org/10.1016/S0882-6145\(04\)21008-6](https://doi.org/10.1016/S0882-6145(04)21008-6)
- Thelen, E., Schönner, G., Scheier, C., & Smith, L. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *The Behavioral and Brain Sciences*, 24(1), 1–34; discussion 34–86. <https://doi.org/10.1017/s0140525x01003910>
- Thellman, S. (2021). *Social Robots as Intentional Agents* (Vol. 811) [Linköping University Electronic Press]. <https://doi.org/10.3384/diss.diva-178806>
- Thompson, E. (2010). *Mind in Life. Biology, Phenomenology, and the Sciences of Mind*. Harvard University Press.
- Thompson, E., & Stapleton, M. (2009). Making sense of sense-making: Reflections on enactive and extended mind theories. *Topoi*, 28(1), 23–30. <https://doi.org/10.1007/s11245-008-9043-2>
- Thornberg, R., & Charmaz, K. (2014). Grounded Theory and Theoretical Coding. In U. Flick (Ed.), *The SAGE Handbook of Qualitative Data Analysis* (pp. 153–169). SAGE Publications Ltd. <https://doi.org/10.4135/9781446282243.n11>
- Thrift, N. (2004). Intensities of feeling: Towards a spatial politics of affect. *Geografiska Annaler: Series B, Human Geography*, 86(1), 57–78. <https://doi.org/10.1111/j.0435-3684.2004.00154.x>
- Thürmel, S. (2015). The Participatory Turn: A Multidimensional Gradual Agency Concept for Human and Non-human Actors. In C. Misselhorn (Ed.), *Collective Agency and Cooperation in Natural and*

- Artificial Systems* (pp. 45–60). Springer International Publishing. https://doi.org/10.1007/978-3-319-15515-9_3
- Tomkins, S. (2008). *Affect Imagery Consciousness: The Complete Edition: Two Volumes*. Springer.
- Turner, J. H. (2009). *The Sociology of Emotions: Basic Theoretical Arguments*. 1(4), 340–354. <https://doi.org/10.1177/1754073909338305>
- Valentin, D., Abdi, H., O’Toole, A. J., & Cottrell, G. W. (1994). Connectionist models of face processing: A survey. *Pattern Recognition*, 27(9), 1209–1230. [https://doi.org/10.1016/0031-3203\(94\)90006-X](https://doi.org/10.1016/0031-3203(94)90006-X)
- van Gelder, T. (1995). What might cognition be, if not computation? *Journal of Philosophy*, 92(7), 354–382.
- van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, 21(5), 615–628. <https://doi.org/10.1017/S0140525X98001733>
- van Vugt, H. C., Hoorn, J. F., Konijn, E. A., & de Bie Dimitriadou, A. (2006). Affective affordances: Improving interface character engagement through interaction. *International Journal of Human-Computer Studies*, 64(9), 874–888. <https://doi.org/10.1016/j.ijhcs.2006.04.008>
- VanLehn, K. (2006). The Behavior of Tutoring Systems. *Int. J. Artif. Intell. Ed.*, 16(3), 227–265.
- VanLehn, K., Jordan, P. W., Rosé, C. P., Bhembé, D., Böttner, M., Gaydos, A., Makatchev, M., Pappuswamy, U., Ringenberg, M., Roque, A., Siler, S., & Srivastava, R. (2002). The architecture of why2-atlas: A coach for qualitative physics essay writing. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2363, 158–167. https://doi.org/10.1007/3-540-47987-2_20
- VanLehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R., Taylor, L., Treacy, D., Weinstein, A., & Wintersgill, M. (2005). The Andes Physics Tutoring System: Lessons Learned. *International Journal of Artificial Intelligence in Education*, 15(3), 147–204.
- VanLehn, K., Van De Sande, B., Shelby, R., & Gershman, S. (2010). The Andes physics tutoring system: An experiment in freedom. *Studies in Computational Intelligence*, 308, 421–443. https://doi.org/10.1007/978-3-642-14363-2_21
- Varela, F. J. (1997). Patterns of life: Intertwining identity and cognition. *Brain and Cognition*, 34(1), 72–87. <https://doi.org/10.1006/brcg.1997.0907>
- Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. MIT Press.
- Varga, S. (2016). Interaction and extended cognition. *Synthese*, 193(8), 2469–2496. <https://doi.org/10.1007/s11229-015-0861-7>
- Veletsianos, G. (2010). Contextually relevant pedagogical agents: Visual appearance, stereotypes, and first impressions and their impact on learning. *Computers and Education*, 55(2), 576–585. <https://doi.org/10.1016/j.compedu.2010.02.019>
- Veletsianos, G., & Miller, C. (2008). Conversing with pedagogical agents: A phenomenological exploration of interacting with digital entities. *British Journal of Educational Technology*, 39(6), 969–986. <https://doi.org/10.1111/j.1467-8535.2007.00797.x>
- Veletsianos, G., Miller, C., & Doering, A. (2010). EnAli: A Research and Design Framework for Virtual Characters and Pedagogical Agents. *Journal of Educational Computing*, 41(2), 171–194.
- Veletsianos, G., & Russell, G. S. (2014). Pedagogical Agents. In M. J. Spector, D. M. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (fourth ed., pp. 759–769). Springer New York. https://doi.org/DOI.10.1007/978-1-4614-3185-5_61
- Vernon, D. (2014). *Artificial Cognitive Systems. A Primer*. The MIT Press. <https://mitpress.mit.edu/books/artificial-cognitive-systems>

