MEDiTATe: a First Step of a Journey from BDI to Neuroscience, and Back

Angelo Ferrando[1][0000−0002−8711−4670], Andrea Gatti[2][0009−0003−0992−4058], and Viviana Mascardi[2][0000−0002−2261−9926]

[1] University of Modena-Reggio Emilia, Italy, angelo.ferrando@unimore.it
[2] University of Genova, Italy, andrea.gatti@edu.unige.it, viviana.mascardi@unige.it

Abstract. The literature that analyzes how neuroscience inspired AI, and viceversa, mainly takes a machine learning point of view: however, the connections between neuroscience findings and intelligent software agents modeled after the Belief-Desire-Intention (BDI) architecture are many, and deserve to be addressed and understood. In order to provide a framework to our exploration, and make it more concrete, we introduce the BDI-inspired MEDiTATe conceptual architecture encompassing theory of **M**ind, **E**motions, **D**eep **TA**lk, and small **T**alk.
MEDiTATe is intended as a principled means to analyze the connections between neuroscience and BDI approaches in a systematic way, and to interact with neuro-scientists by sharing a common terminological ground. The contribution of this paper is indeed to survey the relevant scientific literature and organize the findings of this review coherently with the MEDiTATe conceptual architecture.
Nonetheless, most modules of MEDiTATe have been, or may be, implemented using a well known framework for BDI agents, Jason. In this sense, the possibility to move MEDiTATe from the conceptual level to the practical one is backed up by existing software tools.
MEDiTATe features *small talk* and *deep talk* that we conjecture to be related but distinct cognitive functions, each with its own purpose and possibly dedicated different brain areas. We expect that MEDiTATe – once fully developed – may support the study of these functions and of their relations with other, better understood, cognitive processes, possibly inspiring experiments by neuro-scientists to validate the hypothesis. In fact, in our long-term vision, MEDiTATe should offer to computer scientists and neuro-scientists a shared gym for experimenting models and theories of brain functioning.

Keywords: MEDiTATe, Deep Talk, Small Talk, Neuroscience, Mind, Emotions, Beliefs-Desires-Intentions, BDI, Cognitive Agents

1 Introduction

Since its conception in the mid-1950, Artificial Intelligence (AI) – envisioned by John McCarthy as *the science and engineering of making intelligent machines*

– was interconnect with sciences studying human intelligence from a biological, functional, medical, and psychological perspective.

Even before the term AI was born, the studies carried out by McCulloch and Pitts on artificial neural networks were directly rooted in neuroscience [64], and Reinforcement Learning is inspired by animal learning psychology [94].

These examples show that, often, computer science and AI rely on discoveries by psychologists and neuroscientists. Some times, however, computer science and AI anticipate discoveries made later on. One example is the idea that memory might consist of a fast access, short term component, and a slower access, long term one. The cache computer memory, implementing this idea, was first developed by Wilkes in 1965 [102], but systematic and coherent models of short term and long term human memory appeared only later [92,6].

Finally, in some cases AI systems show unanticipated similarities with human cognitive functions, suggesting that the exploration of how the AI system works might lead to a better understanding of how the brain works [97].

Some reviews analyze how neuroscience inspired AI, and viceversa [47,54], but – unfortunately, but not surprisingly – they mainly assume that artificial intelligence is machine learning. In this review paper we complement those works via a principled discussion of the connections between neuroscience findings and intelligent software agents modeled after the Belief-Desire-Intention (BDI) architecture [79]. We limit our investigation to theory of mind, emotions and language, and we envision MEDiTATe (theory of **M**ind, **E**motions, **D**eep **TA**lk, and small **T**alk) that extends the BDI architecture and provides a conceptual framework for our investigation.

Two innovative elements characterize this paper. On the neuroscience side, we consider *small talk* and *deep talk* as two distinct cognitive functions, pursuing different goals. While this is rooted on psychological studies [5,88,56,66], we formulate the hypothesis that – in the same way as different parts of the brain are involved in fast and slow thinking [29], in short and long term memory, etc – the brain's areas specifically devoted to small talk and deep talk are different, and this is reflected in the MEDiTATe architecture. On the Engineering Multiagent Systems side, we envision that being based in solid scientific studies from both computer science and neurosciences, MEDiTATe may represent the first step towards the development of an effective playground for experimenting not only sophisticated models of cognitive software agents and of their communication mechanisms, but also theories on the brain functioning.

The MEDiTATe vision is grounded in recent scientific literature and has a strong practical flavor: working prototypes of most of its components have been – or might be – implemented in Jason [12], one very popular implementation of the AgentSpeak(L) language [78] for programming BDI agents.

2 From BDI to Neuroscience

The work by Georgeff and Rao on BDI agents was inspired by the philosophical studies on intentionality by Brentano [17], Dennet [33], Bratman [16]. To the best of our knowledge, they did not directly take neuroscience findings into

account. Nevertheless, generation and management of Beliefs, Desires, Goals, Intentions, Plans, namely of the key components of the BDI architecture, are brain functionalities, supported by specific areas in the brain.

In this section we make the connection between Beliefs, Desires, Goals, Intentions, Plans and brain functionalities explicit: for each of them we discuss what it serves for (its ***function***), the ***major anatomical structures involved*** in the brain, according to recent literature and accurate brain maps[3], and one ***feasible implementation in Jason***. The Theory of Mind, Emotions, Deep and Small Talk components are presented in Section 3, following the same schema.

2.1 Beliefs

Although the most immediate counterpart of Beliefs is memory, memory also involves unconscious procedural information which has no "twin" in the BDI architecture. Budson and Price's [18] provide a clear introduction to human memory by classifying memory systems in explicit (associated with conscious awareness) and declarative (that can be consciously recalled), versus implicit (associated with change in behavior) and nondeclarative (unconscious). They also present four different kinds of memory: episodic, semantic, procedural, and working.

