Knowledge Qualification through Argumentation

Loizos Michael and Antonis Kakas

Department of Computer Science
University of Cyprus
{loizosm,antonis}@cs.ucy.ac.cy

Abstract. We propose a framework that brings together two major forms of default reasoning in Artificial Intelligence: default property classification in static domains, and default property persistence in temporal domains. Emphasis in this work is placed on the *qualification problem*, central when dealing with default reasoning, and in any attempt to integrate different forms of such reasoning.

Our framework can be viewed as offering a semantics to two natural problems: *(i)* that of employing default static knowledge in a temporal setting, and *(ii)* the dual one of temporally projecting and dynamically updating default static knowledge.

The proposed integration is introduced through a series of example domains, and is then formalized through argumentation. The semantics follows a pragmatic approach. At each time-point, an agent predicts the next state of affairs. As long as this is consistent with the available observations, the agent continues to reason forward. In case some of the observations cannot be explained without appealing to some *exogenous* reason, the agent revisits and revises its past assumptions.

We conclude with some formal results, including an algorithm for computing complete admissible argument sets, and a proof of *elaboration tolerance*, in the sense that additional knowledge can be gracefully accommodated in any domain.

1 Introduction

An important aspect of intelligence is the ability to reason, and draw conclusions about properties of one's environment that are not directly visible. In the area of logic-based reasoning, it is assumed that such conclusions are drawn by applying some knowledge base comprised of logic rules. To account for the inherent barrier of representing all possible knowledge for any but the simplest domains, *default* logic rules can be used in a knowledge base, so that conclusions are still drawn in the absence of sufficient information, but can be retracted in the presence of evidence to the contrary.

Two major forms of default reasoning have been extensively studied on their own in Artificial Intelligence, but have rarely been addressed in the same formalism. These are *default property classification* as applied to inheritance systems [1,2], and *default persistence* central to temporal reasoning in theories of Reasoning about Actions and Change (RAC) [3,4,5]. Here we consider the question [6]: 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

E. Erdem, F. Lin, and T. Schaub (Eds.): LPNMR 2009, LNCS 5753, pp. 209–222, 2009.

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, the static and temporal forms of default reasoning.

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 [7,8] has shown the importance for RAC theories to properly account for these two forms of qualification, so that exogenous qualification is employed only when observations cannot be accounted for by endogenous qualification of default persistence and the causal laws. When integrating default static and temporal theories, 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.

The mutual qualification of the default static and temporal theories can be understood in two dual ways: On the one hand, temporal reasoning is extended to include default static knowledge, which acts as a global, but defeasible, constraint that qualifies the temporal evolution of a domain. On the other hand, default static reasoning is extended to a temporal setting, where persistence, the effects of actions, and observations across the time-line can qualify the default laws that are used for reasoning.

In particular, we study how four different types of information present in such an integrated framework interact with, and can qualify, each other: *(i)* information generated by default persistence of fluents, *(ii)* action laws that can qualify default persistence, *(iii)* default static laws of fluent relationships that can qualify these action laws, and *(iv)* observations that can (exogenously) qualify any of these types of information. 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 default static theory. In turn, admissible states qualify the actions laws and the temporal projection they generate.

The semantics of the proposed integration is motivated and introduced through a series of example domains in Section 2, and formalized through an argumentation-based framework in Section 3. The framework follows a pragmatic approach. An agent holds certain beliefs about the state of the world at a certain time-point, executes some actions, and possibly makes some observations. It then predicts the next state of the world, by devising an admissible argument set, taking into account all the pieces of knowledge at its disposal, and resolving conflicts based on the strength of the corresponding arguments. If an admissible argument set that agrees with all available observations is found, the agent assumes that the computed state is the actual state of the world, and moves forward, repeating the process at the next time-point. This is repeated until the state predicted by the agent based solely on its knowledge, and without appealing to some

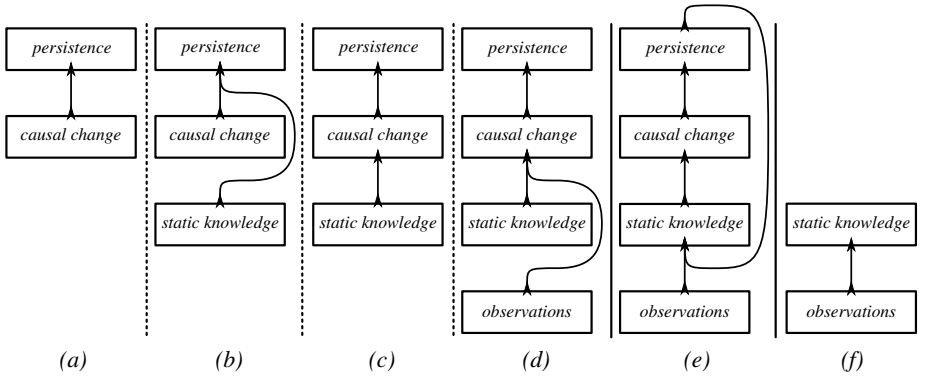

Fig. 1. Solutions to the knowledge qualification problem. Arrows point from the knowledge type that qualifies to the knowledge type being qualified. Root nodes in the graphs correspond to strict knowledge, and internal nodes correspond to default knowledge (qualified by its parent nodes).