- Villalobos, M., & Silverman, D. (2018). Extended functionalism, radical enactivism, and the autopoietic theory of cognition: Prospects for a full revolution in cognitive science. *Phenomenology and the Cognitive Sciences*, 17(4). <https://doi.org/10.1007/s11097-017-9542-y>
- Vygotsky, L. S. (1978). *Mind in Society: Development of Higher Psychological Processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.
- Wachsmuth, I. (2015). Embodied Cooperative Systems: From Tool to Partnership. In C. Misselhorn (Ed.), *Collective Agency and Cooperation in Natural and Artificial Systems* (pp. 63–79). Springer International Publishing. https://doi.org/10.1007/978-3-319-15515-9_4
- Walter, H., Adenzato, M., Ciaramidaro, A., Enrici, I., Pia, L., & Bara, B. G. (2004). Understanding intentions in social interaction: The role of the anterior paracingulate cortex. *Journal of Cognitive Neuroscience*, 16(10), 1854–1863. <https://doi.org/10.1162/0898929042947838>
- Walter, S. (2014a). *Kognition*. Philipp Reclam jun.
- Walter, S. (2014b). Situated Cognition: A Field Guide to Some Open Conceptual and Ontological Issues. *Review of Philosophy and Psychology*, 5(2), 241–263. <https://doi.org/10.1007/s13164-013-0167-y>
- Ward, D., Silverman, D., & Villalobos, M. (2017). Introduction: The Varieties of Enactivism. *Topoi*, 36(3), 365–375. <https://doi.org/10.1007/s11245-017-9484-6>
- Ward, D., & Stapleton, M. (2012). Es are good. In F. Paglieri (Ed.), *Consciousness in Interaction. The role of the natural and social context in shaping consciousness* (Vol. 33, Issue 2, pp. 89–104). John Benjamins Publishing. <https://doi.org/10.1075/aicr.86.06war>
- Ward, L. M. (2002). *Dynamical cognitive science*. MIT Press.
- Weber, A., & Varela, F. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences*, 1(2), 97–125. <https://doi.org/10.1023/A:1020368120174>
- Weber, M. (1978). *Economy and Society: An Outline of Interpretive Sociology*. University of California Press.
- Wenger, E. (1987). *Artificial Intelligence and Tutoring Systems: Computational and Cognitive Approaches to the Communication of Knowledge*. Morgan Kaufmann Publishers Inc. <https://dl.acm.org/doi/book/10.5555/42185>
- Wetherell, M. (2012). *Affect and Emotion. A New Social Science Understanding*. SAGE Publications Ltd.
- Wetherell, M. (2013). Affect and discourse – What’s the problem? From affect as excess to affective/discursive practice. *Subjectivity*, 6(4), 349–368. <https://doi.org/10.1057/sub.2013.13>
- Wetherell, M. (2015). Trends in the Turn to Affect: A Social Psychological Critique. *Body and Society*, 21(2), 139–166. <https://doi.org/10.1177/1357034X14539020>
- Wheeler, M. (2011). Mind in life or life in mind? Making sense of deep continuity. *Journal of Consciousness Studies*, 18(5–6), 148–168.
- Whisker, C. (2018). Review: Adele E. Clarke, Carrie Friese & Rachel S. Washburn (2018). Situational Analysis: Grounded Theory After the Interpretive Turn. *Forum: Qualitative Social Research*, 19(3). <https://doi.org/10.17169/fqs-19.3.3138>
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4), 625–636.
- Wilson, R. A. (1994). Wide Computationalism. *Mind*, 103(411), 351–372. <https://doi.org/10.1093/mind/103.411.351>
- Wilson, R. (2005). Collective memory, group minds, and the extended mind thesis. *Cognitive Processing*, 6(4), 227–236. <https://doi.org/10.1007/s10339-005-0012-z>
- Wilson, R., & Clark, A. (2009). How to Situate Cognition. Letting Nature Take Its Course. In P. Robbins & M. Aydede (Eds.), *The Cambridge Handbook of Situated Cognition* (pp. 55–77). Cambridge University Press. <https://doi.org/10.1017/CBO9780511816826.004>

