

From Externalities to Norms

Jan Broersen, Frank Dignum, Davide Grossi, Rosja Mastop, Paolo Turrini

August 24, 2007

One fundamental issue of social choice theory [2] is how to aggregate the preferences of individual agents in order to form decisions to be taken by society as a whole. However, once we want to take into account the capabilities and strategic reasoning power of agents, as we do in multi-agent systems, mere social choice functions are not enough to explain how and (especially) why individual interests are aggregated in the way they are. In this context, norms should be seen as emergent social constructions that enable us to enforce socially desirable outcomes (cfr. [3]).

In particular there are situations in which individual preferences are not compatible, and coalitions compete to achieve a given social order. A typical case is that of an agent's capability to positively or negatively affect the realization of other agents' preferences. In the economic literature these phenomena are known as externalities ([5]). In our paper we will view the enactment of norms as aimed at the regulation of such externalities. By enacting a norm we mean *the introduction of a normative constraint on individual and collective choices in a multiagent system*. Following Anderson [1], we model the constraints using propositional constants *viol*.

We are specifically concerned with cases where the collective perspective is at odds with the individual perspective. That is, cases where we think that letting everybody pick their own best action regardless of other's interest gives a non-optimal result. The main question we are dealing with is then: how do we determine which norms, if any, are to be imposed?

To answer this question, the paper presents a language to talk about externalities and the process of regulating them. In order to represent abilities of agents we will employ coalition logic [6], and we will model an agent's desires as a collection of subsets of possible worlds. Once we have these two objects, their interaction (i.e. "Coalition C has a strategy to fulfil coalition Y desires") comes natural. In particular we can define that an externality is an action that may interfere with other agents' abilities to reach their preferred states. We model the enactment of norms as model updates, that change the valuation of the special proposition *viol* in accordance to the externality that we would like to forbid.

Consider, for example, the situation in the classical prisoner dilemma as depicted in Table 1. Row can decide to cooperate or to defect. We can

	Column	Cooperate	Defect
Row		(2, 2)	(0, 3)
	Cooperate	(2, 2)	(0, 3)
	Defect	(3, 0)	(1, 1)

Table 1: Prisoners' dilemma

	Column	Left	Right
Row		(3, 3)	(0, 0)
	Left	(3, 3)	(0, 0)
	Right	(0, 0)	(3, 3)

Table 2: Collision Game

observe that, reasoning strategically, Row has a negative externality towards Column by deciding to defect, while it has a positive externality towards him by deciding to cooperate. For column it is exactly the same.

In this situation, a legislator that wants to achieve the socially optimal state (cooperate,cooperate), should declare that defecting is forbidden, thereby labeling the combinations of moves (defect, defect), (defect,cooperate) and (cooperate, defect) as violations. Note that such a norm is in the interest of the players themselves, since they are better off than if they would have pursued the Nash Equilibrium, ending up in the (defect, defect) state.

Other interesting examples concern conventional norms. A classical one is the 'collision' game between Row and Column that have to choose between driving left or right (see Table 2). In this game the outcomes are good for both in case both make the same choice (i.e. they both decide to drive to the left), they are bad for both otherwise (i.e. one decides to drive left, the other right). Here both players have a negative externality towards the other by choosing the wrong. Note how a negative externality of one player towards the other turns out to be a negative externality of the player towards himself. The preferred outcomes of both players are in fact the same. A norm helping both players to reach an optimal outcome would be one that labels as violations combinations of discordant choices. Moreover, in this kind of games Row will never know what is the best thing to choose, since the choice of Column is independent from his. Simply avoiding what is forbidden solves the problem.

It only makes sense to introduce norms in a multiagent system if agents are not able to reach a good outcome, or risk a bad outcome, if they just fulfill their individual preferences. Also, agents in a system need to be able to indeed comply to norms [9]. And, finally, norms are not isolated entities; they are closely related to the preferences agents have [2]. It turns out then,

that non-trivial norms (those that rule out some yet possible behaviour) always regulate some externality.

The effectivity of a norm is the extent to which a given enacted obligation will be fulfilled in a multiagent system. We argue, following the theory of Kohlberg’s moral stages [4], that differentiating agents within normative types can help to understand whether a norm will be in fact effective. We can think of agents that only act according to their preferences, and thus will not follow any regulation contrasting with them, and agents that have a societal view, and that will act maximizing what is best for the coalitions they belong to. To formalize this, we adjust Preference Upgrade Logic ([8]) to allow for different strategies of preference change.

Finally, we show for several specific desirable properties of multi-agent systems, which externalities to regulate. We look at various ways of institutionalizing optimal norms, for instance by Consensus Rules, Majority Rules and Dictatorship Rules [7].

References

- [1] A.R. Anderson. A reduction of deontic logic to alethic modal logic. *Mind*, 67:100–103, 1958.
- [2] K.J. Arrow. *Social Choice and Individual Values*. Yale University Press, 1970.
- [3] J. Coleman. *Foundations of Social Theory*. Belknap Harvard, 1990.
- [4] L. Kohlberg, C. Levine, and A. Hower. *Moral stages: a current formulation and a response to critics*. Basel, NY, 1983.
- [5] Eric S Maskin. The invisible hand and externalities. *American Economic Review*, 84(2):333–37, 1994.
- [6] M.Pauly. *Logic for Social Software*. ILLC Dissertation Series, 2001.
- [7] M. Pauly. Axiomatizing collective judgment sets in a minimal logical language. *Synthese*, 2:233–250, 2007.
- [8] J. van Benthem and F.Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logics*, 14, 2004.
- [9] G.H. von Wright. *Norm and Action: A Logical Enquiry*. Rutledge and Kegan Paul, 1963.