Formalising Oughts and Practical Knowledge without Resorting to Action Types

Extended Abstract

Jan Broersen
Utrecht University
Utrecht, The Netherlands
J.M.Broersen@uu.nl

Aldo Iván Ramírez Abarca
Utrecht University
Utrecht, The Netherlands
a.i.ramirezabarca@uu.nl

ABSTRACT

We show how the logical modelling of puzzles concerning epistemic oughts, such as those put forward by Horty & Pacuit, are solved without introducing action types. We accomplish this within *stit* logic ([2], [6]), by lifting a deontic ordering over histories to the level of actions that an agent can knowingly perform. Through the relation of 'can' and 'practical knowledge', we arrive at objective and subjective axiomatizable versions of the ought-to-do modality.

KEYWORDS

Epistemic logic; Deontic logic; Logic of action; Logic of responsibility

ACM Reference Format:
Jan Broersen and Aldo Iván Ramírez Abarca. 2018. Formalising Oughts and Practical Knowledge without Resorting to Action Types. In *Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), Stockholm, Sweden, July 10–15, 2018,* IFAAMAS, 3 pages.

1 EPISTEMIC DEONTIC STIT LOGIC

In their recent work [5], Horty & Pacuit argue that we cannot have a logical theory of oughts, knowledge, and action without resorting to action types[1]. Accordingly, they motivate the introduction of types into epistemic deontic *stit* logic by presenting 3 puzzles for which standard semantics of knowledge and ought-to-do ([6]) fail. Although they develop a semantics that solves these puzzles, it comes with major technical disadvantages inherent to the type construction. Our approach –free of types– manages to solve the puzzles by adding a new deontic operator. Before presenting our proposal, we recover the basic definitions of the logic used. The reader is referred to [6] and [7] for comprehensive definitions.

Definition 1.1 (Syntax). Given a finite set *Ags* of agent names and a countable set of propositions P, with $p \in P$ and $\alpha \in Ags$, the grammar for the formal language $\mathcal{L}_{\text{KCo}}$ is given by:

$$\varphi \quad := \quad p \mid \neg\varphi \mid \varphi \wedge \psi \mid \Box\varphi \mid [\alpha\ \texttt{Cstit}]\varphi \mid K_\alpha\varphi \mid \odot[\alpha\ \texttt{Cstit}]\varphi$$

[1] This echoes similar claims by Lorrini and Herzig [4], and is in accordance with the broad sentiment in the DEL community and in the ATL [1] community. As we see it, the difference between types and tokens is that, while an action token is the particular performance of an action by an agent at a specific moment (*opening the window of my bedroom at 8 a.m. on Monday...*), action types refer to categories or patterns of actions that are instantiated in tokens (*to open a window*).

Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), M. Dastani, G. Sukthankar, E. André, S. Koenig (eds.), July 10–15, 2018, Stockholm, Sweden.

$\Box\varphi$ will express historical necessity of φ ($\Diamond\varphi$ abbreviates $\neg\Box\neg\varphi$); $[\alpha\ \texttt{Cstit}]\varphi$ stands for 'agent α sees to it that φ'; K_α is the epistemic operator for α; and $\odot[\alpha\ \texttt{Cstit}]$ is meant to represent what α ought to do ([6]).

Definition 1.2 (Frames). A *finite-choice epistemic utilitarian KCo-frame* is a tuple $\langle T, \sqsubset, \textbf{Choice}, \{\sim_\alpha\}_{\alpha\in Ags}, \textbf{Value}\rangle$, where T is a non-empty set of moments and $\sqsubset$ is a strict partial ordering on T satisfying 'no backward branching'. Each maximal $\sqsubset$-chain is called a *history*. H denotes the set of all histories, and for each $m \in T$, $H_m := \{h \in H \mid m \in h\}$. Tuples $\langle m, h\rangle$ are called *situations* iff $m \in T$, $h \in H$, and $m \in h$. **Choice** is a function that maps each agent α and moment m to a *finite*[2] partition $\textbf{Choice}^m_\alpha$ of H_m. The cells of such a partition represent α's available actions at m. **Value** is a function that assigns to each history h a real number representing the *utility* of h. For each agent α, $\sim_\alpha$ is an epistemic equivalence relation on the set of situations.