- Wilson, R. A., & Foglia, L. (2017). Embodied Cognition. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.
- Woods, M., Paulus, T., Atkins, D. P., & Macklin, R. (2016). Advancing Qualitative Research Using Qualitative Data Analysis Software (QDAS)? Reviewing Potential Versus Practice in Published Studies using ATLAS.ti and NVivo, 1994–2013. *Social Science Computer Review*, 34(5), 597–617. <https://doi.org/10.1177/0894439315596311>
- Woolf, B. P. (2009). Building Intelligent Interactive Tutors. In *Building Intelligent Interactive Tutors*. Elsevier Inc. <https://doi.org/10.1016/B978-0-12-373594-2.X0001-9>
- Xu, J.-X. (2012). Initial State Learning. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 1566–1568). Springer US. https://doi.org/10.1007/978-1-4419-1428-6_537
- Ye, P., Wang, T., & Wang, F. Y. (2018). A survey of cognitive architectures in the past 20 years. *IEEE Transactions on Cybernetics*, 48(12), 3280–3290. <https://doi.org/10.1109/TCYB.2018.2857704>
- Ziemke, T. (2003). What's that Thing Called Embodiment? *Proceedings of the Annual Meeting of the Cognitive Science Society*, 1305–1310.

8 Appendix

8.1 List of Figures

Figure 2: Four-component architecture; this illustration is an adapted version of [Nkambou et al. 2010, p.4., fig. 1.1] that has been modified for the purposes of this master's thesis.....	113
Figure 3: AutoTutor architecture; this illustration is an adapted version of [D'Mello & Graesser 2012, p. 23:12, fig. 2] that has been modified for the purposes of this master's thesis.....	139
Figure 4: Simplified tutoring session structure; design of tutor actions and dialogue moves. Grey fields represent modes in which the tutor cannot be interrupted by the student (lecture mode). The elements comprising the question-repetition-format (based on EMT-dialogue) are depicted in white.....	150
Figure 1: Simplified context map of paradigm shifts and turns in cognitive science.....	11
Figure 5: Process of data collection and analysis (simplified).....	157
Figure 6: Conceptual Analysis Road Map.....	166
Figure 7: Identified 'social variants' of situated cognition paradigms and associated theories concerned with social interaction and/or cognition.....	197

8.2 List of Tables

Table 1: Concepts related to social interaction identified in philosophy: ToM.....	22
Table 2: Concepts related to social interaction identified in philosophy: Intersubjectivity.....	23
Table 3: Concepts related to social interaction identified in neuroscience and psychology.....	25
Table 4: Concepts related to social interaction identified in AI (including artificial life, behaviour-based, evolutionary, and social robotics, as well as theory/philosophy of AI).....	29
Table 5: Concepts related to social interaction identified in computer science (including the fields of human-robot and human-computer interaction).....	32
Table 6: Overview of the curriculum.....	147
Table 7: Summary regarding fruitfulness of paradigms and theories/concepts with regard to social human-ITS interaction.....	200

8.3 Abstract (English)

This master's thesis is concerned with the phenomenon of social interaction and aims to investigate conceptualisations of social interaction within cognitive science. It furthermore investigates a specific case of human-computer interaction as a social kind of social interaction; the interaction between a human learner and a type of educational technology called 'Intelligent Tutoring System' (ITS). Concretely, the following three research questions (RQ) are addressed:

- RQ1: What conceptions of social interaction currently exist in classical/computational and situated/4E cognitive science paradigms?
- RQ2: How do power and related affect influence and shape the social interaction with ITS regarding: i) social roles; ii) interaction and the (situational) structures it is embedded in; and iii) social relations?
- RQ3: Which of the conceptions from RQ1 are relevant for and can be fruitfully applied to the case of social interaction with ITS?