Episodic Memory – ***Function:*** Episodic memory refers to the explicit and declarative memory system used to recall personal experiences framed in our own context. ***Major anatomical structures involved:*** Prefrontal cortex, medial temporal lobes, anterior thalamic nucleus, mammillary body, fornix. ***Feasible implementation in Jason:*** Beliefs with `personal, long-term` annotation.

Semantic Memory – ***Function:*** Semantic memory refers to our general store of conceptual and factual knowledge not related to any specific memory. It is a declarative and explicit memory system. ***Major anatomical structures involved:*** Inferolateral temporal lobes. ***Feasible implementation in Jason:*** Beliefs with `factual, long-term` annotation.

Procedural Memory – ***Function:*** Procedural memory refers to the ability to learn behavioral and cognitive skills and algorithms that are used at an automatic, unconscious level. Procedural memory is nondeclarative but during acquisition may be either explicit or implicit. ***Major anatomical structures involved:*** Basal ganglia, cerebellum, supplementary motor area. ***Feasible implementation in Jason:*** No explicit and direct BDI twin exists for procedural memory, as the BDI architecture does not integrate "cognitive skills used in an automatic way". However, in Jason internal actions may represent a feasible way for the agent to run an algorithm "without thinking about it", so in an "unconscious", "automatic" way. Jason internal actions are implemented in Java and support is given, e.g., for binding of logical variables. This paves the way to model (simulated, but also real, in principle) actions like "driving in a known

[3] See for example `https://dana.org/resources/neuroanatomy-the-basics/`.

road with light traffic". What cannot be easily supported by Jason, w.r.t. human procedural memory, is the ability to learn such internal actions, and to move them from System 2 (deliberative, slow) to System 1 (unconscious, fast) [29]. We are not aware of proposals dealing with this capability in the BDI literature.

Working Memory – ***Function:*** Working memory is an explicit and declarative memory system combining the fields of attention, concentration, and short-term memory. It refers to the ability to temporarily maintain and manipulate information that one needs to keep in mind. ***Major anatomical structures involved:*** Prefrontal cortex, Broca's area, Wernicke's area (limited to phonologic working memory). ***Feasible implementation in Jason:*** Beliefs with `short-term` annotation instead of `long-term` one.

2.2 Desires and Goals

Function: Pleasure serves to motivate individuals to pursue rewards necessary for fitness, and rewards involve a composite of several psychological components: liking (core reactions to hedonic impact), wanting (motivation process of incentive salience), and learning (Pavlovian or instrumental associations and cognitive representations) [10]. Intuitively, goals are usually states we want but have difficulty achieving even when we know they are achievable. Discriminating between desires and goals in neurobiology is difficult, as desires may be seen as one of the two goals' dimensions, *the will*, with the other dimension being *the way* [9]. ***Major anatomical structures involved:*** Prefrontal cortex. ***Feasible implementation in Jason:*** The support that Jason offers to representing goals and to managing them during the agent's reasoning cycle directly comes from the AgentSpeak(L) operational semantics, and is described in the Jason related resources. As far as liking is concerned, besides ad-hoc beliefs that model what agents like and dislike, or annotations to beliefs, there is no directly supported counterpart in Jason. We may consider preferences associated with goals, along the lines of [22,23,24]. When learning comes into play, we may mention the recent proposals to integrate Reinforcement Learning (RL) in Jason [7,8,100,13,76,72]. While the idea of injecting some RL into BDI agents dates back to the beginning of the millennium [69,61,60,1,77,59,95], implementations in Jason became available only recently.

2.3 Intentions and Plans

Function: Intentions operate at the interface of thought and action, translating cognitive states into detailed motor coordination. Jahanshahi [52] and Brass and Haggard [15] suggests that intentions consist of a "what to do" component, a decision "when to act", and an inhibitory process (a "whether" element in Brass and Haggard model). Apparently, plans should be easier to characterize than intentions. Their behavioral and psychological intuition is clear, and their computational counterpart is even clearer: a plan is a sequence of actions, and planning is a process that considers actions and their sequential interdependence in terms of the desirability of their outcomes. However, planning *remains one*

of the most elusive cognitive processes at the neural level [63]. ***Major anatomical structures involved:*** Prefrontal cortex. ***Feasible implementation in Jason:*** Jason agents – coherently with AgentSpeak(L) – consist of a belief set and a plan set; when relevant (namely, triggered by the current event selected by the *event selection function*) and applicable (namely, characterized by a context that is a logical consequence of the current beliefs) plans are selected for execution, they become intentions. Intentions are data structures used at runtime by the Jason interpreter, and correspond to stacks of partially instantiated plans. While structures named plans and intentions are already integrated in Jason, no dynamic, first-principles planning is supported by design. However, both old [32,99,91] and recent [65,104] proposals for extending the basic BDI model with dynamic planning exist. Some of them target Jason or its JaCaMo extension [11] as their implementation framework [20].

3 From Neuroscience to MEDiTATe

In this section, we focus on the "from neuroscience to MEDiTATe" direction by looking at brain science outcomes that do not fit the original BDI architecture, but that might be integrated into its MEDiTATe extension. The MEDiTATe components, as well as a rough sketch of the input, output, and working interpreter, are shown in Figure 1. For all the MEDiTATe components we highlighted their cognitive function (top), the brain areas involved (medium), and feasible implementations in Jason (bottom), if available.

