In the procedings of CogRob'08: the 6th International Cognitive Robotics Workshop, July 21-22, 2008, Patras, Greece, www.cse.yorku.ca/cogrob08

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 problem of *qualification*, central in default reasoning and in any attempt to integrate different forms of this type of reasoning. We examine previous semantics developed independently for the two separate forms of default reasoning, and illustrate how these naturally lead to the solution that we propose in integrating the two. The resulting integration gives rise to domains where four different types of knowledge interact and qualify each other in an intricate manner. Through a series of examples we show how this knowledge qualification leads to intuitive conclusions. We prove that our framework of integration is *elaboration tolerant*: extending a consistent domain with additional action occurrences, causal laws, or static knowledge does not render the domain inconsistent. The conclusions that are drawn are always adjusted so as to gracefully accommodate the extra knowledge.

1 Introduction

Tweety is watching as we prepare to shoot Fred. We load the gun, we wait, and then shoot the gun. Will we conclude that Tweety will fly away as birds normally do when they hear a loud noise that shooting a loaded gun normally produces? It depends on whether Tweety can fly or not! If all we know about Tweety is that it is a bird, we then expect to see it flying, but if we also know that it is a penguin we will not expect to see it flying, even if we hear a loud noise produced by the act of firing. What can we conclude if after the act of shooting we observe that Tweety is still on the ground? That Tweety is not a typical bird, or that the gun did not make a loud noise when fired, or even that the gun was not loaded at the time of shooting? Can we indeed conclude anything at all after such an unexpected observation?

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 [5, 10], and *default persistence* central to temporal reasoning in theories of Reasoning about Action and Change (RAC) [4, 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. One aspect of the qualification problem is to express within the theory the knowledge required to properly qualify and block the default inference under exceptional situations. This *endogenous form* of qualification is implicit in the theory, driven by auxiliary observations that enable the known qualifying information to be applied. For example, known exceptional classes in the case of default property inheritance, or known action laws (and their ramifications) in the case of default persistence, qualify respectively these two forms of default reasoning.

But this task of completely representing within a given theory the qualification knowledge is impractical and indeed undesirable, as we want to jump to default conclusions based on a minimal set of information available. We, therefore, also need to allow for default conclusions to be qualified unexpectedly from observed information that is directly (or explicitly) contrary to them. In this *exogenous* form of qualification the theory itself cannot account for the qualification of the default conclusion, but our observations tell us explicitly that this is so and we attribute the qualification to some unknown reason.

Recent work [6, 12] has shown the importance for RAC theories to properly account for these two forms of qualification, so that an exogenous qualification is employed only when observations cannot be accounted for by an endogenous qualification of the causal laws and default persistence. In our problem of integrating the default reasoning of property classification into RAC, this means that we need to ensure that the two theories properly qualify each other endogenously, so that the genuine cases of exogenous qualification can be correctly recognized. In particular, 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. In the Fred meets Tweety scenario described above, for example, the normal default that "penguins cannot fly" would act as an implicit qualification for the causal law that "a loud noise causes birds to fly", but not so when either Tweety is not known to be a penguin, or it is known to be a super-penguin (super-penguins being an exception to the default that penguins cannot fly).

More generally, we study how four different types of information present in such an integrated 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. Hence, in our proposed integrated framework, temporal projection with the observations help to determine the admissible states of the static default theory. In turn, admissible states qualify the actions laws and the temporal projection they generate.

Section 2 examines the qualification problem as studied in the two separate domains and its form for the proposed integration. Section 3 gives the formal semantics of the integration framework and the central result that ensures its elaboration tolerance. Section 4 briefly dis-

[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

cusses related and future work.

2 Knowledge Qualification

Through a series of examples, we present in this section the issues that arise when examining the qualification of knowledge, and place in context the various problems and solutions considered so far. We remark that we generally use the term *qualification* in a broader sense than that used in the context of Reasoning about Action and Change.

Here and throughout the paper we employ the syntax of the action description language $\mathcal{ME}$ [6] for temporal domain descriptions, and a pseudo-syntax based on that of propositional logic for representing static theories describing default or strict domain constraints. Strict static knowledge is represented in propositional logic. Default static knowledge is represented in terms of default rules of the form *"$\phi \rightsquigarrow \psi$"*, where ϕ, ψ are propositional formulas. In this pseudo-syntax we specify the relative strength between two default rules by statements of the form *"rule (i) overrides rule (j)"*. Formulas which contain variables are a shorthand representation of all formulas obtained by substituting the variables over a finite domain of constants.

