Revising Beliefs and Intentions in Stochastic Environments

Nima Motamed[1], Natasha Alechina[2,1], Mehdi Dastani[1], and Dragan Doder[1]

[1] Utrecht University, Utrecht, The Netherlands
[2] Open University, Heerlen, The Netherlands

Autonomous agents operating in dynamic environments need a method for keeping track of their beliefs, updating these as new information is received. But an agent also carries *intentions* they are committed to bring about, and these also need to be updated on the basis of new information or deliberation. While there is an enormous amount of literature on the former topic of *belief revision*, the topic of *intention revision* received, in comparison, less attention [2, 18, 9, 1, 13, 16, 6, 10, 4, 15, 19, 17].

In [14], we presented the *first* framework for the joint revision of beliefs and intentions in stochastic environments. We defined beliefs and intentions in a probabilistic temporal logic interpreted on Markov Decision Processes (MDPs), giving the ability to express beliefs about the uncertainty in actions' outcomes, as well as complex temporal intentions. Since the paper's publication, we have obtained sharper complexity results for the logic's satisfiability problem: it is contained in EXPSPACE, and is NEXP-hard. We in fact conjecture that the precise complexity of the logic is tied in with an open problem of complexity theory, related to deciding sentences of real arithmetic.

We proposed a set of rationality postulates for revision operators in this logic. Using these rationality postulates, we also obtained semantic understanding of revision through Katsuno & Mendelzon-style representation theorems [11]. Working with stochastic environments and highly expressive beliefs and intentions about probability and time brings new challenges compared to prior work in the field, both technical and conceptual. In terms of technical challenges, giving a representation theorem is complicated by the fact that in general, there are infinitely many MDPs satisfying certain beliefs, due to the presence of probability, and that we have no general guarantees on the expressibility of sets of MDPs. These features are vital to the original theorems of [11]. To overcome this challenge, we employed results and methods from [5] to still obtain representation theorems. But to make our work more practically applicable, we went further by defining two novel postulates which additionally provide representation theorems for a specific class of operators that are conceptually simple and generally *computable*, properties which are not guaranteed by the results of [5].

In future work, we plan to extend our framework to allow for *iterated* revision, à la [3]. We also wish to develop a representation theorem with a modified version of (I1), which would allow success of revision only given coherent new intentions, as is also the case in the work of [17]. Such revision could follow the approach as taken in the *consistent AGM revision* of [7, 8], where revision with inconsistent statements maintains the original beliefs, or the more nuanced approach of [12], in which revision by a formula does not need to immediately entail it, but instead only increase its plausibility every time it is revised with.

References

1. Castelfranchi, C., Paglieri, F.: The role of beliefs in goal dynamics: prolegomena to a constructive theory of intentions. Synthese **155**(2), 237–263 (2007)
2. Cohen, P.R., Levesque, H.J.: Intention is choice with commitment. Artificial intelligence (1990), http://www.sciencedirect.com/science/article/pii/0004370290900555
3. Darwiche, A., Pearl, J.: On the logic of iterated belief revision. Artificial Intelligence **89**(1), 1–29 (1997). https://doi.org/https://doi.org/10.1016/S0004-3702(96)00038-0, https://www.sciencedirect.com/science/article/pii/S0004370296000380
4. van Ditmarsch, H., de Lima, T., Lorini, E.: Intention change via local assignments. In: Dastani, M., El Fallah Seghrouchni, A., Hübner, J., Leite, J. (eds.) Languages, Methodologies, and Development Tools for Multi-Agent Systems. pp. 136–151. Springer, Berlin, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22723-3_8
5. Falakh, F.M., Rudolph, S., Sauerwald, K.: AGM belief revision, semantically (2023), arXiv:2112.13557
6. Grant, J., Kraus, S., Perlis, D., Wooldridge, M.J.: Postulates for revising BDI structures. Synth. **175**(Supplement-1), 39–62 (2010)
7. Hansson, S.O.: A basis for AGM revision in bayesian probability revision. J. Philos. Log. **52**(6), 1535–1559 (2023)
8. Hansson, S.O.: Iterated AGM revision based on probability revision. J. Log. Lang. Inf. **32**(4), 657–675 (2023)
9. van der Hoek, W., Jamroga, W., Wooldridge, M.: Towards a theory of intention revision. Synthese **155**(2), 265–290 (2007). https://doi.org/10.1007/s11229-006-9145-6
10. Icard, T., Pacuit, E., Shoham, Y.: Joint revision of belief and intention. In: Proceedings of the Twelfth International Conference on Principles of Knowledge Representation and Reasoning. p. 572–574. KR'10, AAAI Press (2010)
11. Katsuno, H., Mendelzon, A.O.: Propositional knowledge base revision and minimal change. Artificial Intelligence **52**(3), 263–294 (1991). https://doi.org/10.1016/0004-3702(91)90069-V
12. Konieczny, S., Grespan, M.M., Pérez, R.P.: Taxonomy of improvement operators and the problem of minimal change. In: Proceedings of the Twelfth International Conference on Principles of Knowledge Representation and Reasoning. p. 161–170. KR'10, AAAI Press (2010)
13. Lorini, E., Herzig, A.: A logic of intention and attempt. Synthese **163**(1), 45–77 (2008). https://doi.org/10.1007/s11229-008-9309-7
14. Motamed, N., Alechina, N., Dastani, M., Doder, D.: Revising beliefs and intentions in stochastic environments. In: Larson, K. (ed.) Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24. pp. 3513–3522. International Joint Conferences on Artificial Intelligence Organization (8 2024). https://doi.org/10.24963/ijcai.2024/389, https://doi.org/10.24963/ijcai.2024/389, main Track
15. Shapiro, S., Sardiña, S., Thangarajah, J., Cavedon, L., Padgham, L.: Revising conflicting intention sets in BDI agents. In: van der Hoek, W., Padgham, L., Conitzer,

V., Winikoff, M. (eds.) International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2012. pp. 1081–1088 (2012)
16. Shoham, Y.: Logical theories of intention and the database perspective. Journal of Philosophical Logic **38**(6), 633–647 (2009). https://doi.org/10.1007/s10992-009-9116-8
17. van Zee, M., Doder, D., van der Torre, L., Dastani, M., Icard, T., Pacuit, E.: Intention as commitment toward time. Artificial Intelligence **283**, 103270 (2020). https://doi.org/10.1016/j.artint.2020.103270
18. Wooldridge, M.J.: Reasoning about Rational Agents. MIT Press (2000)
19. van Zee, M., Doder, D.: Agm-style revision of beliefs and intentions. In: ECAI. Frontiers in Artificial Intelligence and Applications, vol. 285, pp. 1511–1519. IOS Press (2016)