ECAI 2008
M. Ghallab et al. (Eds.)
IOS Press, 2008

doi:10.3233/978-1-58603-891-5-747

Fred meets Tweety

Antonis Kakas[1] and **Loizos Michael**[2] and **Rob Miller**[3]

Abstract. We propose a framework that brings together two major forms of default reasoning in Artificial Intelligence: applying default property classification rules in static domains, and default persistence of properties in temporal domains. Particular attention is paid to the central problem of *qualification*. We illustrate how previous semantics developed independently for the two separate forms of default reasoning naturally lead to the integration that we propose, and how this gives rise to domains where different types of knowledge interact and qualify each other while preserving *elaboration tolerance*.

1 Introduction

Tweety is watching as we load the gun, wait, and then shoot Fred. Should we conclude that Tweety will fly away as birds normally do when they hear a loud noise as that normally produced by shooting a loaded gun? It depends on whether Tweety can fly! This belief, in turn, depends on whether Tweety is only known to be a bird, or also known to be a penguin. What can we conclude about Fred if after the act of shooting Fred we observe that Tweety is still on the ground?

In this problem of *"Fred meets Tweety"* we need to bring together two major forms of default reasoning that have been extensively studied on their own in A.I., but have rarely been addressed in the same formalism. These are *default property classification* as applied to inheritance systems [4, 10], and *default persistence* central to temporal reasoning in theories of Reasoning about Action and Change (RAC) [3, 9, 11]. How can a formalism synthesize the reasoning encompassed within each of these two forms of default reasoning?

Central to these two (and indeed all) forms of default reasoning is the *qualification problem*: default conclusions are qualified by information that can block the application of the default inference. Recent work has shown the importance for RAC theories to properly account for different forms of qualification [5, 12]. In our problem of integrating the default reasoning of property classification into RAC, we study how a *static* default theory expressing known default relationships between fluents can endogenously qualify the reasoning about actions and change, so that the application of causal laws and default persistence is properly adjusted by this static theory.

2 Knowledge Qualification

One of the first knowledge qualification problems formally studied in A.I. relates to the *Frame Problem* (see, e.g., [11]) of how the causal change properly qualifies the default persistence. In the archetypical Yale Shooting Problem domain [3], a turkey named Fred is initially alive, and one asks whether it is still alive after loading a gun, waiting, and then shooting Fred. The lapse of time cannot cause the gun to become unloaded. Default persistence is qualified only by known events and known causal laws linked to these events.

The consideration of indirect action effects gave rise to the *Ramification Problem* (see, e.g., [7]) of how these effects are generated and qualify persistence. Static knowledge expressing domain constraints was introduced to encode such indirect action effects. In early solutions to the Ramification Problem a direct action effect would cause this static knowledge to be violated, unless a minimal set of indirect action effects were also assumed so as to maintain consistency [7, 8]. Thus, given the static knowledge that "dead birds do not walk", the shooting action causing Fred to be dead would also indirectly cause Fred to stop walking, qualifying the persistence of the latter property.

Subsequent work examined default causal knowledge, bringing to focus the *Qualification Problem*[4] (see, e.g., [12]) of how such default causal knowledge is qualified by domain constraints. In some solutions to the Qualification Problem, the role of static knowledge within the domain description was identified as that of *endogenously* qualifying causal knowledge, as opposed to aiding causal knowledge in qualifying persistence [5]. Observations after action occurrences also qualify causal change when the two are in conflict, a problem known as the *Exogenous Qualification Problem* (see, e.g., [5]).

Independently of the above, another qualification problem was examined in the context of *Default Static Theories* [10] that consider how observed facts qualify default static knowledge. In the typical domain one asks whether Tweety is able to fly, when it is *only* known to be a bird. In the absence of any explicit information on whether Tweety is able to fly, the theory predicts that it is, but retracts this prediction once the extra fact that Tweety is a penguin is added.

In this paper we investigate temporal domains that incorporate (possibly) *default* static theories. The technical challenge lies in understanding how the four types of knowledge in a domain, three of which may now be default, interact and qualify each other. To illustrate some of these interactions we employ the syntax of the action description language $\mathcal{ME}$ [5]. Strict static knowledge is encoded in propositional logic. Default static knowledge is encoded in terms of default rules of the form *"$\phi \rightsquigarrow \psi$"*, where ϕ, ψ are propositional formulas; an informal reading of such default rules suffices for this section. Formulas with variables are used as a shorthand notation for the set of all of their groundings over a finite domain of constants.