exogenous reason, is in conflict with the observations. In such a case, the agent revisits its past assumptions, and revises them so as to resolve conflicts with the observations.

We establish some formal results in Section 4. From a computational point of view, we provide an algorithm that is guaranteed to compute a complete admissible argument set. From an epistemological point of view, we show that domains interpreted under our proposed semantics are *elaboration tolerant*, in the sense that they can be extended with, and gracefully accommodate, *arbitrary* pieces of knowledge. In particular, the semantics enjoys a *free-will* property, so called because an agent may choose to execute any action, without this causing an inconsistency and compromising its ability to reason. We conclude in Section 5, where we briefly discuss 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 Actions and Change.

For illustration purposes, we employ the syntax of the action description language $\mathcal{ME}$ [7] 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 classical 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 an appropriate 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 [2]

or argumentation [9] can be used. For the example domains in this section, it is sufficient for the reader to use the informal reading of the static theories for their semantics.

2.1 Past Investigations of Knowledge Qualification

One of the first knowledge qualification problems formally studied in Artificial Intelligence relates to the *Frame Problem* (see, e.g., [5]) of how the causal change properly qualifies the default persistence; see Fig. 1(a). 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 for some time, 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., [10]) of how indirect action effects are generated and qualify persistence; see Fig. 1(b). Static knowledge expressing relationships between different properties (i.e., domain constraints) was introduced to encode indirect 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 effects were also assumed in order to maintain consistency [10,11]. Thus, given the static knowledge that "dead birds do not walk", the shooting action that causes 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., [8]) of how such default causal knowledge is qualified by domain constraints; see Fig. 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 [7]. 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. A typical domain is shown below.

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*

static theory:

¬(¬*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 Fred keeps walking. If, on the other hand, a model implies *"¬GunBroken* `holds-at` *1"*, then no 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 cases are valid.

Perhaps the next natural step was realizing that observations after causal change, also, qualify the latter when the two are in conflict, a problem known as the *Exogenous Qualification Problem* (see, e.g., [7]); see Fig. 1(d). Consider, for example, the last 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* [2] that consider how observed facts qualify default static knowledge; see Fig. 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.

	static theory:
Bird(Tweety)	*(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 retraction happens if instead of the fact *"Penguin(Tweety)"*, the fact *"¬CanFly(Tweety)"* is added. In either case the static theory is qualified, and yields to explicit facts or stronger evidence to the contrary.

2.2 Putting Fred and Tweety in the Same Scene

In this paper we investigate how temporal domains can incorporate default static theories, or dually, how static theories should be revised when interpreted in a temporal setting, in the presence of default persistence, default causal change, and observations. The technical challenge lies in understanding how the four types of domain knowledge, three of which may now be default, interact and qualify each other; see Fig. 1(e).

We view observations as part of the non-defeasible part in default static 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 default static 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"*, and a static theory as in the Tweety example above. By viewing each time-point in isolation, we can conclude that *"CanFly(Tweety)"* holds only at time-point 4, but not at time-point 1. This cannot be extended into a temporal model

without violating the (strict) persistence. Rather, *"Penguin(Tweety)* `holds-at` *1"* should persist everywhere, as if *"Penguin(Tweety)"* were observed at every time-point. This persistence, then, qualifies the static theory at every time-point, and implies that *"¬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 temporal part of 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 type of qualification is illustrated in the next example 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*

static theory:

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

The 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"*.

This domain illustrates also a new aspect of the qualification problem. 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 default static 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 default static 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 can be accommodated by interpreting the particular causal law of interest *"Spell(x)* `causes` *CanFly(x)"* as a default rule in the static theory, and giving it priority over other default rules of the static theory with the contrary conclusion.[1] This interpretation need not be explicated. It suffices to mark *strong* causal laws as such, and then let their effects qualify the static theory, much in the same way that observations and persistence do. Because of this qualification, then, *"CanFly(Tweety)"* will hold in the set of admissible states associated with the time-point immediately following the occurrence of the action *"Spell(Tweety)"*, allowing the action's effect to come about and override the static theory's usual default conclusion that *"¬CanFly(Tweety)"*.