We do not reproduce here the formal syntax for these theories. In particular, the formal semantics of our approach, given in the next section, will not depend on the specific form of the static theories, and different frameworks such as Default Logic [10] or argumentation [1] can be used. In this section it is sufficient for the reader to use the informal reading of the theories for their semantics.

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; see Figure 1(a). In the archetypical Yale Shooting Problem domain [4], 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 richer domains gave rise to the *Ramification Problem* (see, e.g., [7]) of how indirect action effects are generated and qualify persistence; see Figure 1(b). Static knowledge expressing relationships (or domain constraints) between different properties was introduced to encode these indirect effects. Then, 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 effects were also assumed in order 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, thus qualifying the persistence of Fred walking.

Subsequent work examined default causal knowledge, bringing to focus the *Qualification Problem* (see, e.g., [12]) of how such default causal knowledge is qualified by domain constraints; see Figure 1(c). In some solutions to the Qualification Problem, static knowledge within the domain description was identified as the knowledge that *endogenously* qualified causal knowledge, as opposed to as an aid to causal knowledge in qualifying persistence [6]. The Ramification Problem was now addressed by the explicit addition of causal laws, and the development of a richer semantics to account for their interaction. The following example domain illustrates a typical case.

Shoot(x) `causes` *FiredAt(x)*
FiredAt(x) `causes` *¬Alive(x)*
¬Alive(x) `causes` *¬Walks(x)*
Alive(Fred) `holds-at` *1*
Walks(Fred) `holds-at` *1*
Shoot(Fred) `occurs-at` *2*

<u>*static theory:*</u>

¬(¬Alive(x) and Walks(x))
¬(GunBroken and FiredAt(x))

Fix a model implying *"GunBroken* `holds-at` *1"*. Then we reason that the static theory (of domain constraints) qualifies the direct effect of the action *"Shoot(Fred)"* on *"FiredAt(Fred)"*, and hence it also prevents the indirect effect *"¬Walks(Fred)"* from being triggered. Thus, the default persistence of Fred walking is not qualified, and we conclude that Fred keeps walking. If, on the other hand, a model implies *"¬GunBroken* `holds-at` *1"*, then neither causal law is qualified by the static theory. Note that the effect *"¬Alive(Fred)"* is not qualified despite the observation *"Walks(Fred)* `holds-at` *1"*; the causal knowledge *"¬Alive(Fred)* `causes` *¬Walks(Fred)"* provides an escape route to this qualification. Hence, the default persistence of *"Walks(Fred)"* is qualified, and Fred is not walking after time-point 2. Models derived according to either of the two cases are valid.

Perhaps the next natural step was realizing that observations after action occurrences also qualify causal change when the two conflict, a problem known as the *Exogenous Qualification Problem* (see, e.g., [6]); see Figure 1(d). Consider, for example, the previous domain extended by the observation *"¬FiredAt(Fred)* `holds-at` *4"*. Even though the effect of the *"Shoot(Fred)"* is not, as we have seen, necessarily qualified by the static theory alone, the explicit observation that the action's direct effect is not produced leads us to conclude that it was necessarily qualified. The interaction with the endogenous qualification of the causal laws by the static theory comes from the fact that *"GunBroken"* together with the static theory qualifies the action law, and provides, thus, an explanation of the observed action failure. So, if we wish to minimize the unknown exogenous cases of qualification, we would conclude that *"GunBroken"* holds, as this is the only known way to endogenously account for the observed failure.

Independently of the study of qualification in a temporal setting, another qualification problem was examined in the context of *Default Static Theories* [10] that consider how observed facts qualify default static knowledge; see Figure 1(f). In the typical domain, represented below, one asks whether a bird named Tweety has the ability to fly, when the *only* extra given knowledge is that Tweety is a bird.

