The development of autonomous agents is a central goal of artificial intelligence. A salient feature of autonomous agents is their ability to exhibit goal-directed behaviour, i.e., to commit to goals and search for plans to attain them. In order to plan, an agent must think about the possible outcomes of its actions and make the best choice about what to do. But to think before acting, an agent needs an internal representation of its world; a mental simulation of it, whose manipulation serves as a substitute for action.

This thesis is concerned with (i) learning such internal representations from experience and (ii) planning with them.

Regarding learning, we consider an agent exposed to a partially observable domain, with which the agent has never interacted before, and about which the agent wishes to learn both what she can observe and how her actions can affect it. A partially observable environment is one in which the entire state of the environment is not visible to the agent at all times. The world itself, as perceived by humans, is an excellent example of a partially observable environment; what we observe is limited by our visual field. We assume that the agent can learn about this domain from experience gathered by taking actions in the domain and observing their results. We present learning algorithms capable of learning as much as possible (in a well-defined sense) about both what is directly observable and what actions do in the domain, given the learner's observational constraints. We distinguish the levels of domain knowledge attained by each algorithm, and characterize the type of observations required to reach such knowledge. The algorithms use dynamic epistemic logic (DEL) to represent the learned domain information symbolically. The presented work extends previous research by Bolander and Gierasimczuk (2018), which developed learning algorithms based on DEL to learn in fully observable domains.

Regarding planning, we consider an agent that already has a representation of its environment. The agent is assumed to inhabit a social, multi-agent world. In order to plan in such a world, the agent needs to take into account, not only her own capabilities and knowledge, but also the capabilities and knowledge of other agents. This type of planning requires theory-of-mind (ToM) reasoning: the ability to reason about how the world is perceived by others, what they believe and what they intend to do. We introduce a new model for multi-agent planning, supporting ToM reasoning, which we call first-order epistemic planning. Epistemic planning using propositional DEL was proposed by Bolander and Andersen (2011), and first-order epistemic planning is an extension of this framework. FODEL allows for more compact representations of multi-agent worlds than regular DEL. FODEL also adds to DEL the ability to represent abstract knowledge in a natural way. We show that FODEL satisfies a number of desirable technical properties for representation languages (soundness, completeness, and decidability over worlds with finitely many agents). We then study first-order epistemic planning problems, and show that, in many cases, we can decide whether such a planning problem can be solved or not. In other words, we show that the plan existence problem for relevant types of first-order epistemic planning problems is decidable.

Finally, we also study FODEL as a formalism to describe social network dynamics involving ToM reasoning. We show how several network dynamics considered in the literature are modelled naturally in this framework, and examine the expressivity of FODEL relative to other network dynamics formalisms.