[1] We remind the reader that our goal here is not to provide semantics for static theories, and that using an informal reading in all presented example domains suffices for their semantics.

Such strong actions[2] (like *"Spell(x)"*) take the world out of the normal default state of affairs (where penguins cannot fly) into an exceptional, from the point of view of the static theory, state (where Tweety, a penguin, 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 by persistence until some later action occurrence (e.g., *"UndoSpell(Tweety)"*) brings the world back into its normal state. In our domain, the second occurrence of *"ClapHands"* is while in an exceptional state; thus, the causal change 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. 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 the observation *"Fly(Tweety)* `holds-at` *5"* is that Tweety flies for some *exogenous* reason (e.g., it is on a plane), and thus it is not known how the static theory is qualified, but only that it is somehow exogenously qualified. 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 by the static theory, 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 *how* this qualification happens. Consider the Fred meets Tweety domain [6] below, and assume we wish to know whether Fred is alive after firing at it. One concludes that Fred is dead from time-point 2 onwards, and also that Tweety is flying. What happens, however, if one were to observe *"¬Fly(Tweety)* `holds-at` *4"*? Could one still conclude that Fred is dead? Interestingly enough, the answer depends on why Tweety would not fly after Fred would be shot! The observation by itself does not explain why the causal laws that would normally cause Tweety to fly did not do so.

Shoot(x) `causes` *FiredAt(x)*
FiredAt(x) `causes` *¬Alive(x)*
Shoot(x) `causes` *Noise*
Noise `causes` *Fly(x)*
Noise `causes` *¬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) $\rightsquigarrow$ *¬CanFly(x)*
(2) Penguin(x) or Turkey(x) $\rightarrow$ *Bird(x)*
(3) Bird(x) $\rightsquigarrow$ *CanFly(x)*
rule (1) overrides rule (3)
(4) ¬CanFly(x) $\rightarrow$ *¬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

[2] The set of strong actions is domain-dependent, and it is the domain designer's task to identify them and to mark them as such in the domain provided to an agent for reasoning.

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 action failed altogether. Different conclusions on Fred's status might be reached depending on the explanation.

3 Argumentation Semantics

Motivated by the discussion in Section 2, we propose in this section a formal semantics for the qualification problem in the context of integrating default static and temporal theories. Argumentation offers a natural framework for this purpose, as it allows the easy specification of different types of knowledge as arguments, and the specification of their relative strengths as preferences imposed over these arguments. This, in turn, provides a clean formalization for the non-monotonic nature of knowledge qualification. A list of numerous non-monotonic logics that have been (re-) formulated in terms of argumentation, and a discussion of how argumentation offers a uniform formalism for understanding non-monotonic reasoning, can be found in [12] and references therein.

We emphasize that our proposed semantics does not hinge on any particular syntax or semantics used in the previous and this section for illustration purposes. In particular, we take a black-box approach to the syntax and semantics of default static theories, and assume simply that we have access to their models, without concerning ourselves with how these models are derived. For the temporal part of our semantics, we follow a pragmatic approach. We first focus on defining how an agent can reason from what holds in the current state of affairs to what will hold in the subsequent one. We discuss later how this single-step approach can be extended across the entire time-line.

We assume a time structure defined over the non-negative integers. Fix a positive integer T, and a state of affairs $\mathcal{E}$ that is believed to hold at time-point $T-1$. Given a domain $\mathcal{D}$ expressed in some syntax, and interpreted according to some semantics, one derives a set of arguments of what holds at time-point T. Again, we take a black-box approach here, and do not concern ourselves with how these arguments are derived.

Definition 1. *Denote by $\mathcal{U}_{\mathcal{D},\mathcal{E},T}$ the **argument universe for** domain $\mathcal{D}$ **at** time-point T **given** state $\mathcal{E}$. $\mathcal{U}_{\mathcal{D},\mathcal{E},T}$ comprises* $\texttt{stat}$*, and arguments of the form* $\texttt{argm}(L)$*, namely* $\texttt{assm}(L)$*,* $\texttt{pers}(L)$*,* $\texttt{ngen}(L)$*,* $\texttt{sgen}(L)$*,* $\texttt{exog}(L)$*, as determined by the causal, static, and narrative parts of the given domain $\mathcal{D}$, and assuming the state $\mathcal{E}$ holds at $T-1$.*

Assumption arguments $\texttt{assm}(L)$ are necessary only at time-point 0, where, in fact, they can be thought of as a special case of generation arguments. Beyond this, assumptions are useful only for ease of presentation, and perhaps from a computational point of view in abstracting the past by postulating that something holds without a proof. Persistence arguments $\texttt{pers}(L)$ exist exactly if L holds in the current state of affairs $\mathcal{E}$.