Bird(Tweety)

<u>*static theory:*</u>

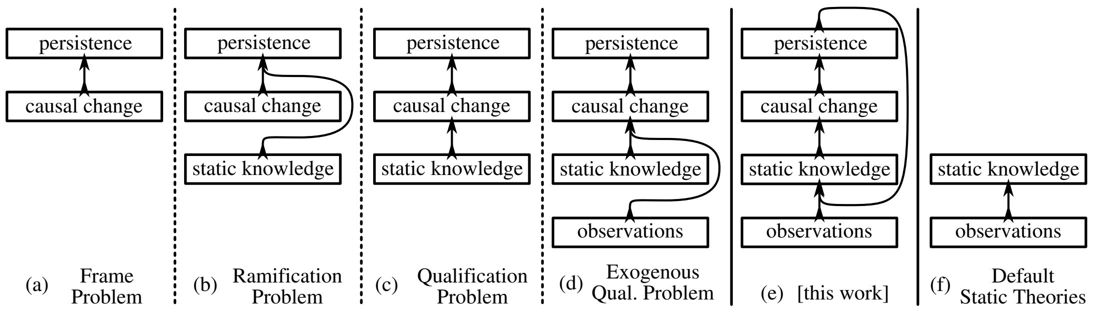
(1) Penguin(x) $\rightsquigarrow$ *¬CanFly(x)*
(2) Penguin(x) $\rightarrow$ *Bird(x)*
(3) Bird(x) $\rightsquigarrow$ *CanFly(x)*
rule (1) overrides rule (3)

In the absence of any explicit information on whether Tweety has the ability to fly, the theory predicts *"CanFly(Tweety)"*. Once extended with the fact *"Penguin(Tweety)"*, however, *"CanFly(Tweety)"* is retracted. The same happens if instead of *"Penguin(Tweety)"*, the fact *"¬CanFly(Tweety)"* is added. In either case the static theory is qualified, and yields to explicit facts or stronger evidence.

2.1 Putting Fred and Tweety in the Same Scene

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; see Figure 1(e).

We view observations as part of the non-defeasible part in static default theories, thus primarily taking the role of qualifying the static knowledge, which then in turn will qualify the causal knowledge as described above. Due to the temporal aspect of a domain, however, a point-wise interpretation of observations as facts in the static default theory is insufficient, *even* in domains with no causal laws and, thus, strict persistence. Consider a temporal domain with the observations *"Penguin(Tweety)* `holds-at` *1"* and *"Bird(Tweety)* `holds-at` *4"*,



(a) Frame Problem (b) Ramification Problem (c) Qualification Problem (d) Exogenous Qual. Problem (e) [this work] (f) Default Static Theories

Figure 1. Various solutions to the problem of knowledge qualification. Arrows point from the type of knowledge that qualifies to the type of knowledge that is being qualified. Leaf nodes in the graphs correspond to strict knowledge, and internal nodes correspond to default knowledge (qualified by its children nodes).

and a static theory as in the Tweety example above. By viewing each time-point in isolation, we can only conclude that *"CanFly(Tweety)"* holds at time-point 4, but not at time-point 1. This cannot be extended into a temporal model without violating the (strict) persistence. Instead, *"Penguin(Tweety)* `holds-at` *1"* should persist everywhere, as if *"Penguin(Tweety)"* was observed at every time-point. These *virtual (or assumed) observations* then qualify the static theory at every time-point, implying *"¬CanFly(Tweety)"*. Analogously, if the observation *"CanFly(Tweety)* `holds-at` *7"* is included in the domain, the observation persists everywhere and qualifies the default conclusion of the static theory that the penguin Tweety cannot fly.

Assume, now, that observations and persistence have appropriately qualified the static theory *at each time-point* T, so that the theory's default extensions (models) determine the set of *admissible* states at T. Through these sets of admissible states, the qualified static knowledge then qualifies the change that the theory attempts to generate through its causal knowledge. Given a time point T, it is natural that causal knowledge will be qualified by admissible states as determined *immediately after* T. This is illustrated in the next domain.

ClapHands `causes` *Noise*
Noise `causes` *Fly(x)*
Noise `causes` *¬Noise*
Spell(x) `causes` *CanFly(x)*
Penguin(Tweety) `holds-at` *1*
ClapHands `occurs-at` *3*
Spell(Tweety) `occurs-at` *5*
ClapHands `occurs-at` *7*

<u>*static theory:*</u>

(1) Penguin(x) $\rightsquigarrow$ *¬CanFly(x)*
(2) Penguin(x) $\rightarrow$ *Bird(x)*
(3) Bird(x) $\rightsquigarrow$ *CanFly(x)*
rule (1) overrides rule (3)
(4) Spell(x) $\rightsquigarrow$ *CanFly(x)*
rule (4) overrides rule (1)
(5) ¬CanFly(x) $\rightarrow$ *¬Fly(x)*

The default persistence of *"Penguin(Tweety)* `holds-at` *1"* implies that *"¬CanFly(Tweety)"* holds in each set of admissible states up to time-point 5. In particular, this conclusion holds immediately after *"ClapHands* `occurs-at` *3"*, and qualifies through the static theory the causal generation of *"Fly(Tweety)"* by the action *"ClapHands"*.