ClapHands `causes` *Noise*	<u>*static theory:*</u>
Noise `causes` *Fly(x)*	*(1) Penguin(x)* $\rightsquigarrow \neg$*CanFly(x)*
Noise `causes` $\neg$*Noise*	*(2) Penguin(x)* $\rightarrow$ *Bird(x)*
Penguin(Tweety) `holds-at` *1*	*(3) Bird(x)* $\rightsquigarrow$ *CanFly(x)*
ClapHands `occurs-at` *3*	*rule (1) overrides rule (3)*
ClapHands `occurs-at` *7*	*(4)* $\neg$*CanFly(x)* $\rightarrow \neg$*Fly(x)*

The default persistence of *"Penguin(Tweety)* `holds-at` *1"* implies, through the static theory, that *"¬CanFly(Tweety)"* holds everywhere. This, then, qualifies the causal generation of *"Fly(Tweety)"* by the

[1] University of Cyprus, P. O. Box 20537, CY-1678, Cyprus.
e-mail: antonis@ucy.ac.cy
[2] Harvard University, Cambridge, MA 02138, U.S.A.
e-mail: loizos@eecs.harvard.edu
[3] University College London, London WC1E 6BT, U.K.
e-mail: rsm@ucl.ac.uk

[4] Not to be confused with the broader sense of the term qualification we use.

action *"ClapHands"* at time-points 3 and 7. If, on the other hand, the observation *"Fly(Tweety)* holds-at *5"* is added, then the static theory is qualified itself, and does no longer qualify the causal generation of *"Fly(Tweety)"*. Note, however, that Tweety flies for an *exogenous* reason. If an action at time-point 6 were to cause Tweety to stop flying, this would *release* the static theory's default conclusion that penguins do not fly. The action *"ClapHands* occurs-at *7"* would then be qualified and would not cause Tweety to fly again.

What would happen if *"Noise"* was caused at time-point 3 because Fred, a turkey that is initially alive, was shot; and we only knew that Tweety is a bird? Then, we would conclude that Fred is dead from time-point 3 onwards, and also that Tweety is flying. If, however, one observes *"¬Fly(Tweety)* holds-at *4"*, then whether Fred is dead depends on why Tweety did not fly after Fred was shot! The observation by itself does not explain why the causal laws that would normally cause Tweety to fly were qualified. An endogenous explanation would be that Tweety is a penguin, and *"Fly(Tweety)"* is qualified from being caused. An exogenous explanation would be that Tweety could not fly due to exceptional circumstances (e.g., an injury). However, Tweety might not have flown because the shooting action failed to cause a noise, or because it failed altogether. Different conclusions on Fred's status might be appropriate depending on the explanation.

3 Formal Semantics of Integration

Four different types of information present in a framework of RAC interact and qualify each other: *(i)* information generated by default persistence, *(ii)* action laws that qualify default persistence, *(iii)* static default laws of fluent relationships that can qualify these action laws, and *(iv)* observations that can qualify any of these. This hierarchy of information comes full circle, as the bottom layer of default persistence of observations (which carry the primary role of qualification) can also qualify the static theory. Due to the cyclical nature of the qualifications, we develop the formal semantics in two steps.

For the *temporal semantics* we follow the semantics of $\mathcal{ME}$ [5], which accounts for the qualification of causal knowledge by a given *strict* static theory. Causal knowledge in $\mathcal{ME}$ is qualified so as to ensure that the static theory is never violated at the observable time scale. We extend that semantics by proposing that the qualification comes from an external set $\alpha(T)$ of admissible states that might depend on the time-point T. Thus, we end up with a semantics that given an externally provided ***admissibility requirement*** α, computes the temporal evolution of states so as to ensure that the state of the world at time-point T always lies within the set of admissible states $\alpha(T)$. The details of the temporal semantics of $\mathcal{ME}$ are largely orthogonal to the next step of determining how α is computed.

An ***externally qualified model of*** a domain description D ***given*** an admissibility requirement α is any mapping of time-points to states such that *(1)* the world is initially in an admissible state; *(2)* it changes in an admissible manner; and it holds that *(3.i)* literals not caused to change persist, and *(3.ii)* caused change is realized.

