Using Values and Norms to Model Realistic Social Agents

Rijk Mercuur, Virginia Dignum, Catholijn Jonker

Delft University of Technology

In the past years a series of papers has highlighted the importance of developing agents which are more socially realistic [7, 6, 5, 10]. Realism, here, means that the agent reproduces human behavior. The aforementioned proponents of social agents furthermore claim that modeling humans with standard agent frameworks (e.g., BDI-agents [13]) filtered according to some 'social awareness' modules is not realistic enough. Humans do not first construct a list of possible actions and then realize there are others in the world. Humans use social ingredients, which play a central role in the decision making of the agent. According to Dignum et al. [6], a few possible candidates for such social ingredients are human values, norms, identity, and motives. This paper focuses on two social ingredients: values and norms. Values are understood as 'what one finds important in life', for example, privacy, wealth or fairness [11]. Norms are understood as a standard or rule in society [2].

We have looked at a few studies that indeed use norms and values as the main driver of actions [4, 1, 8]. Although these studies help in *constructing* theories on social agents, they do not *check* this theory against independent quantitative data on human behavior. Without this check it is unclear if these theories provide the social realism we need. This raises the question: how useful are value-based and normative agents for modeling *realistic* social agents?

This paper synthesizes and evaluates three theories: a theory on values, a theory on norms and a theory that combines both values and norms. We use the same theory on values as described in Mercuur et al. [9]:

V.1 There are ten different basic values, those empirically established by Schwartz [12]. Two examples are wealth and fairness.
V.2 The importance one attributes to a value is correlated with the importance one attributes to other values Schwartz [12]. For example, the values of wealth and fairness are negatively correlated.
V.3 Humans are heterogeneous in the values they find important.
V.4 A human considers a subset of these values to determine his or her actions.
V.5 Values are the prime cognitive determiner for actions.

Our simple theory on norms (e.g., no norm violation) comprises:

N.1 A statement is a norm if and if only it comprises an actor, action, a deontic elements and a condition that states when it applies.
N.2 A norm exists, for a particular person, when that persons beliefs other people do or expect that norm.
N.3 The action a human does is the same as what they perceive as the norm.

And lastly we look at a simple theory that combines norms and values:

VN.1 Some humans always act according to their norms and other humans always act according to their values.

We compared the behavior of humans to that of agents by simulating a psychological experiment: the ultimatum game (UG). In the UG, two players (human or agent) negotiate over a fixed amount of money ('the pie'). Player 1, the *proposer*, demands a portion of the pie, with the remainder offered to Player 2. Player 2, the *responder*, can choose to accept or reject this proposed split. If the responder chooses to 'accept', the proposed split is implemented. If the responder chooses to 'reject', both players get no money. One round comprises of one such demand and reply.

We consider two scenarios: the one-round UG and the multi-round UG. In the one-round UG, humans demand on average 56% of the pie and accept about 81% of the demands [3]. One popular explanation of why humans make these particular demands and accepts is that they have evolved to do this by repeated interaction. This paper checks if simulating our agents to interact for many rounds eventually results in the demands and accepts human make in one-shot games. We found that our theory of norms cannot reproduce these demands and accepts. An intuitive explanation of this is that the normative agents can only copy social norms, but have no incentive to make the precise 56% demands and 81% accepts that humans display. In contrast, our theory on values can reproduce human demands and accepts. Lastly, our theory that combines values and norms can reproduce the human demands and accepts as well. The intuitive explanation here is that this theory allows some agents to act out of their values that will guide the other agents (that act out of norms) towards the realistic demands and accepts.

In the multi-round UG, humans are shown to make on average slightly higher demands and slightly more accepts in the 10th round than in the first. This paper checks if simulating our agents for 10 rounds reproduces this rise in demands and accepts. We found that our theory on values cannot reproduce the learning rate humans display, while our theory on norms can. Norms, in contrast to values, provide a dynamic learning element that makes the agents – just like humans – adaptive. Our theory that combines values and norms can reproduce the learning rate in demands as well. It allows some of these agents to act out of their norms and as such produces similar results as our theory on norms.

We conclude that, for our domain, the theory that combines both values and norms seems the most realistic. First, in contrast to our theory on norms, it can reproduce first-round behavior. Second, in contrast to our theory on values, it can reproduce multi-round behavior. Norms add a realistic adaptive element to the agent behavior, while values provide the static ground-work that realistically models the balance humans find between fairness and wealth. Our study shows a first step in how values and norms can be used to come closer to realistic social agents.

Bibliography

[1] Atkinson, K., Bench-capon, T.: Value Based Reasoning and the Actions of Others. In: Proceedings of ECAI. pp. 680–688 (2016)

[2] Bicchieri, C.: The Grammar of Society: The Nature and Dynamics of Social Norms. Cambridge University Press (2005), www.cambridge.org

[3] Cooper, D.J., Dutcher, E.G.: The Dynamics of Responder Behavior in Ultimatum Games: A Meta-Study. Experimental Economics 14(4), 519–546 (2011)

[4] Dechesne, F., Di Tosto, G., Dignum, V., Dignum, F.: No Smoking Here: Values, Norms and Culture in Multi-Agent Systems. Artificial Intelligence and Law 21(1), 79–107 (8 2012), http://link.springer.com/10.1007/s10506-012-9128-5

[5] Dignum, F., Dignum, V., Prada, R., Jonker, C.M.: A Conceptual Architecture for Social Deliberation in Multi-Agent Organizations. Multiagent and Grid Systems 11(3), 147–166 (2015)

[6] Dignum, F., Prada, R., Hofstede, G.: From Autistic to Social Agents. In: Proceedings of the 2014 international conference on Autonomous Agents and Multi-Agent Systems. pp. 1161–1164 (2014), http://dl.acm.org/citation.cfm?id=2617431

[7] Kaminka, G.: Curing robot autism: A challenge. Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems pp. 801–804 (2013)

[8] Mercuur, R., Dignum, F., Kashima, Y.: Changing Habits Using Contextualized Decision Making. In: Advances in Intelligent Systems and Computing. vol. 528 (2017)

[9] Mercuur, R., Dignum, V., Jonker, C.: Modeling Social Preferences with Values. In: Smart Modeling and Simulation for Complex Systems @ IJCAI. p. (in press) (2017)

[10] Norling, E.: Don't Lose Sight of the Forest: Why the Big Picture of Social Intelligence is Essential. In: Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems. pp. 984–987. No. May (2016)

[11] Poel, I.v.d., Royakkers, L.: Ethics, Technology, and Engineering: An Introduction. John Wiley & Sons (2011), http://eu.wiley.com/WileyCDA/WileyTitle/productCd-EHEP002302.html

[12] Schwartz, S.H.: An Overview of the Schwartz Theory of Basic Values. Online Readings in Psychology and Culture 2, 1–20 (2012)

[13] Weiss, G.: Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence. MIT press, second edi edn. (2013)