Normal generation arguments $\texttt{ngen}(L)$ exist when causal change is triggered by some action occurrence, associated with a set of causal laws.[3] These normal generation

[3] As already illustrated in Section 2, certain RAC frameworks follow the approach that direct action effects may trigger other indirect effects. For ease of presentation, we do not make this distinction here, and focus on the more fundamental problem of how action effects (direct and indirect alike) interact with other pieces of knowledge. We note, however, that conditional arguments could be introduced so as to properly accommodate for indirect effects.

arguments are assumed to be qualified by the static theory; intuitively, we think of static theory as a compiled form of normal causal knowledge. Strong generation arguments $\texttt{sgen}(L)$ are similar to normal generation arguments, but they exist when the action effects are produced through strong causal laws, as these are defined in a domain. These strong generation arguments do not yield to the static theory, but, rather, override it.

Exogenous arguments $\texttt{exog}(L)$ exist when L is observed to hold at time-point T. Observations in a domain do not capture causal or static knowledge explaining *why* the environment reaches a particular state. Instead, they postulate that something holds for reasons *exogenous* to the causal and static theory. As such, observations are linked to the exogenous arguments of our framework in a one-to-one correspondence.

The static argument `stat` serves to indicate that the static theory is to be taken into account. This may give rise to additional conclusions, but also opens up the possibility for some of these conclusions to be questioned in lieu of stronger counter-arguments.

The relative strengths of various types of knowledge are captured by imposing preferences between the corresponding arguments; see Definition 2. As expected, assumptions are qualified by every other type of knowledge, while observations (i.e., exogenous reasons) qualify all other types of knowledge. The static theory qualifies normal causal change, but it is qualified by strong causal change, persistence, and observations. Finally, persistence is qualified by causal change. We have made the working assumption that strong causal change is incomparable in strength to normal causal change; that is, their only difference is with respect to the static theory. This assumption is retractable, and does not affect any of the definitions or results that follow in any important way.

Definition 2. *Define a* ***preference*** *relation $\succ$ between pairs of arguments, so that for every literal L, the arguments on the left are preferred over those on the top in the table:*

	$\texttt{stat}$	$\texttt{assm}(\overline{L})$	$\texttt{ngen}(\overline{L})$	$\texttt{pers}(\overline{L})$	$\texttt{sgen}(\overline{L})$	$\texttt{exog}(\overline{L})$
$\texttt{stat}$		$\succ$	$\succ$			
$\texttt{assm}(L)$						
$\texttt{ngen}(L)$		$\succ$		$\succ$		
$\texttt{pers}(L)$	$\succ$	$\succ$				
$\texttt{sgen}(L)$	$\succ$	$\succ$		$\succ$		
$\texttt{exog}(L)$	$\succ$	$\succ$	$\succ$	$\succ$	$\succ$	

Since the static theory is qualified by other pieces of knowledge, we need to assume that the static theory is associated with a revision mechanism. Following our black-box approach, we make no assumptions on what this mechanism is, beyond its existence.

Definition 3 (Dynamic Revision of Static Theory). *It is assumed that there exists a fixed revision function $\texttt{rev}(\cdot,\cdot)$ that given a static theory and a set of literals, revises the static theory so that it entails all literals in the set. Given an argument set $\mathcal{A} \subseteq \mathcal{U}_{\mathcal{D},\mathcal{E},T}$, define $\mathcal{Q}(\mathcal{A}) \triangleq \{L \mid \texttt{exog}(L) \in \mathcal{A}$ or $\texttt{sgen}(L) \in \mathcal{A}$ or $\texttt{pers}(L) \in \mathcal{A}\}$. The* ***associated*** *static theory $\mathcal{S}_{\mathcal{D},\mathcal{A}}$* ***of*** *an argument set $\mathcal{A} \subseteq \mathcal{U}_{\mathcal{D},\mathcal{E},T}$* ***under*** *a domain $\mathcal{D}$ with a static theory $\mathcal{S}_{\mathcal{D}}$, is defined to be $\texttt{rev}(\mathcal{S}_{\mathcal{D}}, \mathcal{Q}(\mathcal{A}))$.*

Those types of knowledge in an argument set that are preferred over the static theory are captured by $\mathcal{Q}(\mathcal{A})$, and are used to qualify the static theory. This revised static theory is then used to draw conclusions and possibly qualify other types of knowledge.

