



















## REFERENCES

- [1] R. P. Abelson. 1979. Difference between belief and knowledge systems. *Cognitive Science* 3 (1979), 355–366.
- [2] N. Ajmeri, H. Guo, P. K. Murukannaiah, and M. P. Singh. 2020. Elessar: Ethics in Norm-Aware Agents. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020)*. IFAAMAS, 16–24.
- [3] C. Allen, G. Varner, and J. Zinser. 2000. Prolegomena to any future artificial moral agent. *Journal of Experimental and Theoretical Artificial Intelligence* 12, 3 (2000), 251–261.
- [4] R. C. Arkin, P. Ulam, and A. R. Wagner. 2012. Moral Decision Making in Autonomous Systems: Enforcement, Moral Emotions, Dignity, Trust, and Deception. *Proc. IEEE* 100, 3 (2012), 571–589.
- [5] P. Balbiani, D. Fernández-Duque, and E. Lorini. 2016. A Logical Theory of Belief Dynamics for Resource-Bounded Agents. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems (AAMAS 2016)*. ACM, 644–652.
- [6] A. Baltag, L. Moss, and S. Solecki. 1998. The Logic of Public Announcements, Common Knowledge and Private Suspicions. In *Proceedings of the Seventh Conference on Theoretical Aspects of Rationality and Knowledge (TARK'98)*, Itzhak Gilboa (Ed.). Morgan Kaufmann, San Francisco, CA, 43–56.
- [7] N. Belnap, M. Perloff, and M. Xu. 2001. *Facing the future: agents and choices in our indeterminist world*. Oxford University Press, New York.
- [8] C. Benz Müller, X. Parent, and L. W. N. van der Torre. 2020. Designing normative theories for ethical and legal reasoning: LogiKey framework, methodology, and tool support. *Artificial Intelligence* 287 (2020).
- [9] F. Berreby, G. Bourgne, and J.-G. Ganascia. 2017. A Declarative Modular Framework for Representing and Applying Ethical Principles. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2017)*. ACM, 96–104.
- [10] P. Blackburn, M. de Rijke, and Y. Venema. 2001. *Modal Logic*. Cambridge University Press, Cambridge.
- [11] C. Castelfranchi and F. Paglieri. 2007. The role of beliefs in goal dynamics: prolegomena to a constructive theory of intentions. *Synthese* 155, 2 (2007), 237–263.
- [12] S. Cranefield, M. Winikoff, V. Dignum, and F. Dignum. 2017. No Pizza for You: Value-based Plan Selection in BDI Agents. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI 2017)*. ijcai.org, 178–184.
- [13] L. A. Dennis, M. Fisher, M. Slavkovik, and M. Webster. 2016. Formal verification of ethical choices in autonomous systems. *Robotics and Autonomous Systems* 77 (2016), 1–14.
- [14] J. Dewey. 1939. *Theory of Valuation*. University of Chicago Press.
- [15] F. Dietrich and C. List. 2013. A reason-based theory of rational choice. *Noûs* 47, 1 (2013), 104–134.
- [16] F. Dietrich and C. List. 2017. What matters and how it matters: A choice-theoretic representation of moral theories. *Philosophical Review* 126, 4 (2017), 421–479.
- [17] V. Dignum. 2017. Responsible Autonomy. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI 2017)*. ijcai.org, 4698–4704.
- [18] A. Etzioni and O. Etzioni. 2017. Incorporating Ethics into Artificial Intelligence. *The Journal of Ethics* 21 (2017), 403–418.
- [19] R. Fagin, J. Halpern, Y. Moses, and M. Vardi. 1995. *Reasoning about Knowledge*. MIT Press, Cambridge.
- [20] P. M. Gollwitzer. 1996. The volitional benefits of planning. In *The psychology of action*. Guilford Press, 287–312.
- [21] S. O. Hansson. 2001. *The Structure of Values and Norms*. Cambridge University Press.