Intuitively, we expect *"Spell(Tweety)* `occurs-at` *5"* to override the static theory's default conclusion *"¬CanFly(Tweety)"* from holding at time-points following time-point 5. Note, however, that up to now we have assumed that the static default theory is stronger than the causal knowledge, and that it qualifies any change implied by the latter. But this is not the case now, since we wish to specify that some causal information is stronger than the static default theory. How, then, can we ensure that the causal generation of *"CanFly(Tweety)"* by *"Spell(Tweety)"* will not be qualified in this particular case?

This requirement is accommodated by including the particular causal law of interest *"Spell(x)* `causes` *CanFly(x)"* as a default rule *"Spell(x)* $\rightsquigarrow$ *CanFly(x)"* in the static theory, and giving this rule priority over other default rules of the static theory with the contrary conclusion.[4] The action occurrence *"Spell(Tweety)"* is also automatically included as a fact in the default theory, so that together with the default rule they imply *"CanFly(Tweety)"*. This conclusion holds in the set of admissible states associated with the time-point at which the action *"Spell(Tweety)"* occurred, namely time-point 5, which then allows the action's effect *"CanFly(Tweety)"* to override the static theory's usual default conclusion *"¬CanFly(Tweety)"*.

Such "strong" actions[5] (like *"Spell(x)"*) take the world out of the normal default state (where penguins cannot fly) into an exceptional, from the point of view of the static theory, state (where Tweety can fly). The rest of the default conclusions of the static theory still apply in this exceptional state (following time-point 5), conditioned on the exception (that Tweety can fly) that the "strong" action has brought about. This exception holds until some later action occurrence (of *"UndoSpell(Tweety)"*) brings the world back into its normal state. In our domain, then, the action *"ClapHands* `occurs-at` *7"* is not qualified, and Tweety (a penguin able to fly) flies after time-point 7.

Consider now replacing *"Spell(Tweety)* `occurs-at` *5"* in the domain above with the observation *"Fly(Tweety)* `holds-at` *5"*. By persistence, this observation qualifies the static theory so that *"Fly(Tweety)"* holds in each set of admissible states at time-points strictly after 3. Note that it is not known how the static theory is qualified, but only that it is somehow *exogenously* qualified. This does not hold for time-points up to and including time-point 3, since the occurrence of the action *"ClapHands"* at time-point 3 can now account for the change from *"¬Fly(Tweety)"* by qualifying its persistence, as the static theory does not now qualify *"ClapHands* `occurs-at` *3"*. Note that the interpretation of *"Fly(Tweety)* `holds-at` *5"* is that Tweety flies for some *exogenous* reason (e.g., it is on a plane). If an action at time-point 6 were to cause Tweety to stop flying, then this would *release* the static theory's default conclusion that penguins do not fly, so that the subsequent action *"ClapHands* `occurs-at` *7"* would be qualified and would not cause Tweety to fly again.

A somewhat orthogonal question to the one of *when* causal knowledge is qualified by the static theory, is that of *how* this qualification happens. Assume we wish to know if Fred is alive after firing at it. In the following domain one concludes that Fred is dead from time-point 2 onwards, and also that Tweety is flying. What happens, however, if one observes *"¬Fly(Tweety)* `holds-at` *4"*? Can one still conclude that Fred is dead? Interestingly enough, the answer depends on why Tweety did not fly after Fred was shot! The observation by it-

[4] We remind the reader that our goal here is not to provide semantics for static theories, and that using an informal reading suffices for their semantics.

[5] "Strong" actions are domain-dependent, and it is the domain designer's task to identify them and to extend the static theory with appropriate extra rules.

self does not explain why the causal laws that would normally cause Tweety to fly were qualified.

Shoot(x) `causes` *FiredAt(x)*
FiredAt(x) `causes` $\neg$*Alive(x)*
Shoot(x) `causes` *Noise*
Noise `causes` *Fly(x)*
Noise `causes` $\neg$*Noise*
Alive(Fred) `holds-at` *1*
Turkey(Fred) `holds-at` *1*
Bird(Tweety) `holds-at` *1*
Shoot(Fred) `occurs-at` *2*

static theory:
(1) Penguin(x) or Turkey(x)
$\rightsquigarrow$ $\neg$*CanFly(x)*
(2) Penguin(x) or Turkey(x)
$\rightarrow$ *Bird(x)*
(3) Bird(x) $\rightsquigarrow$ *CanFly(x)*
rule (1) overrides rule (3)
(4) $\neg$*CanFly(x)* $\rightarrow$ $\neg$*Fly(x)*

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). In either case we would presumably conclude that Fred is dead. However, Tweety might not have flown because the shooting action failed to cause a noise, or even because the shooting action failed altogether. Different conclusions on Fred's status might be reached depending on the explanation.