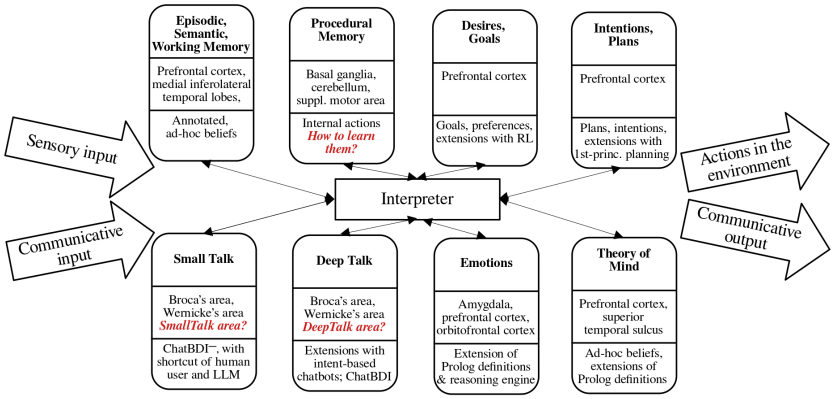

Fig. 1. MEDiTATe architecture.

3.1 Theory of Mind

Function: Theory of Mind (ToM) is the ability to reason about mental states, such as beliefs, desires, and intentions, in order to explain and predict people's

behavior [3], and to plan how to behave in social situations [48].Neuroimaging findings suggest that there are several core regions in the brain, including parts of the prefrontal cortex and superior temporal sulcus, that contribute to ToM reasoning [21]. ***Major anatomical structures involved:*** Prefrontal cortex, superior temporal sulcus. ***Feasible implementation in Jason:*** The idea that autonomous agents and robots need a ToM to engage into social interactions with humans and among themselves is as old as the idea of agent itself [30,31], and it is still objective of active research [90]. Many works explore how intelligent agents may exhibit a ToM [81] and the BDI architecture is a very natural framework for this investigation [46,14]. Various proofs of concept have been developed in Jason or JaCaMo [87,71,19,28,67,89,105], often by annotating extensional beliefs or exploiting the Prolog intensional definition of beliefs with abduction and other ToM-related rules. This suggests that an implementation (or better, an approximation) of ToM in Jason is feasible.

3.2 Emotions

Function: Emotions play a myriad of roles at intrapersonal, interpersonal, and social and cultural levels [82,49]. At the intrapersonal level, they prepare us for behavior with minimal thinking [27] and associate memories with the emotions experienced at those times the facts occurred, allowing us to create "emotional connections" among disparate facts [101]. At the interpersonal level, they send non verbal signals to others and influence others and our social interactions [35]. Finally, the development and transmission of attitudes, values, beliefs, and norms related to emotions, is part of cultural transmission and operates then at the cultural level [62]. The brain areas devoted to managing emotions have been studied for more than thirty years [57,58,82], with the amygdala playing a major role in processing emotions and linking them to memories, learning, and sensing, and – due to the complexity of emotions and of their relation with cognitive processes – with many other central and peripherical areas involved. ***Major anatomical structures involved:*** Amygdala, prefrontal cortex, orbitofrontal cortex. ***Feasible implementation in Jason:*** Various extensions of the BDI architecture and of its underpinning formal model have been proposed over the last years, aimed at integrating emotions [73,75,93,4]. Sánchez and Cerezo's survey is a good starting point for overviewing the literature on the topic [85]. Not surprisingly, Jason is often used as a handy and flexible tool to experiment with BDI emotional agents [96,2,25].

3.3 Deep and Small Talk

In this section we put forward the most visionary and unexplored component on MEDiTATe, namely the one related with language, and we differentiate between talking deep, and talking small. We keep the distinction because it is very relevant from a computational point of view, although – to the best of our knowledge – no neurological studies have been specifically performed on localization of these two functions. According to the Cambridge Dictionary, *small talk* is a

conversation about things that are not important, often between people who do not know each other well[4]. This is often used in contrasts with *deep talk*, meaning a conversation involving increasingly greater self-disclosure[5]. The language area involved in turning thoughts into words is Broca's area, while Wernicke's areas is involved in language understanding and processing. The angular gyrus processes concrete and abstract concepts and plays a role in verbal working memory during retrieval of verbal information. While not all the scientists agree on the localization of language functions [98], the Broca-Wernicke's theory still holds a dominant position in neurosciences [84].

Talk Deep – ***Function:*** While it is not always possible to engage into deep, intimate and self-disclosing talk, recent experimental studies show that people feel more connected to deep conversation partners than shallow conversation partners [56,66]. Deep talk may strengthen social connections, besides leaving lasting memories [26]. ***Major anatomical structure involved.*** Not explored; we name it "TalkDeep area". ***Feasible implementation in Jason:*** The literature on BDI implementations of conversational agents and dialogue systems is almost rich [103,68,34,50], but just a few recent papers use Jason as implementation language. In a set of papers published between 2021 and 2023 [38,37,36,39], Engelmann et al. present Dial4JaCa. Dial4JaCa integrates JaCaMo and Dialogflow [44], an intent-based chatbot platform developed by Google. VEsNA [42] exploits Dial4JaCa to bridge a human user speaking in natural language, and a Virtual Reality (VR) environment. We mention these works here, because they exploit a chatbot platform driven by the recognition of "intents" of the user, and keep the control of the conversation on the Jason side: "what to say, why, and when" is the result of a Jason-driven rational process based on the users intentions, that we associate with deep thinking. Our AAMAS 2025 work on ChatBDI [43] provides an integration of BDI agents and LLMs that may serve as talk deep in its default implemented setting, where sentences by the human user are sent to agents for reasoning, and answers are sent to LLMs for being properly expressed in natural language. The reasoning stage might be however bypassed, for a talk small conversation. The current version of ChatBDI available at `https://github.com/VEsNA-ToolKit/chatbdi` is implemented using Jason, JaCaMo, Nomic-embed-text [70], and CodeGemma [106]. KQML [40], is used as intermediate language between agents and LLMs. ChatBDI 'chattifies' BDI agents by equipping them with LLM-based 'language actuators'. The work by Frering et al. [41] is similar to ChatBDI, but lacks its generality.