The admissibility requirement is determined by the static theory *after being qualified* by the combined effect of observations and persistence. We model this effect by considering ***virtual extensions of*** a domain D that contain additional ***virtual observations***. Virtual observations are not meant to capture abnormal situations, but rather persistence of known observations from other time-points. The minimization of virtual observations that we impose later guarantees that known observations persist only as needed to achieve this effect.

At every time-point T, we consider the static theory and the observations (including virtual ones) at T. The extensions of this default theory determine a particular set of admissible states $\alpha(T)$. An ***internally qualified model of*** a domain description D is an externally qualified model of D given this admissibility requirement α.

Given a domain description D, we consider its virtual extensions that have internally qualified models. Among those, we choose the ones with a minimal set of virtual observations. The internally qualified models of these virtual extensions of D are the ***models of*** D.

Observations in our semantics act as the knowledge that bootstraps reasoning. Since every other type of knowledge is amenable to qualification, a strong elaboration tolerance result can be established.

Theorem 1 (Elaboration Tolerance Theorem) *Let D be a consistent domain, D' a domain with no observations, and $D \cup D'$ their union, where the static theories of D and D' are merged together to form the single static theory of $D \cup D'$. We assume that the static theory of $D \cup D'$ is consistent. Then, $D \cup D'$ is a consistent domain.*

4 Concluding Remarks

We have presented an integrated formalism for reasoning with both default static and default causal knowledge, two problems that have been extensively studied in isolation from each other. The proposed solution applies to domains where the static knowledge is "stronger" than the causal knowledge, and qualifies excessive change caused by the latter. A more detailed exposition of our developed formalism, including a tentative solution of how to encode causal laws that are "stronger" than the static knowledge, appears in [6].

Our future research agenda includes further investigation of such "strong" causal knowledge, and of how "strong" static knowledge can generate extra (rather than block) causal change. We also plan to develop computational models corresponding to the presented theoretical framework, using, for example, ideas from argumentation.

Although we are unaware of any previous work explicitly introducing Fred to Tweety, much work has been done on the use of default reasoning in inferring causal change. In the context of the Qualification Problem see [2, 12]. For distinguishing between default and non-default causal rules in the context of the Language $\mathcal{C}+$ see [1].

REFERENCES

[1] S. Chintabathina, M. Gelfond, and R. Watson, 'Defeasible Laws, Parallel Actions, and Reasoning about Resources', in *Proc. of Commonsense'07*, pp. 35–40, (2007).

[2] P. Doherty, J. Gustafsson, L. Karlsson, and J. Kvarnström, 'TAL: Temporal Action Logics Language Specification and Tutorial', *ETAI*, **2**(3–4), 273–306, (1998).

[3] S. Hanks and D. McDermott, 'Nonmonotonic Logic and Temporal Projection', *AIJ*, **33**(3), 379–412, (1987).

[4] J. Horty, R. Thomason, and D. Touretzky, 'A Skeptical Theory of Inheritance in Nonmonotonic Semantic Networks', *AIJ*, **42**(2–3), 311–348, (1990).

[5] A. Kakas, L. Michael, and R. Miller, 'Modular-E: An Elaboration Tolerant Approach to the Ramification and Qualification Problems', in *Proc. of LPNMR'05*, pp. 211–226, (2005).

[6] A. Kakas, L. Michael, and R. Miller, 'Fred meets Tweety', in *Proc. of CogRob'08*, (2008).

[7] F. Lin, 'Embracing Causality in Specifying the Indirect Effects of Actions', in *Proc. of IJCAI'95*, pp. 1985–1991, (1995).

[8] F. Lin and R. Reiter, 'State Constraints Revisited', *J. of Logic and Comp.*, **4**(5), 655–678, (1994).

[9] J. McCarthy and P. Hayes, 'Some Philosophical Problems from the Standpoint of Artificial Intelligence', *Mach. Intel.*, **4**, 463–502, (1969).

[10] R. Reiter, 'A Logic for Default Reasoning', *AIJ*, **13**(1–2), 81–132, (1980).

[11] M. Shanahan, *Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia*, MIT Press, 1997.

[12] M. Thielscher, 'The Qualification Problem: A Solution to the Problem of Anomalous Models', *AIJ*, **131**(1–2), 1–37, (2001).