

Title : Extending Formal Models of the Doctrine of Double Effect  
with Emotions  
Authors : Naveen Sundar Govindarajulu, Rikhiya Ghosh and Selmer  
Bringsjord  
Speaker : Naveen Sundar Govindarajulu

Abstract: We have previously modeled the Doctrine of Double Effect (DDE) in a formal computational logic [1]. While DDE can account for a large range of human behavior in moral dilemmas (situations in which all available actions have positive and negative effects), the doctrines have to be further extended to account for how humans behave (and ought to behave). One such extension has produced a version of DDE that can model self-sacrifice [2]. We now extend these models with support for emotional content that can mix with other cognitive states such as belief and intention used in prior models of DDE. A simple version of DDE stipulates that agents can perform the only available action in a dilemma iff (i) the positive effects are much larger than the negative effects; (ii) the agent does not intend any negative effect but intends some of the positive effects; (iii) none of the negative effects are used to cause any of the positive effects; and (iv) the action is not forbidden. Emotions play an important role in that they themselves can be considered to be positive and negative effects. Agents can also use emotions to shape their desires and intentions. Our new model takes into account the above two dynamics.

[1] Naveen Sundar Govindarajulu and Selmer Bringsjord (2017). "On Automating the Doctrine of Double Effect." In Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne, Australia. <https://doi.org/10.24963/ijcai.2017/658>

[2]. Naveen Sundar Govindarajulu, Selmer Bringsjord, Rikhiya Ghosh, and Matthew Peveler (Forthcoming 2017). "Beyond The Doctrine Of Double Effect: A Formal Model Of True Self-Sacrifice." In Proceedings of The International Conference on Robot Ethics and Safety Standards. October, 2017. Lisbon, Portugal.