Definition 4 (Argument Set Entailment and Completeness). *An argument set $\mathcal{A} \subseteq \mathcal{U}_{\mathcal{D},\mathcal{E},T}$ **entails** a literal L, denoted $\mathcal{A} \models L$, if either $\texttt{argm}(L) \in \mathcal{A}$, or $\texttt{stat} \in \mathcal{A}$ and L holds in all those models of $\mathcal{S}_{\mathcal{D},\mathcal{A}}$ that are consistent with every literal G such that $\texttt{argm}(G) \in \mathcal{A}$. An argument set $\mathcal{A} \subseteq \mathcal{U}_{\mathcal{D},\mathcal{E},T}$ **minimally entails** a literal L if $\mathcal{A} \models L$ and there exists no argument set $\mathcal{A}' \subset \mathcal{A}$ such that $\mathcal{A}' \models L$.*

*An argument set $\mathcal{A} \subseteq \mathcal{U}_{\mathcal{D},\mathcal{E},T}$ is **complete for** a fluent F if either $\mathcal{A} \models F$ or $\mathcal{A} \models \overline{F}$. An argument set $\mathcal{A} \subseteq \mathcal{U}_{\mathcal{D},\mathcal{E},T}$ is **complete** if $\mathcal{A}$ is complete for every fluent. A complete argument set $\mathcal{A} \subseteq \mathcal{U}_{\mathcal{D},\mathcal{E},T}$ **entails a state** $\mathcal{E}'$ if $\mathcal{A}$ entails every literal in $\mathcal{E}'$.*

Definitions for attacks and admissibility are given next, in a manner that closely follows corresponding definitions in the literature (see, e.g., [13]). We emphasize this point, since it allows one to use existing and well-studied argumentation frameworks, and exploit computational models that have been developed for those (see, e.g., [12]).

Definition 5 (Attacking Relation). *An argument set $\mathcal{A}_1 \subseteq \mathcal{U}_{\mathcal{D},\mathcal{E},T}$ **attacks** an argument set $\mathcal{A}_2 \subseteq \mathcal{U}_{\mathcal{D},\mathcal{E},T}$ (**on** the literal L) if $\mathcal{A}_1 \models \overline{L}$ and $\mathcal{A}_2 \models L$, and there exist argument sets $\mathcal{A}_1^m \subseteq \mathcal{A}_1$ and $\mathcal{A}_2^m \subseteq \mathcal{A}_2$ such that the following conditions hold:*

(i) $\mathcal{A}_1^m$ minimally entails $\overline{L}$ and $\mathcal{A}_2^m$ minimally entails L;
(ii) if an argument in $\mathcal{A}_2^m$ is preferred over an argument in $\mathcal{A}_1^m$, then an argument in $\mathcal{A}_1^m$ is preferred over an argument in $\mathcal{A}_2^m$.

Definition 6 (Admissibility). *An argument set $\mathcal{A} \subseteq \mathcal{U}_{\mathcal{D},\mathcal{E},T}$ is **admissible** if the following conditions hold:*

(i) $\mathcal{A}$ does not attack itself;
(ii) $\mathcal{A}$ attacks every argument set $\mathcal{A}' \subseteq \mathcal{U}_{\mathcal{D},\mathcal{E},T}$ that attacks $\mathcal{A}$.

We now have the necessary machinery to formalize the integration of default static and temporal theories for the single time-step case. For a domain $\mathcal{D}$, an agent starts with a state $\mathcal{E}$ at time-point $T-1$, and a set of available arguments $\mathcal{U}_{\mathcal{D},\mathcal{E},T}$, and constructs a complete admissible argument set $\mathcal{A} \subseteq \mathcal{U}_{\mathcal{D},\mathcal{E},T}$. In turn, $\mathcal{A}$ entails a state that — according to the information available to the agent — is the state of affairs at time-point T. It is straightforward to extend this to multiple time-steps, where the prediction from a time-step serves as input to the next time-step. For convenience, and without loss of generality, we assume that there are no observations or causal effects at time-point 0.

Definition 7 (Pre-Models). *An **interpretation** $\mathcal{M}$ **of** a domain $\mathcal{D}$ is a total mapping from time-points to states, so that for every time-point $T \geq 0$, $\mathcal{M}(T)$ denotes the state associated with T. A **pre-model** $\mathcal{M}$ **of** a domain $\mathcal{D}$ **supported by** a mapping α is an interpretation of $\mathcal{D}$ such that $\mathcal{M}(0)$ is a model of $\mathcal{S}_{\mathcal{D}}$, and for every time-point $T > 0$, there is a complete admissible argument set $\alpha(T) \subseteq \mathcal{U}_{\mathcal{D},\mathcal{M}(T-1),T}$ that entails $\mathcal{M}(T)$.*

Recall that argument sets may contain exogenous arguments. Intuitively, these are used when the reason for which something holds is unknown, yet it is known that it does hold, since it was observed so. Appealing to such exogenous reasons should be minimized across the set of all argument sets used to construct a temporal model.