3 Formal Semantics of Integration

Due to the cyclical nature of the qualifications amongst different types of knowledge, we develop the formal semantics in two steps, starting from the *temporal semantics*. Thus, we start by assuming that the static theory is somehow qualified, and do not, for now, examine how this is achieved. This effectively breaks the cycle of qualifications, and reduces Figure 1(e) to Figure 1(c).We will then base our semantics on that of $\mathcal{ME}$ [6], from which we borrow the syntax.

A ***state*** is a complete and consistent set of positive or negative fluent literals in our problem domain language. A ***state change*** is a pair of states, comprised of an ***initial*** state, and a ***resulting*** state.

Definition 1 (Causal Node) *A **causal node** (or simply **node**) is a tuple $N = \langle S, B, P\rangle$, where S is a state, B is a set of action constants, and P is an active process log. Let $\mathcal{A}$ be a set of state changes. A pair $\langle\langle S_1, B_1, P_1\rangle, \langle S_2, B_2, P_2\rangle\rangle$ of causal nodes is an **admissible change under** $\mathcal{A}$ iff $\langle S_1, S_2\rangle$ is a state change in $\mathcal{A}$.*

Consider the domain description D^* of the last example in the previous section, which will serve as a running example in this section. Intuitively, then, one possible causal node associated with time-point 2 in D^* is $N_0^* = \langle S_0^*, B_0^*, \emptyset\rangle$, where $B_0^* = \{\textit{Shoot(Fred)}\}$, and the literals $\neg$*FiredAt(Fred)*, *Alive(Fred)*, *Turkey(Fred)*, $\neg$*Noise*, *Bird(Tweety)* are amongst those satisfied by (or belonging to) S_0^*.

A process $\mathrm{proc}(L)$ is ***triggered at*** a causal node $\langle S, B, P\rangle$ ***w.r.t.*** D iff the body C of a causal law *"C* `causes` *L"* holds in $S \cup B$. When the literal L is positive F (resp., negative $\neg F$), the triggered process $\mathrm{proc}(F) = \upharpoonleft F$ (resp., $\mathrm{proc}(\neg F) = \downharpoonleft F$) is ***initiating*** (resp., ***terminating***). All processes P_t triggered at a causal node $\langle S, B, P\rangle$ become part of the active process log, and the ***process successor of*** $\langle S, B, P\rangle$ is the (unique) causal node $\langle S, \emptyset, P \cup P_t\rangle$. In the example D^* above, the processes $\upharpoonleft$*FiredAt(Fred)* and $\upharpoonleft$*Noise* are (the only ones) triggered at N_0^* w.r.t. D^*, since the action constant in the bodies of the causal laws *"Shoot(x)* `causes` *FiredAt(x)"* and *"Shoot(x)* `causes` *Noise"* belongs in B_0^* when $x = \textit{Fred}$. Thus, the process successor of N_0^* w.r.t. D^* is $N_1^* = \langle S_0^*, \emptyset, \{\upharpoonleft\textit{FiredAt(Fred)}, \upharpoonleft\textit{Noise}\}\rangle$.

Processes in the active process log get resolved. A causal node $\langle S', \emptyset, P'\rangle$ is a ***resolvant of*** a causal node $N = \langle S, \emptyset, P\rangle$ iff either *(i)* $S' = S$ and $P' = P = \emptyset$, or *(ii)* $P' \subset P$, and S' differs from S on exactly those fluents in $P \setminus P'$, and is such that it satisfies F (resp., $\neg F$) when an initiating (resp., terminating) process for F is in $P \setminus P'$. Any non-empty subset of the processes can be resolved in a single step, so that multiple resolvants might be obtained. This captures the possibly asynchronous resolution of processes — unresolved processes remain in the process log and get resolved later. In our example, $N_2^* = \langle S_2^*, \emptyset, \{\upharpoonleft\textit{Noise}\}\rangle$ is *one of* the resolvants of N_1^*, where S_2^* differs from S_0^* only in that it satisfies *FiredAt(Fred)*.