- [22] J. Harsanyi. 1982. Utilitarianism and Beyond. In *Morality and the theory of rational behaviour*, A. K. Sen and B. Williams (Eds.). Cambridge University Press, Cambridge.
- [23] E. Hemaspaandra. 1996. The Price of Universality. *Notre Dame Journal of Formal Logic* 37, 2 (1996), 174–203.
- [24] J. F. Horty. 2001. *Agency and Deontic Logic*. Oxford University Press, Oxford.
- [25] I. Levi. 1990. *Hard Choices: Decision Making Under Unresolved Conflict*. Cambridge University Press.
- [26] F. Liu. 2011. *Reasoning about Preference Dynamics*. Springer.
- [27] E. Lorini. 2013. Temporal STIT logic and its application to normative reasoning. *Journal of Applied Non Classical Logics* 23, 4 (2013), 372–399.
- [28] E. Lorini. 2015. A logic for reasoning about moral agents. *Logique & Analyse* 58, 230 (2015), 177–218.
- [29] M. Miceli and C. Castelfranchi. 2000. The role of evaluation in cognition and social interaction. In *Advances in consciousness research. Human cognition and social agent technology*, K. Dautenhahn (Ed.). John Benjamins Publishing Company, 225–261.
- [30] A. Moors, P. C. Ellsworth, K. Scherer, and N. Frijda. 2013. Appraisal Theories of Emotion: State of the Art and Future Development. *Emotion Review* 5, 2 (2013), 119–124.
- [31] P. K. Murukannaiah, N. Ajmeri, C. M. Jonker, and M. P. Singh. 2020. New Foundations of Ethical Multiagent Systems. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020)*. IFAAMAS, 1706–1710.
- [32] T. Nagel. 1980. The Limits of Objectivity. In *The Tanner Lectures on Human Values*. Cambridge University Press, 77–139.
- [33] M. Rodríguez-Soto, M. López-Sánchez, and J. A. Rodríguez-Aguilar. 2020. A Structural Solution to Sequential Moral Dilemmas. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2018)*. IFAAMAS, 1152–1160.
- [34] J. Searle. 2001. *Rationality in Action*. MIT Press, Cambridge.
- [35] A. Sen. 1985. Well-Being, Agency and Freedom: The Dewey Lectures 1984. *The Journal of Philosophy* 82, 4 (1985), 169–221.
- [36] A. Sen. 1987. *On Ethics and Economics*. Basil Blackwell.
- [37] M. Serramia, M. López-Sánchez, J. A. Rodríguez-Aguilar, M. Rodríguez, M. J. Wooldridge, J. Morales, and C. Ansótegui. 2018. Moral Values in Norm Decision Making. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2018)*. IFAAMAS, 1294–1302.
- [38] J. van Benthem. 2016. Tracking Information. In *J. Michael Dunn on Information Based Logics*, K. Bimbó (Ed.). Springer, 363–389.
- [39] J. van Benthem, D. Fernández-Duque, and E. Pacuit. 2014. Evidence and plausibility in neighborhood structures. *Annals of Pure and Applied Logic* 165, 1 (2014), 106–133.
- [40] J. van Benthem and F. Liu. 2007. Dynamic logic of preference upgrade. *Journal of Applied Non Classical Logics* 17, 2 (2007), 157–182.
- [41] D. Vanderelst and A. F. T. Winfield. 2018. An architecture for ethical robots inspired by the simulation theory of cognition. *Cognitive Systems Research* 48 (2018), 56–66.
- [42] W. Wallach and C. Allen. 2008. *Moral Machines: Teaching Robots Right from Wrong*. Oxford University Press.
- [43] A. F. T. Winfield, C. Blum, and W. Liu. 2014. Towards an Ethical Robot: Internal Models, Consequences and Ethical Action Selection. In *Proceedings of the 15th Annual Conference on Advances in Autonomous Robotics Systems (TAROS 2014) (LNCS, Vol. 8717)*. Springer, 85–96.