Talk Small – ***Function:*** Experiments from psychologists, sociologists and neuroscientists show that small talk with "weak ties" generates well-being [5,88,86]. ***Major anatomical structure involved.*** Not explored; we name it "TalkSmall area". ***Feasible implementation in Jason:*** In the Generative AI and Large

[4] `https://dictionary.cambridge.org/dictionary/english/small-talk`.

[5] `https://www.linkedin.com/pulse/why-deep-meaningful-conversations-important-ray-williams-mpjbc/`.

Language Models era, our vision of talking small is "talking as an LLM would talk". While this does not mean that an interaction with an LLM always looks like being shallow, or "chit-chat", LLMs are not reasoners [55], and are not even speakers because they lack goals and intentions [45,74,83]. We claim that not being intention-driven speakers prevents LLMs from talking deep. Still, they generate very fluent and believable sentences, making them suitable for talking small. Hence, a feasible Jason implementation would integrate LLM-based generation of sentences as actions that agents may perform without needing to "reason on what to say" and, most importantly, without needing to recall the contents of the conversation. A restricted version of ChatBDI, let us name it ChatBDI$^-$, may serve this purpose, if we just suppress the delivery of user messages to the BDI brain, and we only use the BDI infrastructure to interface users and LLMs, inside a MAS. It can be however used for a double purpose, switching between talking deep and talking small depending on the classification of the user's sentence as 'serious' or 'chit-chat'. Besides ChatBDI, in [51], Ichida et al. exploit LLMs and reinforcement learning to bootstrap the reasoning capabilities of NatBDI agents, which is not what we need. In [80], Ricci et al. envision generative BDI architectures, namely architectures based on the BDI model integrating generative AI technologies, but no implemented integration in Jason is available.

4 Conclusions

Albeit just sketched, all the MEDiTATe modules shown in Figure 1 are rooted on neuro-scientific or agent-oriented literature; however, the main challenge in implementing MEDiTATe is not in the implementation of its components, but *in fully understanding their connections* (on the neuro-scientific side) and *in seamlessly integrating them* (from the agent-oriented software engineering side). As the title of the paper says, this is a first step in the journey of bridging neuroscience outcomes and achievements in the BDI research field. A first – even small – step is always needed to start a journey and, to the best of our knowledge, no systematic analysis of the BDI and neuro-scientific literature had been carried out so far. While accommodating results from neuroscience into the MEDiTATe conceptual framework, we aimed at achieving two different goals: making the MEDiTATe vision more coherent, and looking for gaps in the neuro-scientific literature, where some implemented tools exist that have no counter-part in the brain. This is what happened with deep and small talk.

Despite the many open challenges, we believe that MEDiTATe may represent a framework where achievements from experts in different disciplins can find a natural positioning, and may offer a more controllable, explainable, and transparent approach for testing those hypotheses than emerging in silico experimentation using deep learning-based encoding models [53].

Acknowledgments. This work was partially supported by *ENGINES – ENGineering INtElligent Systems around intelligent agent technologies*, funded by the Italian MUR program PRIN 2022 under grant number 20229ZXBZM.

References

1. Airiau, S., Padgham, L., Sardiña, S., Sen, S.: Enhancing the adaptation of BDI agents using learning techniques. Int. J. Agent Technol. Syst. **1**(2), 1–18 (2009). https://doi.org/10.4018/JATS.2009040101, https://doi.org/10.4018/jats.2009040101
2. Alfonso, B., Vivancos, E., Botti, V.J.: Toward formal modeling of affective agents in a BDI architecture. ACM Trans. Internet Techn. **17**(1), 5:1–5:23 (2017)
3. Apperly, I.A.: What is "theory of mind"? concepts, cognitive processes and individual differences. Quarterly journal of experimental psychology **65**(5), 825–839 (2012)
4. Argente, E., del Val Noguera, E., Pérez-García, D., Botti, V.J.: Normative emotional agents: A viewpoint paper. IEEE Trans. Affect. Comput. **13**(3), 1254–1273 (2022)
5. Ascigil, E., Gunaydin, G., Selcuk, E., Sandstrom, G.M., Aydin, E.: Minimal social interactions and life satisfaction: The role of greeting, thanking, and conversing. Social Psychological and Personality Science (2023). https://doi.org/10.1177/19485506231209793, https://doi.org/10.1177/19485506231209793
6. Baddeley, A.: Working memory: Theories, models, and controversies. Annual Review of Psychology Volume 63, 2012 **63**, 1–29 (2012). https://doi.org/https://doi.org/10.1146/annurev-psych-120710-100422
7. Badica, A., Badica, C., Ivanovic, M., Mitrovic, D.: An approach of temporal difference learning using agent-oriented programming. In: 20th International Conference on Control Systems and Computer Science, CSCS 2015, Bucharest, Romania, May 27-29, 2015. pp. 735–742. IEEE (2015). https://doi.org/10.1109/CSCS.2015.71, https://doi.org/10.1109/CSCS.2015.71
8. Badica, C., Becheru, A., Felton, S.: Integration of jason reinforcement learning agents into an interactive application. In: Jebelean, T., Negru, V., Petcu, D., Zaharie, D., Ida, T., Watt, S.M. (eds.) 19th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC 2017, Timisoara, Romania, September 21-24, 2017. pp. 361–368. IEEE Computer Society (2017). https://doi.org/10.1109/SYNASC.2017.00065, https://doi.org/10.1109/SYNASC.2017.00065
9. Berkman, E.T.: The neuroscience of goals and behavior change. Consulting Psychology Journal: Practice and Research **70**(1), 28 (2018)
10. Berridge, K.C., Kringelbach, M.L.: Pleasure systems in the brain. Neuron **86**(3), 646–664 (2015). https://doi.org/10.1016/j.neuron.2015.02.018, https://pmc.ncbi.nlm.nih.gov/articles/PMC4425246/
11. Boissier, O., Bordini, R.H., Hübner, J.F., Ricci, A.: Multi-agent oriented programming: programming multi-agent systems using JaCaMo. MIT Press (2020)
12. Bordini, R.H., Hübner, J.F., Wooldridge, M.J.: Programming multi-agent systems in AgentSpeak using Jason. J. Wiley (2007)
13. Bosello, M., Ricci, A.: From programming agents to educating agents - A jason-based framework for integrating learning in the development of cognitive agents. In: Dennis, L.A., Bordini, R.H., Lespérance, Y. (eds.) Engineering Multi-Agent Systems - 7th International Workshop, EMAS 2019, Montreal, QC, Canada, May 13-14, 2019, Revised Selected Papers. Lecture Notes in Computer Science, vol. 12058, pp. 175–194. Springer (2019). https://doi.org/10.1007/978-3-030-51417-4_9, https://doi.org/10.1007/978-3-030-51417-4_9