Definition 2 (Causal Chain) *Let D be a domain description, $\mathcal{A}$ a set of state changes, and N_0 a causal node. A **causal chain rooted at** N_0 **w.r.t.** D is a (finite) sequence $N_0, N_1, \ldots, N_{2n}$ of causal nodes such that for each $k : 0 \le k \le n-1$, N_{2k+1} is a process successor of N_{2k} w.r.t. D and N_{2k+2} is a resolvant of N_{2k+1}, and such that every resolvant of the process successor of N_{2n} has the same state as N_{2n}. A causal chain $N_0, N_1, \ldots, N_{2n}$ is **admissible under** $\mathcal{A}$ **up to** N_{2k} iff the pair $\langle N_{2(j-1)}, N_{2j}\rangle$ of causal nodes is an admissible change under $\mathcal{A}$ for every $j : 1 \le j \le k \le n$, and either (i) $k = n$; or (ii) $\langle N_{2k}, N_{2(k+1)}\rangle$ is not an admissible change under $\mathcal{A}$. In the former case the causal chain is **fully admissible under** $\mathcal{A}$.*

Causal chains capture, thus, the triggering and resolution of (indirect) effects, until the state stabilizes. One causal chain rooted at N_0^* w.r.t. D^* is $N_0^*, \ldots, N_6^*$, where: $P_3^* = \{\upharpoonleft\textit{Noise}, \downharpoonleft\textit{Alive(Fred)}\}$; S_4^* differs from S_2^* only in that it satisfies *Noise*, $\neg$*Alive(Fred)*; $P_5^* = \{\downharpoonleft\textit{Noise}, \upharpoonleft\textit{Fly(Fred)}, \downharpoonleft\textit{Alive(Fred)}, \upharpoonleft\textit{Fly(Tweety)}\}$; and S_6^* differs from S_4^* only in that it satisfies $\neg$*Noise*, *Fly(Fred)*, *Fly(Tweety)*. The causal chain does not continue further; the process successor N_7^* of N_6^* contains in its process log P_7^* only the process $\downharpoonleft$*Alive(Fred)*, and all resolvants of N_7^* have the same state as N_6^*.

Each causal chain corresponds to a possible evolution path of the state of affairs at a fixed time point, as implied by a domain's causal knowledge. The static knowledge determines, through the notion of admissible change that it defines, whether a change between consecutive states in an evolution path is indeed allowed. If all possible evolution paths contain a non-admissible change, then the static theory suggests that the causal knowledge of the domain is flawed, and that the evolution of the state of affairs has stopped at *an unknown point* before reaching a non-admissible change (Condition *(ii)* below).

Definition 3 (Proper Causal Descendant) *Let D be a domain description, $\mathcal{A}$ a set of state changes, and N_0, N two causal nodes. N is a **proper causal descendant of** N_0 **w.r.t.** D **under** $\mathcal{A}$ iff either:*

(i) there exists a causal chain $N_0, N_1, \ldots, N_{2n}$ rooted at N_0 w.r.t. D that is fully admissible under $\mathcal{A}$ such that $N = N_{2n}$; or

(ii) there exists no causal chain rooted at N_0 w.r.t. D that is fully admissible under $\mathcal{A}$, and there exists $k : 0 \le k \le n-1$ and a causal chain $N_0, N_1, \ldots, N_{2n}$ rooted at N_0 w.r.t. D that is admissible under $\mathcal{A}$ up to N_{2k} such that $N = N_{2j}$ for some $j : 0 \le j \le k$.

It can be verified that the causal node N_6^* defined earlier is contained in each causal chain rooted at N_0^* w.r.t. D^*, with S_6^* satisfying, amongst others, the literals *Fly(Fred)* and *Turkey(Fred)*. Intuitively, the set $\mathcal{A}$ of state changes that corresponds to the static theory of D^* includes no state change with a resulting state that simultaneously satisfies *Fly(x)* and *Turkey(x)*. Hence, no pair $\langle N^*, N_6^*\rangle$ is an admissible change under $\mathcal{A}$, and, thus, no causal chain rooted at N_0^* is fully admissible under $\mathcal{A}$. So, Condition *(ii)* of Definition 3 is used.