This master's thesis comprises three parts; one for each RQ. Part I corresponds to RQ1, Part II to RQ2, and Part III to RQ3.

Part I reviews major paradigm shifts and turns within the field. Social cognition and the issues of methodological individualism, body-social problem, and human-centredness are introduced. Social interaction and related concepts across cognitive science disciplines are described, followed by a conceptual investigation of social interaction across cognitive science paradigms. Classical paradigms (computational paradigms) and situated approaches (situated/4EA cognition) are examined, whereby the focus of Part I lies on situated cognition paradigms.

Part II encompasses a comprehensive reporting of a research design proposal for addressing RQ2. The field of Intelligent Tutoring Systems (ITS) is introduced. A theoretical review of social science conceptions and definition of power and affect is provided. The concrete methodological approach is reported. This includes the concrete research setting and how the ITS AutoTutor was used to create tutoring sessions to be investigated, the concrete design and content of these sessions (i.e. the curriculum and learning goals), as well as methods for data collection and analysis, which comprise a combination of Constructivist Grounded Theory and Situational Analysis.

Part III addresses RQ3. This comprises a conceptual discussion of social interaction with ITS from an embodied, embedded, extended, distributed, and enactive cognition perspective. Finally, conclusions across Part I and II are drawn, including limitations and future work.

8.4 Abstract (German)

Diese Masterarbeit befasst sich mit dem Phänomen der sozialen Interaktion und zielt darauf ab, Konzeptualisierungen der sozialen Interaktion innerhalb der Kognitionswissenschaft zu untersuchen. Darüber hinaus wird ein spezifischer Fall von Mensch-Computer-Interaktion als eine soziale Art von sozialer Interaktion untersucht: die Interaktion zwischen einem menschlichen Lernen und einer spezielle Art von computerbasiertem Lernsystem, das als „Intelligent Tutoring System“ (ITS) bezeichnet wird. Konkret werde die folgenden drei Forschungsfragen (FF) behandelt:

- FF1: Welche Konzeptionen von sozialer Interaktion existieren derzeit in klassischen und situierten/„4E“ kognitionswissenschaftlichen Paradigmen
- FF2: Wie beeinflussen und gestalten Macht und damit verbundene Affekte die soziale Interaktion mit ITS in Bezug auf: i) soziale Rollen; ii) Interaktion und die (situativen) Strukturen, in die sie eingebettet ist; und iii) soziale Beziehungen?
- FF3: Welche der Konzepte aus FF1 sind für die soziale Interaktion mit ITS relevant und können auf diese fruchtbar angewendet werden?

Diese Masterarbeit besteht aus drei Teilen; einer für jede Fragestellung. Teil I bezieht sich auf FF1, Teil II auf FF2 und Teil III auf FF3.

Teil I beschreibt die wichtigsten Paradigmenwechsel in den Kognitionswissenschaften. Soziale Kognition und die Themen des methodologischen Individualismus, des Körper-Soziales-Problems und der Menschenzentrierung werden vorgestellt. Soziale Interaktion und damit zusammenhängende Konzepte in kognitionswissenschaftlichen Disziplinen werden beschrieben. Das Hauptaugenmerk von Teil I liegt auf der konzeptionellen Untersuchung von sozialer Interaktion in den kognitionswissenschaftlichen Paradigmen. Untersucht werden klassische Paradigmen und situierte Ansätze, wobei der Schwerpunkt von Teil I auf situierten Paradigmen liegt.

Teil II umfasst eine umfassende Berichterstattung über ein Forschungsdesign zur Beantwortung von FF2. Das Feld der „Intelligent Tutoring Systems“ wird vorgestellt. Es wird ein theoretischer Überblick über sozialwissenschaftliche Konzepte und Definitionen von Macht und Affekt gegeben. Es wird über den konkreten methodischen Ansatz berichtet: Forschungssetting und die Art und Weise, wie das ITS AutoTutor verwendet wurde, sowie die konkreten Lerninhalte (Curriculum). Die Methoden zur Datenerhebung und -analyse, die eine Kombination aus konstruktivistischer Grounded Theory und Situationsanalyse umfassen, werden im Detail beschrieben.

Teil III umfasst eine konzeptionelle Diskussion der sozialen Interaktion mit ITS aus der Perspektive der verkörperten, eingebetteten, erweiterten, verteilten und enaktiven Kognition (FF3).