Thinking before acting: intentions, logic, rational choice
Roy, O.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Abstract

In this thesis I propose a theory of decision making that does justice to the idea that human agents can form future-directed intentions, but which at the same time capitalizes on the resources of contemporary theories of instrumental rationality and dynamic epistemic logic. The result is a more all-encompassing picture of practical reasoning for planning agents. I show that such a broad approach genuinely enriches existing models of rational decision making, as well as the philosophical theory of intentions.

In Chapter 2 I show that the introduction of future-directed intentions does indeed broaden the explanatory scope of decision-theoretic models. The volitive commitment of future-directed intentions allows one to go beyond traditional decision-theoretic reasoning by “breaking ties” between equally desirable options, and thus provides a straightforward anchor for personal coordination.

In Chapter 3 I consider coordination, mostly in “Hi-Lo” games. I show that intentions do indeed anchor coordination in these games, in a way that naturally generalizes their “tie-breaking” effect in single agent contexts. At the end of the chapter I look at how intentions can anchor coordination in the general case. This allows to revisit important claims in the planning theory concerning “shared agency”, and in particular to circumscribe better the extent of this phenomenon.

In Chapter 4 I turn to two facets of the reasoning-centered commitment of intentions, namely the filtering of options and the focus on means. I show that they can be studied by means of two simple operations which transform decision- and game-theoretic models. In contexts of strategic interaction, these operations acquire an important social character, that has not yet been studied in philosophy of action.

In Chapter 5 I use dynamic epistemic logic to bring the considerations of the previous chapters under a single umbrella. I show that an important aspect of the volitive commitment used to account for coordination with intentions has an echo in the filtering of options that I define in Chapter 4. This observation triggers a natural generalization of the idea of filtering, which takes into account
the information that agents have about their own intentions and the intentions of others. By the end of the chapter I explore two other issues at the intersection of planning agency and instrumental rationality, namely the condition under which intention-based transformations of decision problems foster coordination and become "enabled" by the elimination of dominated strategies.

In Chapter 6 I look back at this theory from a philosophical point of view, and investigate the question of how the norms of consistency and coherence which apply to intentions can be explained. In contemporary philosophy of action there are two main takes on this issue, called the "cognitivist" and "agency" approaches. Here I explore an alternative one, hybrid pragmatism, which stands half-way between cognitivism and the agency approach. It is based on the notion of "acceptance in deliberation", a cognitive state which has so far attracted little attention. I argue that hybrid pragmatism is a plausible alternative to the two main contemporary approaches, and that its use of acceptances provides a more complete picture of how future-directed intentions make their way into practical reasoning.