For the semantics of ought-to-do, [6] uses a weak dominance ordering $\leq$ on $\textbf{Choice}^{m_*}_\alpha$: for $L, L' \subseteq H_{m_*}$, $L \leq L'$ iff $\textbf{Value}(h) \leq \textbf{Value}(h')$ for every $h \in L, h' \in L'$. We write $L \prec L'$ iff $L \leq L'$ and $L' \not\leq L$. With this ordering, an optimal set of actions $\textbf{Optimal}^{m_*}_\alpha$ is defined as $\{L \in \textbf{Choice}^{m_*}_\alpha ; \nexists L' \in \textbf{Choice}^{m_*}_\alpha \text{ s. t. } L \prec L'\}$.

Definition 1.3 (Models). A finite-choice epistemic utilitarian KCo-frame $\langle T, \sqsubset, \textbf{Choice}, \{\sim_\alpha\}_{\alpha\in Ags}, \textbf{Value}\rangle$ is extended to a model by adding a valuation function $\mathcal{V} : P \to 2^{T\times H}$ assigning to each proposition the set of situations relative to which it is true. Extending the usual semantics for formulas involving classical connectives, we recursively define

$$\begin{array}{lll}
\langle m, h\rangle \models p & \Leftrightarrow & \langle m, h\rangle \in \mathcal{V}(p)\\
\langle m, h\rangle \models \Box\varphi & \Leftrightarrow & \forall h' \in H_m, \langle m, h'\rangle \models \varphi\\
\langle m, h\rangle \models [\alpha\ \texttt{Cstit}]\varphi & \Leftrightarrow & \forall h' \in \textbf{Choice}^m_\alpha(h), \langle m, h'\rangle \models \varphi\\
\langle m, h\rangle \models K_\alpha\varphi & \Leftrightarrow & \langle m, h\rangle \sim_\alpha \langle m', h'\rangle \text{ implies } \langle m', h'\rangle \models \varphi\\
\langle m, h\rangle \models \odot[\alpha\ \texttt{Cstit}]\varphi & \Leftrightarrow & \forall L \in \textbf{Optimal}^m_\alpha, \langle m, h'\rangle \models \varphi \text{ for all } h' \in L.
\end{array}$$

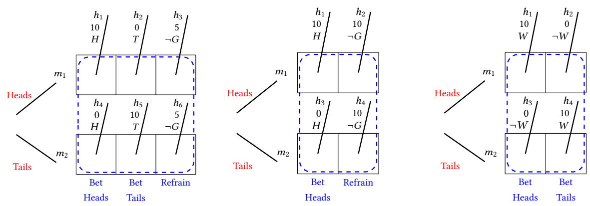
As a convention, we write $|\varphi|^m$ to refer to the set $\{h \in H_m ; \langle m, h\rangle \models \varphi\}$.

1.1 Formal account of the epistemic puzzles

Horty & Pacuit's 3 puzzles all start with a fair coin-flip, the outcome of which is still hidden for a betting agent α.

Example 1.4. Agent α can bet heads, bet tails, or refrain from betting. If α bets correctly, it wins €10. If it bets incorrectly, it does not win anything, and if it refrains from betting, it wins €5. Figure 1a provides a *stit* diagram for Horty & Pacuit's interpretation of the situation, where payoffs represent the values assigned to histories. The blue dotted rectangle represents the information set

[2] Hence the term "finite-choice" in the definition of these frames.



(a) Ex. 1.4 **(b) Ex.** 1.5 **(c) Ex.** 1.6

Figure 1: Epistemic puzzles

of α at any given situation. Since the outcome of the coin-flip is unknown, α cannot distinguish whether it is at moment m_1 or at moment m_2. In this interpretation, a problem ensues because $\langle m_i, h_i \rangle \models K_\alpha \odot [\alpha\ \texttt{Cstit}]G$, which means that α knows that it ought to gamble, even if this is a "risky" move that could lead to a null payoff.

Example 1.5. (Figure 1b) α can only bet heads or refrain. If α bets correctly, it wins €10. If α bets incorrectly, it does not win anything. If α refrains from betting, it *also* wins €10. Intuitively, α ought to refrain from gambling, for it would win the same amount as when betting correctly, but without engaging in possible failure. The problem here is that $\langle m_i, h_i \rangle \not\models K_\alpha \odot [\alpha\ \texttt{Cstit}]\neg G$: the agent does not know that it ought to refrain from gambling.

Example 1.6. (Figure 1c) If α bets correctly, it wins €10. If it bets incorrectly, it does not win anything. In Horty & Pacuit's formalization, the problem is that Kant's principle of 'ought implies can' is not satisfied: $\langle m_i, h_i \rangle \not\models K_\alpha \odot [\alpha\ \texttt{Cstit}]W \rightarrow \Diamond K_\alpha [\alpha\ \texttt{Cstit}]W$. Though α knowingly ought to win, it cannot knowingly do so.