We define now the temporal projection component of the semantics. Let Π be the set of time-points, and Φ the set of positive or negative fluent literals in the language. We assume an ***initial time-point*** $T_{in} \triangleq \min(\Pi)$, but do not assume discreteness or total ordering. Let $\overline{L}$ denote the negation of $L \in \Phi$; thus, if $L = \neg F$, then $\overline{L} = F$.

An ***interpretation*** H is a mapping of each fluent at each time-point to a truth-value. The ***state*** $S(H,T)$ ***at*** T ***w.r.t.*** H is the restriction of H to the time-point T. The ***event base*** $B(D,T)$ ***at*** T ***w.r.t.*** D is the set of action constants $\{A \mid \text{“}A\ \texttt{occurs-at}\ T\text{”} \in D\}$. An ***admissibility requirement*** α maps each time-point to a set of state changes.

A state S is ***stable in*** H ***at*** T ***w.r.t.*** D ***under*** α iff there exists a proper causal descendant $\langle S, \emptyset, P\rangle$ of $\langle S, \emptyset, \emptyset\rangle$ w.r.t. D under $\alpha(T)$. So, stable states do not spontaneously change, and take into account the causal knowledge and the admissibility requirements — the effects of any processes that could have been triggered have already be taken into account, so that no other change is "pending". We ask that the initial state at T_{in} in a temporal model of D satisfies this requirement. The change that occurs at each time-point T is determined by a proper causal descendant $\langle S, \emptyset, P\rangle$ of $\langle S(H,T), B(D,T), \emptyset\rangle$ w.r.t. D under $\alpha(T)$. The change $S \setminus S(H,T)$ that is brought about in the state of affairs is a ***change set of*** H ***at*** T ***w.r.t.*** D ***under*** α.

Definition 4 (Externally Qualified Model) *Let D be a domain description, H an interpretation, $c : \Pi \rightarrow 2^{\Phi}$ a mapping, and α_{st}, α_{ch} two admissibility requirements. H is an **externally qualified model** of D **under** $\langle \alpha_{st}, \alpha_{ch}\rangle$ **supported by** c iff the following hold:*

(1) $S(H, T_{in})$ is stable w.r.t. D under $\alpha_{st}(T_{in})$;

(2) for every $T \in \Pi$, $c(T)$ is a change set of H at T w.r.t. D under α_{ch};

(3) for every $L \in \Phi$, and every $T_1, T_3 \in \Pi$ s.t. $T_1 \prec T_3$:

(i) If H satisfies L at T_1, and there does not exist $T_2 \in \Pi$ s.t. $T_1 \preceq T_2 \prec T_3$ and $\overline{L} \in c(T_2)$, then H satisfies L at T_3;

(ii) If $L \in c(T_1)$, and there does not exist $T_2 \in \Pi$ s.t. $T_1 \prec T_2 \prec T_3$ and $\overline{L} \in c(T_2)$, then H satisfies L at T_3.

Hence, the world is initially in an admissible state of the static default theory (Condition *(1)*), and it changes in an admissible manner (Condition *(2)*) so that: literals not caused to change persist (Condition *(3.i)*), and caused change is realized (Condition *(3.ii)*).

3.1 Defining Admissibility w.r.t. a Static Theory

The static theory determines the admissibility requirements α_{st}, α_{ch} after being qualified by the combined effect of observations and persistence. We model this effect through ***virtual observations***, assumed to be part of a domain description D despite not being explicitly observed. Adding a set (Q) of such observation in D results in a ***virtual extension of*** D ***(by*** Q***)***. If D_1, D_2 are virtual extensions of D by Q_1, Q_2, respectively, and $Q_1 \subset Q_2$, then D_1 is ***preferred over*** D_2.

The domain description $D_1^* = D^*$ is a virtual extension of D^* by $Q_1^* = \emptyset$. The domain description D_2^* obtained from D^* by adding the observations in $Q_2^* = \{\text{“}\neg\textit{Fly(Tweety)}\ \texttt{holds-at}\ T\text{”} \mid T > 2\}$, is a virtual extension of D^* by Q_2^*. Clearly, D_1^* is preferred over D_2^*.

Note that virtual observations are not meant to capture abnormal situations. Instead, a virtual observation at T_{vrt} is simply interpreted as the persistence to T_{vrt} of a known observation at T_{obs}, providing a means for the known observation at T_{obs} to qualify the static theory at T_{vrt}. The minimization of virtual observations guarantees that known observations persist only as needed to achieve this effect.