14. Bosse, T., Memon, Z.A., Treur, J.: A recursive BDI agent model for theory of mind and its applications. Appl. Artif. Intell. **25**(1), 1–44 (2011)
15. Brass, M., Haggard, P.: The what, when, whether model of intentional action. The Neuroscientist **14**(4), 319–325 (2008). https://doi.org/10.1177/1073858408317417, https://doi.org/10.1177/1073858408317417, pMID: 18660462
16. Bratman, M.: Intention, plans, and practical reason (1987)
17. Brentano, F.: Psychology From an Empirical Standpoint. Routledge (1874)
18. Budson, A.E., Price, B.H.: Memory dysfunction. New England Journal of Medicine **352**(7), 692–699 (2005). https://doi.org/10.1056/NEJMra041071, https://www.nejm.org/doi/full/10.1056/NEJMra041071
19. Cantucci, F., Falcone, R.: A computational model for cognitive human-robot interaction: An approach based on theory of delegation. In: WOA. CEUR Workshop Proceedings, vol. 2404, pp. 127–133. CEUR-WS.org (2019)
20. Cardoso, R.C., Ferrando, A., Papacchini, F.: Automated planning and BDI agents: A case study. In: PAAMS. Lecture Notes in Computer Science, vol. 12946, pp. 52–63. Springer (2021)
21. Carrington, S.J., Bailey, A.J.: Are there theory of mind regions in the brain? a review of the neuroimaging literature. Human brain mapping **30**(8), 2313–35 (2009). https://doi.org/doi:10.1002/hbm.20671
22. Casali, A., Godo, L., Sierra, C.: Graded BDI models for agent architectures. In: CLIMA. Lecture Notes in Computer Science, vol. 3487, pp. 126–143. Springer (2004)
23. Casali, A., Godo, L., Sierra, C.: A graded BDI agent model to represent and reason about preferences. Artif. Intell. **175**(7-8), 1468–1478 (2011)
24. Casali, A., Godo, L., Sierra, C.: A language for the execution of graded BDI agents. Log. J. IGPL **21**(3), 332–354 (2013)
25. Chella, A., Lanza, F., Seidita, V.: Decision process in human-agent interaction: Extending Jason reasoning cycle. In: EMAS@AAMAS. Lecture Notes in Computer Science, vol. 11375, pp. 320–339. Springer (2018)
26. Cooney, G., Boothby, E.J., Lee, M.: The thought gap after conversation: Underestimating the frequency of others' thoughts about us. Journal of experimental psychology. General **151**(5), 1069–1088 (2022). https://doi.org/https://doi.org/10.1037/xge0001134
27. Cosmides, L., Tooby, J.: Evolutionary psychology and the emotions (2000)
28. Costantini, S., De Gasperis, G., Migliarini, P., Salutari, A.: Proposal of a empathic multi-agent robot design based on theory of mind. Proceedings of cAESAR (2020)
29. Daniel, K.: Thinking, fast and slow. Farrar, Straus and Giroux (2011)
30. Dautenhahn, K.: Getting to know each other - artificial social intelligence for autonomous robots. Robotics Auton. Syst. **16**(2-4), 333–356 (1995)
31. Dautenhahn, K.: Socially intelligent agents and the primate social brain - towards a science of social minds (1999), aAAI Technical Report FS-00-04
32. De Silva, L., Padgham, L.: Planning on demand in BDI systems. In: ICAPS. pp. 37–40 (2005)
33. Dennett, D.C.: The Intentional Stance. The MIT Press, Cambridge, MA (1987)
34. Dennis, L.A., Oren, N.: Explaining BDI agent behaviour through dialogue. Auton. Agents Multi Agent Syst. **36**(1), 29 (2022)
35. Elfenbein, H.A., Ambady, N.: On the universality and cultural specificity of emotion recognition: a meta-analysis. Psychological bulletin **128**(2), 203 (2002)

