Plans, Planning, and Intention

[Redacted]
[Redacted]

Planning is central to agency. Bratman has proposed a well-known account of the relationship between planning and agency where plans and intention stand in distinctive relations as part of a larger agential architecture [1]. Here, we critique Bratman's approach from the perspective of reinforcement learning (RL), focusing on two of his main claims: the formation of intentions (i) precede planning and terminate deliberation and (ii) prompts a switch from reasoning about ends to reasoning about means.

We contend that RL, a widespread influential computational approach in cognitive science, puts pressure on both theses. In RL, agents make choices and observe their outcomes. Any mismatches between outcomes and expectations are used to update the expected outcomes for subsequent decisions. Agents learn the value of actions via these updates to use in selecting actions in the future. In model-based RL, agents also learn the probability of transitioning between states, such as the probability that state S2 occurs if the agent performs action A in state S1. Planning involves using this model to simulate a series of actions off-line. Given the transitions and associated probabilities, the simulations allow the agent to calculate the expected outcome of potential actions. In this framework, action values are always considered in the context of a model of the environment. Different environments involve distinct contingencies that change the expected value of these actions. To make a decision, the environmental model is used during those simulations, calculating the expected outcomes as part of the decision process.

Planning behaviors are explained by the RL framework. For instance, in multi-stage tasks [2], agents make a series of choices, where choices in the first stage are probabilistically related to what options will be available in the second stage. If agents are using a model to make decisions via simulation, their first-stage choices will reflect the overall reward history for that choice, the contingencies of the second-stage choice, and the probability of the transition between the stages. That is, the initial choice will be considered in the context of the probability of ending up in the second-stage and which options will then be available. Changing the probabilities in the action-state transition or modifying the value of the second-stage choice outcomes will be reflected in the probability that the agent makes a particular choice at the first stage. The influence of first-state outcomes, transition probabilities, and second-stage rewards is a kind of result frequently observed in complex choice tasks [3-6].

We argue that sensitivity to environmental transitions and to differences between actions depending on states, outcomes, and transitions is evidence that agents are planning their actions through simulation in a model. Given the explanatory power of this framework, we suggest the following thesis:

(P-RL): Planning is simulation of outcomes from future decisions using a model of the environment.

This (P-RL) thesis captures the idea that decision making is part of planning in virtue of the processes involved in simulation.

Bratman's commitments are challenged by (P-RL). Bratman's first claim is that intentions terminate deliberation. On (P-RL), planning is deliberation via simulation, involving beliefs about probabilities of transitions between states of the environment due to actions and their desirable outcomes. But then, intentions at best are simultaneous to planning and deliberation. The second claim is that intentions prompt a switch from thinking about ends to thinking about means. On (P-RL), the expected values of options are computed by simulating outcomes using the agent's model, which comprises series of state-action transitions that probabilistically lead to outcomes. Since these are representations of means on which decision-making operates, the (P-RL) thesis implies selection of means and ends simultaneously. This includes simulation at different grains, such as choosing to get a drink and choosing which arm angles and trajectories to select to pick it up, where such fine-grained actions are gathered together as 'options' that enter into the simulation [7].

P-RL is a major shift from standard philosophical thinking about plans. Besides undermining Bratman's account, this shift has significant implications for cognitive ontology, the norms of planning and action, and cognitive control. We end by outlining these briefly.

1. Bratman, M., *Intention, plans, and practical reason*. 1987: Harvard University Press.
2. Doll, B.B., et al., *Model-based choices involve prospective neural activity.* Nature neuroscience, 2015. **18**(5): p. 767-772.
3. Doll, B.B., et al., *The ubiquity of model-based reinforcement learning*. Curr Opin Neurobiol, 2012. **22**(6): p. 1075-81.

4. Gilbert, D.T. and T.D. Wilson, *Prospection: Experiencing the future.* Science, 2007. **317**(5843): p. 1351-1354.
5. Rangel, A., et al., *A framework for studying the neurobiology of value-based decision making.* Nat Rev Neurosci, 2008. **9**(7): p. 545-56.
6. Redish, A.D., *Vicarious trial and error.* Nature Reviews Neuroscience, 2016. **17**(3): p. 147-159.
7. Sutton, R.S., et al., *Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning.* Artificial intelligence, 1999. **112**(1-2): p. 181-211.