Horty & Pacuit solve these puzzles by extending the language with an operator $[\ldots\ \texttt{kstit}]$ encoding *ex interim* knowledge. The semantics for this operator is what involves types, and thus is what brings 2 unfavorable new model constraints: (1) the epistemic relations occur not between moment-history pairs but between moments ([5]), which is semantically limiting, and (2) indistinguishable states offer same types, which Horty & Pacuit cannot characterize syntactically, since action types are not in the object language –this is the 'uniformity of strategies' constraint from ATEL and game theory.

2 OUR PROPOSAL

We propose to define epistemic ought-to-do's by lifting the dominance ordering over choices to an ordering over choices *within information sets*. This gives way to two different versions of ought-to-do: an *objective* one, which is exactly the same as Horty's act utilitarian ought-to-do, and a *subjective* one, which selects the best candidate out of the set of actions that the agent knows how to perform. Therefore, we extend the language of Definition 1.1 with a new operator $\odot_{\mathcal{S}}[\alpha\ \texttt{Cstit}]\varphi$.

In order to present the semantics for $\odot_{\mathcal{S}}[\alpha\ \texttt{Cstit}]\varphi$, we need more definitions. Fix $\alpha \in Ags$ and $m_* \in T$. For $L \subseteq H_{m_*}$ and $m \in T$, we define its *epistemic cluster set* at m, $[L]^m_\alpha$, as the set $\{h \in H_m; \exists h_* \in L \text{ s.t. } \langle m_*, h_* \rangle \sim_\alpha \langle m, h \rangle\}$. As a convention, we write $m \sim_\alpha m'$ if there exist $h \in H_m, h' \in H_{m'}$ such that $\langle m, h \rangle \sim_\alpha \langle m', h' \rangle$. We now define a *subjective* ordering $\preceq_s$ on $\mathbf{Choice}^{m_*}_\alpha$ such that for $L, L' \subseteq H_{m_*}$, $L \preceq_s L'$ iff for every m such that $m_* \sim_\alpha m$, $\mathbf{Value}(h) \leq \mathbf{Value}(h')$ for every $h \in [L]^m_\alpha, h' \in [L']^m_\alpha$. As expected, we write $L \prec_s L'$ iff $L \preceq_s L'$ and $L' \not\preceq_s L$. This ordering allows us to define a subjectively optimal set of actions $\mathbf{S}-\mathbf{optimal}^{m_*}_\alpha := \{L \in \mathbf{Choice}^{m_*}_\alpha;$ there is no $L' \in \mathbf{Choice}^{m_*}_\alpha$ s. t. $L \prec_s L'\}$. The semantics for the formulas of involving $\odot_{\mathcal{S}}[\alpha\ \texttt{Cstit}]$ is then defined by $\langle m, h \rangle \models \odot_{\mathcal{S}}[\alpha\ \texttt{Cstit}]\varphi$ iff $\forall L \in \mathbf{S}-\mathbf{optimal}^m_\alpha, \forall m'$ s.t. $m \sim_\alpha m'$, $[L]^{m'}_\alpha \subseteq |\varphi|^{m'}$. This semantics offers solutions to natural interpretations of Horty & Pacuit's puzzles. For Example 1.4, we capture the assumption that the outcome of the coin-flip is hidden by taking $\sim_\alpha$ to be defined by the following information sets: $\{\langle m_1, h_1 \rangle, \langle m_2, h_4 \rangle\}$, in which α bets heads; $\{\langle m_1, h_2 \rangle, \langle m_2, h_5 \rangle\}$, in which α bets tails; and $\{\langle m_1, h_3 \rangle, \langle m_2, h_6 \rangle\}$, in which α refrains from betting. With our semantics, the problem is solved because although $\langle m_i, h_i \rangle \models K_\alpha \odot [\alpha\ \texttt{Cstit}]G$, we consider this as the knowledge of an *objective* ought-to-do. *Subjectively* speaking, we do not obtain that α knows that it ought to gamble: $\langle m_i, h_i \rangle \not\models \odot_{\mathcal{S}}[\alpha\ \texttt{Cstit}]G$ and thus $\langle m_i, h_i \rangle \not\models K_\alpha \odot_{\mathcal{S}} [\alpha\ \texttt{Cstit}]G$. In Example 1.5, we take the information sets to be $\{\langle m_1, h_1 \rangle, \langle m_2, h_3 \rangle\}$ (bets heads), and $\{\langle m_1, h_2 \rangle, \langle m_2, h_4 \rangle\}$ (refrains from betting). The problem is solved because $\langle m_i, h_i \rangle \models \odot_{\mathcal{S}}[\alpha\ \texttt{Cstit}]\neg G$ and $\langle m_i, h_i \rangle \models K_\alpha \odot_{\mathcal{S}} [\alpha\ \texttt{Cstit}]\neg G$ (notice that $\mathbf{S}-\mathbf{optimal}^{m_1}_\alpha = \{\{h_2\}\}$, and $\mathbf{S}-\mathbf{optimal}^{m_2}_\alpha = \{\{h_4\}\}$). For Example 1.6, we take the information sets to be $\{\langle m_1, h_1 \rangle, \langle m_2, h_3 \rangle\}$ (bets heads), and $\{\langle m_1, h_2 \rangle, \langle m_2, h_4 \rangle\}$ (bets tails). The problem is then solved, for we obtain that $\langle m_i, h_i \rangle \not\models \odot_{\mathcal{S}}[\alpha\ \texttt{Cstit}]W$, which in turn implies that $\langle m_i, h_i \rangle \not\models K_\alpha \odot_{\mathcal{S}} [\alpha\ \texttt{Cstit}]W$ (notice that $\mathbf{S}-\mathbf{optimal}^{m_1}_\alpha = \{\{h_1\}, \{h_2\}\}$, and that $\mathbf{S}-\mathbf{optimal}^{m_2}_\alpha = \{\{h_3\}, \{h_4\}\}$). Therefore, although α knows that it objectively ought to win, it is not the case that it subjectively ought to win.