36. Engelmann, D.C., Cezar, L.D., Panisson, A.R., Bordini, R.H.: A conversational agent to support hospital bed allocation. In: BRACIS (1). Lecture Notes in Computer Science, vol. 13073, pp. 3–17. Springer (2021)
37. Engelmann, D.C., Damasio, J., Krausburg, T., Borges, O.T., Cezar, L.D., Panisson, A.R., Bordini, R.H.: Dial4jaca - A demonstration. In: PAAMS. Lecture Notes in Computer Science, vol. 12946, pp. 346–350. Springer (2021)
38. Engelmann, D.C., Damasio, J., Krausburg, T., Borges, O.T., da Silveira Colissi, M., Panisson, A.R., Bordini, R.H.: Dial4jaca - A communication interface between multi-agent systems and chatbots. In: PAAMS. Lecture Notes in Computer Science, vol. 12946, pp. 77–88. Springer (2021)
39. Engelmann, D.C., Panisson, A.R., Vieira, R., Hübner, J.F., Mascardi, V., Bordini, R.H.: MAIDS - A framework for the development of multi-agent intentional dialogue systems. In: AAMAS. pp. 1209–1217. ACM (2023)
40. Finin, T.W., Fritzson, R., McKay, D.P., McEntire, R.: KQML as an agent communication language. In: Proceedings of the Third International Conference on Information and Knowledge Management (CIKM'94), Gaithersburg, Maryland, USA, November 29 - December 2, 1994. pp. 456–463. ACM (1994). https://doi.org/10.1145/191246.191322, https://doi.org/10.1145/191246.191322
41. Frering, L., Steinbauer-Wagner, G., Holzinger, A.: Integrating belief-desire-intention agents with large language models for reliable human-robot interaction and explainable artificial intelligence. Eng. Appl. Artif. Intell. **141**, 109771 (2025)
42. Gatti, A., Mascardi, V.: Vesna, a framework for virtual environments via natural language agents and its application to factory automation. Robotics **12**(2), 46 (2023)
43. Gatti, A., Mascardi, V., Ferrando, A.: ChatBDI: Think BDI, Talk LLM. In: AAMAS. International Foundation for Autonomous Agents and Multiagent Systems / ACM (2025)
44. Google: Dialogflow (2017), https://cloud.google.com/dialogflow/, accessed on April 22, 2025
45. Gubelmann, R.: Large language models, agency, and why speech acts are beyond them (for now)–a kantian-cum-pragmatist case. Philosophy & Technology **37**(1), 32 (2024)
46. Harbers, M., van den Bosch, K., Meyer, J.C.: Modeling agents with a theory of mind. In: IAT. pp. 217–224. IEEE Computer Society (2009)
47. Hassabis, D., Kumaran, D., Summerfield, C., Botvinick, M.: Neuroscience-inspired artificial intelligence. Neuron **95**(2), 245–258 (2017). https://doi.org/https://doi.org/10.1016/j.neuron.2017.06.011, https://www.sciencedirect.com/science/article/pii/S0896627317305093
48. Ho, M.K., Saxe, R., Cushman, F.: Planning with Theory of Mind. Trends in Cognitive Sciences **26**(11), 959–971 (2022). https://doi.org/https://doi.org/10.1016/j.tics.2022.08.003, https://www.sciencedirect.com/science/article/pii/S1364661322001851
49. Hwang, H., Matsumoto, D.: Functions of emotions. Noba textbook series: Psychology (2018)
50. Ichida, A.Y., Meneguzzi, F.: Modeling a conversational agent using BDI framework. In: Hong, J., Lanperne, M., Park, J.W., Cerný, T., Shahriar, H. (eds.) Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing, SAC 2023, Tallinn, Estonia, March 27-31, 2023. pp. 856–863. ACM (2023). https://doi.org/10.1145/3555776.3577657, https://doi.org/10.1145/3555776.3577657

51. Ichida, A.Y., Meneguzzi, F., Cardoso, R.C.: BDI agents in natural language environments. In: AAMAS. pp. 880–888. International Foundation for Autonomous Agents and Multiagent Systems / ACM (2024)
52. Jahanshahi, M.: Willed action and its impairments. Cognitive neuropsychology **15**(6-8), 483–533 (1998)
53. Jain, S., Vo, V.A., Wehbe, L., Huth, A.G.: Computational language modeling and the promise of in silico experimentation. Neurobiology of Language **5**(1), 80–106 (04 2024). https://doi.org/10.1162/nol_a_00101, https://doi.org/10.1162/nol_a_00101
54. Jiao, L., Ma, M., He, P., Geng, X., Liu, X., Liu, F., Ma, W., Yang, S., Hou, B., Tang, X.: Brain-inspired learning, perception, and cognition: A comprehensive review. IEEE Transactions on Neural Networks and Learning Systems pp. 1–21 (2024). https://doi.org/10.1109/TNNLS.2024.3401711
55. Kambhampati, S.: Can Large Language Models reason and plan? Annals of the New York Academy of Sciences **1534**(1), 15–18 (2024). https://doi.org/10.1111/nyas.15125, http://dx.doi.org/10.1111/nyas.15125
56. Kardas, M., Kumar, A., Epley, N.: Overly shallow?: Miscalibrated expectations create a barrier to deeper conversation. Journal of personality and social psychology **122**(3), 367–398 (2022). https://doi.org/10.1037/pspa0000281, https://pubmed.ncbi.nlm.nih.gov/34591541/
57. Ledoux, J.E.: Cognitive-emotional interactions in the brain. Cognition and Emotion **3**(4), 267–289 (1989). https://doi.org/10.1080/02699938908412709, https://doi.org/10.1080/02699938908412709
58. LeDoux, J.E.: The emotional brain: The mysterious underpinnings of emotional life. Simon and Schuster (1998)
59. Lee, S., Son, Y.J.: Dynamic learning in human decision behavior for evacuation scenarios under bdi framework. In: Proceedings of the 2009 INFORMS Simulation Society Research Workshop. INFORMS Simulation Society: Catonsville, MD. pp. 96–100 (2009)
60. Lokuge, P., Alahakoon, D.: Handling multiple events in hybrid BDI agents with reinforcement learning: A container application. In: Chen, C., Filipe, J., Seruca, I., Cordeiro, J. (eds.) ICEIS 2005, Proceedings of the Seventh International Conference on Enterprise Information Systems, Miami, USA, May 25-28, 2005. pp. 83–90 (2005)
61. Lokuge, P., Alahakoon, D.: Reinforcement learning in neuro BDI agents for achieving agent's intentions in vessel berthing applications. In: 19th International Conference on Advanced Information Networking and Applications (AINA 2005), 28-30 March 2005, Taipei, Taiwan. pp. 681–686. IEEE Computer Society (2005). https://doi.org/10.1109/AINA.2005.293, https://doi.org/10.1109/AINA.2005.293
62. Matsumoto, D., Hwang, H.C.: Assessing cross-cultural competence: A review of available tests. Journal of cross-cultural psychology **44**(6), 849–873 (2013)
63. Mattar, M.G., Lengyel, M.: Planning in the brain. Neuron **110**(6), 914–934 (2022). https://doi.org/https://doi.org/10.1016/j.neuron.2021.12.018, https://www.sciencedirect.com/science/article/pii/S0896627321010357
64. McCulloch, W., Pitts, W.: A logical calculus of the ideas immanent to nervous activity. The Bulletin of Mathematical Biophysics **5**(4), 115–133 (1943)
65. Meneguzzi, F., de Silva, L.: Planning in BDI agents: a survey of the integration of planning algorithms and agent reasoning. Knowl. Eng. Rev. **30**(1), 1–44 (2015)

