Contextual Deontic Cognitive Event Calculi for Ethically Correct Robots

(abstract)

Selmer Bringsjord • Naveen Sundar G. • Bertram Malle • Matthias Scheutz

ver 1025172359CA for ISAIM 2018

The common situation

(1) Jones is obligated to pay back Smith $10 tonight.

(2) Jones can't possibly pay back Smith $10 tonight.

gets symbolized in (the deontic logic) $\mathbf{KT}d$, which subsumes **SDL** (standard deontic logic), where $\Diamond$ represents (logical?) possibility, as:

(1′) $\mathbf{O}\phi$

(2′) $\neg\Diamond\phi$

However, as McNamara (2010) points out, it's long been known that in light of "Kant's Law" we then immediately face a contradiction, for in $\mathbf{KT}d$, $\vdash \mathbf{O}\phi \rightarrow \Diamond\phi$. This conditional is known as 'Kant's Law,' and we call the reasoning involving it and Jones the 'Kant's-Law Paradox' (K-LP). In light of this paradox, a machine or robot intended to operate as a morally competent banker overseeing Jones and his colleagues is, if classical, apparently paralyzed.

An early reaction to K-LP is Thomason's (1981); he proposes a distinction between two *contexts* to purportedly save the day for Kant: a *context of deliberation* vs. a *context of justification*. This reaction has been taken up and carried into (at least paper-and-pencil) logicist AI by van der Torre (2000, 2003), who has introduced so-called *contextual* deontic logic. This otherwise impressive work is afflicted by a number of very serious problems, however. First, contextual deontic logic is merely modal *propositional*, yet many if not the vast majority of meaningful real-world challenges involve quantification.[1] A second fatal flaw from the perspective of the real world is that deontic reasoning is inseparably bound up not only with quantification, but with knowledge, belief, intention, perception, desire, the full gamut of numerous emotions, counterfactual reasoning, planning and prac-

[1] Actually, even the original scenario involving Jones, given at the outset, if the English for (1) and (2) is taken seriously, involves quantification.

tical possibilty flowing therefrom,[2] and communication between agents, both human and machine. In short, having on hand a narrow restricted logic isn't likely to be too helpful when it comes to real-world moral problems; what's needed is a comprehensive *cognitive calculus* with enough representational reach, in the realm of the intensional, to make respectable any such notion as that logic can really and truly do justice to the astounding complexity of morality, and moral reasoning and decision-making. There are at least four additional, serious problems that plague van der Torre's contextual deontic logic. In addition, perhaps most significantly, contextual deontic logic, at least to our knowledge, hasn't been implemented in any artificial agents, let alone robots.

To be constructive, we flip the situation on its head, and view the host of problems plaguing contextual deontic logic as an aspirational list of desiderata to be satisfied by any viable logicist program in machine/robot ethics. After proving that our own extant systems for giving artificial agents sophisticated deontic reasoning (e.g., a dialect of the deontic cognitive event calculus, $\mathcal{D}^e\mathcal{CEC}$) can subsume standard contextual deontic logic (including dyadic dedontic logic, and including the defasible aspects inspired by Reiter's default logic that van der Torre has formalized),[3] we proceed to explain how the desiderata are satisfied in our recent *extensions* of the likes of $\mathcal{D}^e\mathcal{CEC}$.

Our explanation factors in work on the "logification" of context by John McCarthy and collaborators (e.g., Makarios), and, unlike prior work on context and precise moral reasoning, is guided by results obtained by Malle regarding "norm contexts" on the empirical side. Finally, our calculi are *implemented*, and, as we show, are able to regulate real robots on the strength of contextual deontic reasoning that can be applied far and wide.

[2]The proof of absurdity in K-LP, as alert readers will have noted (and we sent a signal with our parenthetical '(logical?)' in the first sentence of the present paper), is itself potentially confused, and points to a disturbing scarcity of intensional operators, since it uses $\Diamond$ to try to capture what is possible for Jones in a practical sense, but $\Diamond$ is in modal logic frequently intended to capture so-called *logical* possibility.

[3]$\mathcal{D}^e\mathcal{CEC}$includes as a proper part the machinery of all the forms of modal-propositional dyadic logic that we are aware of. This is trivial in cases where dyadic deontic logic simply builds e.g. $\mathbf{O}(\phi/\psi)$ out of material or strict implication, as in the old and venerabel move, made long ago by Chellas (1974) in which $\mathbf{O}(\phi/\psi)$ iff $\psi \rightarrow / \Rightarrow \mathbf{O}\phi$. In dyadic deontic logics where such constructions as $\mathbf{O}(\phi/\psi)$ is not reducible, $\mathcal{D}^e\mathcal{CEC}$ offers the wff type

$$\mathbf{O}(a, t, \phi, \alpha, t')$$

where a is and agent, ϕ is a formula of quantified intensional logic, α is an action within the repertoire of a, and t and t' are times.

References

Chellas, B. (1974), Conditional Obligation, *in* 'Logical Theory and Semantic Analysis: Essays Dedicated to Stig Kanger on His Fiftieth Birthday', Springer, Dordrecht, The Netherlands, pp. 23–33. This book is in the Synthese Library, Series Volume 63.

McNamara, P. (2010), Deontic Logic, *in* E. Zalta, ed., 'The Stanford Encyclopedia of Philosophy'. McNamara's (brief) note on a paradox arising from Kant's Law is given in an offshoot of the main entry.
URL: *https://plato.stanford.edu/entries/logic-deontic*

Thomason, R. (1981), Deontic Logic as Founded on Tense Logic, *in* R. Hilpinen, ed., 'New Studies in Deontic Logic: Norms, Actions and the Foundations of Ethics', D. Reidel, pp. 165–176.

van der Torre, L. (2003), 'Contextual Deontic Logic: Normative Agents, Violations and Independence', *Annals of Mathematics and Artificial Intelligence* **37**, 33–63.

van der Torre, L. & Tan, Y.-H. (2000), Contextual Deontic Logic: Violation Contexts and Factual Defeasibility, *in* P. Bonzon & M. Cavalcanti, eds, 'Formal Aspects of Context', Kluwer, Dordrecht, The Netherlands, pp. 143–160.