2.1 Axiomatization and some logical properties

In the unpublished work [3], an axiom system for the full logic that we have presented is shown to be sound and complete with respect to the class of epistemic utilitarian (bi-valued) KCo-models. These models are more general than the ones described above –instead of only one value function, there are two: one for the objective ought-to-do's, and the other for the subjective ones. The property of 'uniformity of strategies' is axiomatized by the schema $\Diamond K_\alpha \varphi \rightarrow K_\alpha \Diamond \varphi$, which corresponds to the frame condition (US): for every situation $\langle m_*, h_* \rangle$, if $\langle m_*, h_* \rangle \sim_\alpha \langle m, h \rangle$ for some $\langle m, h \rangle$, then for every $h'_* \in H_{m_*}$, there exists $h' \in H_m$ such that $\langle m_*, h'_* \rangle \sim_\alpha \langle m, h' \rangle$. The properties of the subjective oughts are characterized by $K_\alpha \Box \varphi \rightarrow \odot_{\mathcal{S}}[\alpha\ \texttt{Cstit}]\varphi$ (subjective necessity), $\odot_{\mathcal{S}}[\alpha\ \texttt{Cstit}]\varphi \rightarrow \Diamond K_\alpha [\alpha\ \texttt{Cstit}]\varphi$ (ought implies can), and $\odot_{\mathcal{S}}[\alpha\ \texttt{Cstit}]\varphi \rightarrow K_\alpha \Box \odot_{\mathcal{S}} [\alpha\ \texttt{Cstit}]\varphi$ (closure). Regarding the interaction of the two versions of ought-to-do, it turns out that neither of them logically implies the other.

3 CONCLUSION

Our work shows that to model practical knowledge, agency, and obligation in *stit* logic, a better alternative than the use of types is to distinguish between objective and subjective ought-to-do's.

REFERENCES

[1] R. Alur, T.A. Henzinger, and O. Kupferman. 2002. Alternating-time Temporal Logic. *J. ACM* 49, 5 (2002), 672–713.
[2] Nuel D Belnap, Michael Perloff, and Ming Xu. 2001. *Facing the future: agents and choices in our indeterminist world.* Oxford University Press on Demand.
[3] J. Broersen and A. Ramírez Abarca. 2017. Completeness and Decidability of a Logic of Subjective Oughts (work in progress). (2017).
[4] Andreas Herzig and Emiliano Lorini. 2010. A Dynamic Logic of Agency I: STIT, Capabilities and Powers. *Journal of Logic, Language and Information* 19, 1 (2010), 89–121.
[5] John Horty and Eric Pacuit. 2017. Action Types in Stit Semantics. *Review of Symbolic Logic* (2017). forthcoming.
[6] John F. Horty. 2001. *Agency and Deontic Logic.* Oxford University Press.
[7] Ming Xu. 2015. Combinations of Stit with Ought and Know. *Journal of Philosophical Logic* 44, 6 (2015), 851–877. DOI: http://dx.doi.org/10.1007/s10992-015-9365-7