Definition 5 (Internally Qualified Model) *An **internally qualified model** M **of** a domain description D' is an externally qualified model of D' under $\langle \alpha_{st}, \alpha_{ch}\rangle$ supported by c, iff for every $T \in \Pi$,*

(1) $\langle S_1, S_2\rangle \in \alpha_{st}(T)$ iff S_1, S_2 are models of the static theory in D' given as non-defeasible facts the literals observed in D' at T;

(2) $\langle S_1, S_2\rangle \in \alpha_{ch}(T)$ iff S_2 is a model of the static theory in D' given as non-defeasible facts (i) the literals observed in D' at each $T' \in (T, T + \varepsilon)$, for some $\varepsilon > 0$; (ii) the literals satisfied by both S_1 and S_2; and (iii) the action constants in $B(D', T)$.

A static theory's models map the theory's propositional symbols to truth-assignments that are compatible with the theory's default extensions. The semantics of these models is treated as a black-box, about which we only assume that a consistent set of input facts is satisfied by all (if any) models of the static theory; this rather benign assumption holds in typical default static theory semantics (e.g., [1, 10]).

Due to the existence of causal laws that may override the static knowledge, we distinguish between two admissibility requirements: *(1)* α_{st} ensures static admissibility at the initial state of affairs, and *(2)* α_{ch} ensures admissible change thereafter, taking into account previously caused exceptions (Condition *(2.ii)*). The two can be reduced to one if causal laws never override the static knowledge.

An internally qualified model of D_2^*, for instance, would imply $\text{“}\neg\textit{Fly(Tweety)}\ \texttt{holds-at}\ T\text{”}$ for every time-point $T > 2$. Indeed, since $\text{“}\neg\textit{Fly(Tweety)}\ \texttt{holds-at}\ T\text{”}$ appears in D_2^* for $T > 2$, then $\alpha_{ch}(T)$ only contains state changes with a resulting state satisfying $\neg$*Fly(Tweety)*. The overall effect, thus, is that the virtual observations in D_2^* qualify the causal knowledge so that Tweety does not fly.

Definition 6 (Model) *A **model** M of a domain description D is an internally qualified model of a virtual extension D' of D such that there exists no virtual extension D'' of D that has an internally qualified model, and such that D'' is preferred over D'.*

The virtual extension D_1^* of D^* has an internally qualified model, where from time-point 2 onwards Fred is dead and not flying, while Tweety is flying. The virtual extension D_2^* of D^* also has an internally qualified model, where Tweety is not flying. The preference of D_1^* over D_2^*, and over any other virtual extension of D^*, implies that the internally qualified models of D_1^* are also the models of D^*.

Note, in particular, that since virtual extensions of a domain are expected to have internally (and hence externally) qualified models, virtual observations in these virtual extensions are forced to respect the default persistence, as per Condition *(3.i)* of Definition 4.

The central role of observations in our semantics, as the knowledge that bootstraps reasoning, is consistent with Figure 1(e). Indeed, since every other type of knowledge is amenable to qualification, the following 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.*

Proof (sketch): Let D_{ext} be the virtual extension of D by $\{\text{“}L\ \texttt{holds-at}\ T\text{”} \mid S(M,T) \text{ satisfies } L\}$, for M a model of D. M can be shown to be an internally qualified model of $D_{ext} \cup D'$, which is a virtual extension of $D \cup D'$. This implies the claim. □

4 Concluding Remarks

We have proposed 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. Our proposed solution applies to domains where the static knowledge is "stronger" than the causal knowledge, and where it is appropriate for the former to qualify excessive change caused by the latter. Of course, these assumptions might not be appropriate for every domain. Our semantics already allows for "strong" causal laws to override static knowledge.

Our agenda for future research includes further investigation of such "strong" causal knowledge (constituting a different configuration in Figure 1.(c)), and of how "strong" static knowledge can generate extra (rather than block) causal change. We would also like to develop computational models corresponding to the theoretical framework presented here, using, for example, ideas from argumentation.

Although we are not aware of any previous work explicitly introducing Fred to Tweety, much work has been done on the use of default reasoning in inferring causal change. Of particular note in the context of the qualification problem are [3, 12]. An interesting approach to distinguishing between default and non-default causal rules in the context of the Language $\mathcal{C}+$ is given in [2].

REFERENCES

[1] A. Bondarenko, P. Dung, R. Kowalski, and F. Toni, 'An Abstract Argumentation-Theoretic Approach to Default Reasoning', *AIJ*, **93**(1–2), 63–101, (1997).

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

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

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

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

[6] 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).

[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).