Definition 8 (Models). *A **model** $\mathcal{M}$ **of** a domain $\mathcal{D}$ is a pre-model of $\mathcal{D}$ supported by some α, such that there exists no pre-model $\mathcal{M}'$ of $\mathcal{D}$ supported by some α' that point-wise contains a subset of the exogenous arguments contained in α.*

In accordance to our pragmatic point of view, we propose that models of a domain be computed through a combination of forward and backward reasoning steps. Although it is beyond the scope of this work to devise a full computational procedure for the developed semantics, we briefly discuss how such a procedure would look like.

Initially the agent reasons forward, starting from some state at time-point 0, and computing the states at time-point 1, time-point 2, and so on. Whenever it executes an action, or makes an observation, it also reasons forward to compute the state of its environment at the next time-point. As long as the argument sets used in this reasoning process contain no exogenous arguments, the corresponding computed states are assumed to be part of some model (since point-wise they trivially minimize the use of exogenous arguments). If when computing the state at some time-point T, the need arises for an argument set to use exogenous arguments, then the agent enters the backward reasoning phase. It revisits the state at time-point $T-1$, and examines what assumptions it has to change so that the state at T can be computed without appealing to exogenous arguments. In the process of doing so, the need may arise to employ exogenous arguments for the argument set that entails the state at time-point $T-1$. If this is the case, the agent revisits the state at time-point $T-2$, and so on, going backwards possibly until time-point 0. Once the exogenous arguments are eliminated, or some minimal use of exogenous arguments is found to be necessary, the forward reasoning resumes.

4 Formal Results

We now discuss some formal properties of our proposed formalism. An algorithm for constructing complete admissible arguments sets is first presented and shown correct.

Given $\mathcal{U}_{\mathcal{D},\mathcal{E},T}$, construct the argument set $\mathcal{A}$ according to the following steps:

(1) Set $\mathcal{A} := \emptyset$, and set $\mathcal{L}$ to be the set of all literals.
(2) While there is $L \in \mathcal{L}$ s.t. $\texttt{pers}(L) \in \mathcal{U}_{\mathcal{D},\mathcal{E},T}$, and $\texttt{exog}(\overline{L}), \texttt{sgen}(\overline{L}), \texttt{ngen}(\overline{L}) \notin \mathcal{U}_{\mathcal{D},\mathcal{E},T}$, set $\mathcal{A} := \mathcal{A} \cup \{\texttt{pers}(L)\}$, and set $\mathcal{L} := \mathcal{L} \setminus \{\overline{L}\}$.
(3) While there is $L \in \mathcal{L}$ s.t. $\texttt{sgen}(L) \in \mathcal{U}_{\mathcal{D},\mathcal{E},T}$ and $\texttt{exog}(\overline{L}) \notin \mathcal{U}_{\mathcal{D},\mathcal{E},T}$, set $\mathcal{A} := \mathcal{A} \cup \{\texttt{sgen}(L)\}$, and set $\mathcal{L} := \mathcal{L} \setminus \{\overline{L}\}$.
(4) While there is $L \in \mathcal{L}$ s.t. $\texttt{exog}(L) \in \mathcal{U}_{\mathcal{D},\mathcal{E},T}$, set $\mathcal{A} := \mathcal{A} \cup \{\texttt{exog}(L)\}$, and set $\mathcal{L} := \mathcal{L} \setminus \{\overline{L}\}$.
(5) Set $\mathcal{A} := \mathcal{A} \cup \{\texttt{stat}\}$.
(6) While there is $L \in \mathcal{L}$ s.t. $\texttt{pers}(L) \in \mathcal{U}_{\mathcal{D},\mathcal{E},T}$, and both $\mathcal{A} \models L$ and there exists $\mathcal{A}^m \subseteq \mathcal{A}$ that minimally entails L, set $\mathcal{A} := \mathcal{A} \cup \{\texttt{pers}(L)\}$, and set $\mathcal{L} := \mathcal{L} \setminus \{\overline{L}\}$.
(7) While there is $L \in \mathcal{L}$ s.t. $\texttt{ngen}(L) \in \mathcal{U}_{\mathcal{D},\mathcal{E},T}$, and either $\mathcal{A} \not\models \overline{L}$ or there exists no $\mathcal{A}^m \subseteq \mathcal{A}$ that minimally entails $\overline{L}$, set $\mathcal{A} := \mathcal{A} \cup \{\texttt{ngen}(L)\}$, and set $\mathcal{L} := \mathcal{L} \setminus \{\overline{L}\}$.
(8) Return $\mathcal{A}$, and terminate.