66. Molla, H., Smadi, S., Lyubomirsky, S., Li, T., de Wit, H.: Immediate and enduring effects of deep and shallow conversations on feelings of closeness in healthy adults (2022), https://doi.org/10.31234/osf.io/p62va
67. Montes, N., Luck, M., Osman, N., Rodrigues, O., Sierra, C.: Combining theory of mind and abductive reasoning in agent-oriented programming. Auton. Agents Multi Agent Syst. **37**(2), 36 (2023)
68. Mustapha, A., Ahmad, M.S., Ahmad, A.: Conversational agents as full-pledged BDI agents for ambient intelligence. In: ISAmI. Advances in Intelligent Systems and Computing, vol. 219, pp. 221–228. Springer (2013)
69. Norling, E.: Folk psychology for human modelling: Extending the BDI paradigm. In: 3rd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2004), 19-23 August 2004, New York, NY, USA. pp. 202–209. IEEE Computer Society (2004). https://doi.org/10.1109/AAMAS.2004.10066, https://doi.ieeecomputersociety.org/10.1109/AAMAS.2004.10066
70. Nussbaum, Z., Morris, J.X., Duderstadt, B., Mulyar, A.: Nomic Embed: Training a reproducible long context text embedder (2024)
71. Panisson, A.R., Sarkadi, S., McBurney, P., Parsons, S., Bordini, R.H.: On the formal semantics of theory of mind in agent communication. In: AT. Lecture Notes in Computer Science, vol. 11327, pp. 18–32. Springer (2018)
72. Parmiggiani, M., Ferrando, A., Mascardi, V.: Together is better! Integrating BDI and RL agents for safe learning and effective collaboration. In: ICAART. p. 12. SCITEPRESS (2025)
73. Pereira, D., Oliveira, E., Moreira, N., Sarmento, L.: Towards an architecture for emotional BDI agents. In: 2005 portuguese conference on artificial intelligence. pp. 40–46 (2005). https://doi.org/10.1109/EPIA.2005.341262
74. Piwek, P.: Are conversational large language models speakers? In: Proc. of the 28th Workshop on the Semantics and Pragmatics of Dialogue - Poster Abstracts (2024)
75. Puica, M.A., Florea, A.M.: Emotional belief-desire-intention agent model: Previous work and proposed architecture. International Journal of Advanced Research in Artificial Intelligence **2**(2), 1–8 (2013)
76. Pulawski, S., Dam, H.K., Ghose, A.: Bdi-dojo: developing robust BDI agents in evolving adversarial environments. In: El-Araby, E., Kalogeraki, V., Pianini, D., Lassabe, F., Porter, B., Ghahremani, S., Nunes, I., Bakhouya, M., Tomforde, S. (eds.) IEEE International Conference on Autonomic Computing and Self-Organizing Systems, ACSOS 2021, Companion Volume, Washington, DC, USA, September 27 - Oct. 1, 2021. pp. 257–262. IEEE (2021). https://doi.org/10.1109/ACSOS-C52956.2021.00066, https://doi.org/10.1109/ACSOS-C52956.2021.00066
77. Qi, G., Bo-ying, W.: Study and application of reinforcement learning in cooperative strategy of the robot soccer based on bdi model. International Journal of Advanced Robotic Systems **6**(2), 15 (2009). https://doi.org/10.5772/6795, https://doi.org/10.5772/6795
78. Rao, A.S.: AgentSpeak(L): BDI agents speak out in a logical computable language. In: 7th European Workshop on Modelling Autonomous Agents in a Multi-Agent World, Eindhoven, The Netherlands, January 22-25, 1996. Lecture Notes in Computer Science, vol. 1038, pp. 42–55. Springer (1996). https://doi.org/10.1007/BFb0031845, https://doi.org/10.1007/BFb0031845
79. Rao, A.S., Georgeff, M.P.: BDI agents: From theory to practice. In: ICMAS. pp. 312–319. The MIT Press (1995)