Theorem 1 (Correctness of Construction). *For every domain $\mathcal{D}$, time-point T, and state $\mathcal{E}$, the algorithm above returns a complete admissible argument set $\mathcal{A} \subseteq \mathcal{U}_{\mathcal{D},\mathcal{E},T}$.*

Proof (sketch). Consider an argument set $\mathcal{A}' \subseteq \mathcal{U}_{\mathcal{D},\mathcal{E},T}$ such that for every literal L it holds that: *(i)* if $\texttt{sgen}(L) \in \mathcal{U}_{\mathcal{D},\mathcal{E},T}$, then $\texttt{pers}(L) \notin \mathcal{A}'$; and *(ii)* if $\texttt{exog}(L) \in \mathcal{U}_{\mathcal{D},\mathcal{E},T}$, then $\texttt{sgen}(L), \texttt{pers}(L) \notin \mathcal{A}'$. By case analysis it can be shown that for every argument $\texttt{argm}(G) \in \mathcal{A}'$, either $\texttt{argm}(G) \in \mathcal{A}$, or $\mathcal{A}$ attacks $\mathcal{A}'$ on literal G. It follows that if $\mathcal{A}$ does not defend an attack from $\mathcal{A}'$, then $\mathcal{A}' \subseteq \mathcal{A}$, which leads to a contradiction.

The case of arbitrary argument sets can be reduced to the special case considered above. Overall, then, $\mathcal{A}$ is admissible. The completeness of $\mathcal{A}$ follows easily. □

We now continue to establish an *elaboration tolerance* property: every domain has a model, as long as its static theory is not inconsistent to begin with.

Theorem 2 (Guaranteed Consistency of Domains). *For every domain $\mathcal{D}$, and every state $\mathcal{E}$ that is a model of $\mathcal{S}_\mathcal{D}$, there exists a model $\mathcal{M}$ of $\mathcal{D}$ such that $\mathcal{M}(0) = \mathcal{E}$.*

Proof (sketch). Theorem 1 immediately implies the existence of pre-models with the claimed property. This, then, implies the existence of a pre-model that minimizes the exogenous arguments, and is, thus, a model of the domain. □

As a special type of elaboration tolerance, we show that our formalism enjoys a *free-will* property: an agent may attempt to execute any sequence of actions in the future, without requiring revision of any of its beliefs about the past. The need for such a property in the context of Reasoning about Actions and Change has been argued in [7].

Theorem 3 (Free-Will Property of Reasoning). *Consider any two domains $\mathcal{D}_1, \mathcal{D}_2$ for which the following conditions hold: (i) neither domain has observations at time-points after T_0, and (ii) the domains differ only on the occurrences of actions whose effects are brought about at time-points after T_0. For every model $\mathcal{M}_1$ of $\mathcal{D}_1$ there exists a model $\mathcal{M}_2$ of $\mathcal{D}_2$ such that for every time-point $T \leq T_0$, $\mathcal{M}_1(T) = \mathcal{M}_2(T)$.*

Proof (sketch). Let $\mathcal{M}_1$ be a pre-model of $\mathcal{D}_1$ supported by α_1, and let $\mathcal{M}_2$ be a pre-model of $\mathcal{D}_2$ supported by α_2. By Theorem 1, it can be shown that α_2 can be chosen so that for every time-point $T \leq T_0$, $\alpha_1(T) = \alpha_2(T)$. Since there are no observations at time-points after T_0, it follows that for every time-point $T > T_0$, neither $\alpha_1(T)$ nor $\alpha_2(T)$ contain any exogenous argument. Since $\mathcal{M}_1$ minimizes the exogenous arguments, so does $\mathcal{M}_2$. Thus, $\mathcal{M}_2$ is a model of $\mathcal{D}_2$. □

Recall that one of the problems for which our proposed integration offers semantics is that of how to temporally project and dynamically update a static theory. We conclude this section by briefly reiterating the stance that our framework takes on this problem.