80. Ricci, A., Mariani, S., Zambonelli, F., Burattini, S., Castelfranchi, C.: The cognitive hourglass: Agent abstractions in the large models era. In: AAMAS. pp. 2706–2711. International Foundation for Autonomous Agents and Multiagent Systems / ACM (2024)
81. Rocha, M., da Silva, H.H., Morales, A.S., Sarkadi, S., Panisson, A.R.: Applying theory of mind to multi-agent systems: A systematic review. In: BRACIS (1). Lecture Notes in Computer Science, vol. 14195, pp. 367–381. Springer (2023)
82. Rolls, E.T.: On the brain and emotion. Behavioral and Brain Sciences **23**(2), 219–228 (2000). https://doi.org/10.1017/S0140525X00512424
83. Rosen, Z.P., Dale, R.: LLMs don't "do things with words" but their lack of illocution can inform the study of human discourse. In: Proceedings of the 46th Annual Meeting of the Cognitive Science Society. pp. 2870–2876 (2024)
84. Rutten, G.J.: Chapter 2 - Broca-Wernicke theories: A historical perspective. In: Hillis, A.E., Fridriksson, J. (eds.) Aphasia, Handbook of Clinical Neurology, vol. 185, pp. 25–34. Elsevier (2022). https://doi.org/https://doi.org/10.1016/B978-0-12-823384-9.00001-3, https://www.sciencedirect.com/science/article/pii/B9780128233849000013
85. Sánchez, Y., Cerezo, E.: Designing emotional BDI agents: good practices and open questions. Knowl. Eng. Rev. **34**, e26 (2019)
86. Sandstrom, G.M., Dunn, E.W.: Is efficiency overrated?: Minimal social interactions lead to belonging and positive affect. Social Psychological and Personality Science **5**(4), 437–442 (2014). https://doi.org/10.1177/1948550613502990, https://doi.org/10.1177/1948550613502990
87. Sarkadi, S., Panisson, A.R., Bordini, R.H., McBurney, P., Parsons, S.: Towards an approach for modelling uncertain theory of mind in multi-agent systems. In: AT. Lecture Notes in Computer Science, vol. 11327, pp. 3–17. Springer (2018)
88. Schroeder, J., Lyons, D., Epley, N.: Hello, stranger? pleasant conversations are preceded by concerns about starting one. Journal of experimental psychology. General **151**(5), 1141–1153 (2022). https://doi.org/10.1037/xge0001118, https://doi.org/10.1037/xge0001118
89. Seidita, V., Sabella, A.M.P., Lanza, F., Chella, A.: Agent talks about itself: an implementation using Jason, CArtAgO and speech acts. Intelligenza Artificiale **17**(1), 7–18 (2023)
90. Sgorbissa, A., Morocutti, L., D'Angelo, I., Recchiuto, C.T.: Machiavellian robots and their theory of mind. IEEE Transactions on Affective Computing pp. 1–18 (2024). https://doi.org/10.1109/TAFFC.2024.3494595
91. de Silva, L., Sardiña, S., Padgham, L.: First principles planning in BDI systems. In: AAMAS (2). pp. 1105–1112. IFAAMAS (2009)
92. Squire, L.R.: Memory systems of the brain: A brief history and current perspective. Neurobiology of Learning and Memory **82**(3), 171–177 (2004). https://doi.org/https://doi.org/10.1016/j.nlm.2004.06.005, https://www.sciencedirect.com/science/article/pii/S1074742704000735, multiple Memory Systems
93. Steunebrink, B.R., Dastani, M., Meyer, J.C.: A formal model of emotion triggers: an approach for BDI agents. Synth. **185**(Supplement-1), 83–129 (2012)
94. Sutton, R.S., Barto, A.G.: Reinforcement learning - an introduction. Adaptive computation and machine learning, MIT Press (1998), https://www.worldcat.org/oclc/37293240
95. Tan, A., Ong, Y., Tapanuj, A.: A hybrid agent architecture integrating desire, intention and reinforcement learning. Expert Syst. Appl. **38**(7), 8477–8487

(2011). https://doi.org/10.1016/J.ESWA.2011.01.045, https://doi.org/10.1016/j.eswa.2011.01.045
96. Taverner, J., Alfonso, B., Vivancos, E., Botti, V.J.: Integrating expectations into jason for appraisal in emotion modeling. In: IJCCI (ECTA). pp. 231–238. SciTePress (2016)
97. Traylor, A., Merullo, J., Frank, M.J., Pavlick, E.: Transformer mechanisms mimic frontostriatal gating operations when trained on human working memory tasks. CoRR **abs/2402.08211** (2024)
98. Tremblay, P., Dick, A.S.: Broca and Wernicke are dead, or moving past the classic model of language neurobiology. Brain and Language **162**, 60–71 (2016). https://doi.org/https://doi.org/10.1016/j.bandl.2016.08.004, https://www.sciencedirect.com/science/article/pii/S0093934X16300475
99. Walczak, A., Braubach, L., Pokahr, A., Lamersdorf, W.: Augmenting BDI agents with deliberative planning techniques. In: PROMAS. Lecture Notes in Computer Science, vol. 4411, pp. 113–127. Springer (2006)
100. Wan, Q., Liu, W., Xu, L., Guo, J.: Extending the BDI model with q-learning in uncertain environment. In: Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence, ACAI 2018, Sanya, China, December 21-23, 2018. pp. 33:1–33:6. ACM (2018). https://doi.org/10.1145/3302425.3302432, https://doi.org/10.1145/3302425.3302432
101. Wang, Q., Ross, M.: Culture and memory. Handbook of cultural psychology **18**, 645–667 (2007)
102. Wilkes, M.V.: Slave memories and dynamic storage allocation. IEEE Trans. Electron. Comput. **14**(2), 270–271 (1965)
103. Wong, W., Cavedon, L., Thangarajah, J., Padgham, L.: Flexible conversation management using a BDI agent approach. In: IVA. Lecture Notes in Computer Science, vol. 7502, pp. 464–470. Springer (2012)
104. Xu, M., Bauters, K., McAreavey, K., Liu, W.: A formal approach to embedding first-principles planning in BDI agent systems. In: SUM. Lecture Notes in Computer Science, vol. 11142, pp. 333–347. Springer (2018)
105. Yan, E., Burattini, S., Hübner, J.F., Ricci, A.: Towards a multi-level explainability framework for engineering and understanding BDI agent systems. In: WOA. CEUR Workshop Proceedings, vol. 3579, pp. 216–231. CEUR-WS.org (2023)
106. Zhao, H., Hui, J., Howland, J., Nguyen, N., Zuo, S., Hu, A., Choquette-Choo, C.A., Shen, J., Kelley, J., Bansal, K., Vilnis, L., Wirth, M., Michel, P., Choy, P., Joshi, P., Kumar, R., Hashmi, S., Agrawal, S., Gong, Z., Fine, J., Warkentin, T., Hartman, A.J., Ni, B., Korevec, K., Schaefer, K., Huffman, S.: CodeGemma: Open code models based on gemma. CoRR **abs/2406.11409** (2024)