The original static theory $\mathcal{S}_\mathcal{D}$ is determined solely by a given domain $\mathcal{D}$. Since initially, at time-point 0, no temporal information (i.e., observations or causal effects) is present, this original static theory need not be revised. Indeed, according to our semantics (cf. Definition 7), the state of affairs at time-point 0 is consistent exactly with this original static theory $\mathcal{S}_\mathcal{D}$. As time progresses, however, observations and causal effects (from strong causal laws) become available. This information needs to be respected,

even if it is not consistent with the original static theory $\mathcal{S}_\mathcal{D}$. Additionally, if something holds in a past state $\mathcal{E}$ and is not caused to stop, its persistence needs, also, to be respected. According to our semantics (cf. Definition 3), all these pieces of information are taken into account to construct the revised static theory $\mathcal{S}_{\mathcal{D},\mathcal{A}}$ for some $\mathcal{A} \subseteq \mathcal{U}_{\mathcal{D},\mathcal{E},T}$; the state of affairs at time-point T is consistent exactly with this revised static theory. Note further that since temporal knowledge might be non-deterministic (e.g., due to non-deterministic causal effects, or due to conflicting observations at the same time-point), so might be the revision of the static theory. Indeed, the choice of the argument set $\mathcal{A}$ corresponds to a choice of one of the possible temporal evolutions of the world, and this, then, determines the revised static theory $\mathcal{S}_{\mathcal{D},\mathcal{A}}$ that corresponds to this choice.

5 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. The semantics was developed through argumentation, and follows a pragmatic point of view that we feel is appropriate for use in real-world settings.

Our agenda for future research includes investigation of scenarios where it is appropriate for static knowledge to generate extra (rather than block) causal change, when the former qualifies the latter. We would also like to develop a full-fledged computational procedure, along the lines already discussed in the preceding sections.

Beyond the work that introduced the problem and discussed some early ideas [6], we are not aware of other previous work that explicitly addresses the problem of integrating default static and temporal reasoning. However, 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 [14,8]. An interesting approach to distinguishing between default and non-default causal rules in the context of the Language $\mathcal{C}+$ is given in [15].

References

1. Horty, J., Thomason, R., Touretzky, D.: A Skeptical Theory of Inheritance in Nonmonotonic Semantic Networks. Artificial Intelligence 42(2-3), 311–348 (1990)
2. Reiter, R.: A Logic for Default Reasoning. Artificial Intelligence 13(1-2), 81–132 (1980)
3. Hanks, S., McDermott, D.: Nonmonotonic Logic and Temporal Projection. Artificial Intelligence 33(3), 379–412 (1987)
4. McCarthy, J., Hayes, P.: Some Philosophical Problems from the Standpoint of Artificial Intelligence. Machine Intelligence 4, 463–502 (1969)
5. Shanahan, M.: Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia. MIT Press, Cambridge (1997)
6. Kakas, A., Michael, L., Miller, R.: Fred meets Tweety. In: Proc. of the 18th European Conference on Artificial Intelligence (ECAI 2008), pp. 747–748 (2008)
7. Kakas, A., Michael, L., Miller, R.: Modular-E: An Elaboration Tolerant Approach to the Ramification and Qualification Problems. In: Baral, C., Greco, G., Leone, N., Terracina, G. (eds.) LPNMR 2005. LNCS (LNAI), vol. 3662, pp. 211–226. Springer, Heidelberg (2005)
8. Thielscher, M.: The Qualification Problem: A Solution to the Problem of Anomalous Models. Artificial Intelligence 131(1-2), 1–37 (2001)

9. Bondarenko, A., Dung, P., Kowalski, R., Toni, F.: An Abstract Argumentation-Theoretic Approach to Default Reasoning. Artificial Intelligence 93(1-2), 63–101 (1997)
10. Lin, F.: Embracing Causality in Specifying the Indirect Effects of Actions. In: Proc. of the 14th International Joint Conference on Artificial Intelligence (IJCAI 1995), pp. 1985–1991 (1995)
11. Lin, F., Reiter, R.: State Constraints Revisited. Journal of Logic and Computation 4(5), 655–678 (1994)
12. Kakas, A., Toni, F.: Computing Argumentation in Logic Programming. Journal of Logic and Computation 9, 515–562 (1999)
13. Kakas, A., Miller, R., Toni, F.: An Argumentation Framework for Reasoning about Actions and Change. In: Gelfond, M., Leone, N., Pfeifer, G. (eds.) LPNMR 1999. LNCS (LNAI), vol. 1730, pp. 78–91. Springer, Heidelberg (1999)
14. Doherty, P., Gustafsson, J., Karlsson, L., Kvarnström, J.: TAL: Temporal Action Logics Language Specification and Tutorial. Electronic Transactions on Artificial Intelligence 2(3-4), 273–306 (1998)
15. Chintabathina, S., Gelfond, M., Watson, R.: Defeasible Laws, Parallel Actions, and Reasoning about Resources. In: Proc. of the 8th International Symposium on Logical Formalizations of Commonsense Reasoning (Commonsense 2007), pp. 35–40 (2007)