Connecting Proof Theory and Knowledge Representation: Sequent Calculi and the Chase with Existential Rules

Tim S. Lyon[1], **Piotr Ostropolski-Nalewaja**[1,2]
1. Technische Universität Dresden
2. University of Wrocław
{timothy_stephen.lyon, piotr.ostropolski-nalewaja}@tu-dresden.de

Abstract

Chase algorithms are indispensable in the domain of knowledge base querying, which enable the extraction of implicit knowledge from a given database via applications of rules from a given ontology. Such algorithms have proved beneficial in identifying logical languages which admit decidable query entailment. Within the discipline of proof theory, sequent calculi have been used to write and design proof-search algorithms to identify decidable classes of logics. In this paper, we show that the chase mechanism in the context of existential rules is in essence the same as proof-search in an extension of Gentzen's sequent calculus for first-order logic. Moreover, we show that proof-search generates universal models of knowledge bases, a feature also exhibited by the chase. Thus, we formally connect a central tool for establishing decidability proof-theoretically with a central decidability tool in the context of knowledge representation.

1 Introduction

Existential Rules and the Chase. The formalism of existential rules is a significant sub-discipline within the field of knowledge representation, offering insightful results within the domain of ontology-based query answering (Baget et al. 2009), data exchange and integration (Fagin et al. 2005), and serving a central role in the study of generic decidability criteria (Feller et al. 2023).[1] Ontology-based query answering is one of the principal problems studied within the context of existential rules, and asks if a query is logically entailed by a given knowledge base (KB) $\mathcal{K} = (\mathcal{D}, \mathcal{R})$, where $\mathcal{D}$ is a database and $\mathcal{R}$ is a finite set of existential rules (Baget et al. 2011). Databases generally consist of positive atomic facts such as $Female(Marie)$ or $Mother(Zuza, Marie)$, while existential rules—which are first-order formulae of the form $\forall \boldsymbol{x}\boldsymbol{y}\beta(\boldsymbol{x}, \boldsymbol{y}) \rightarrow \exists \boldsymbol{z}\alpha(\boldsymbol{y}, \boldsymbol{z})$ with β and α conjunctions of atoms—are used to encode a logical theory or ontology that permits the extraction of implicit knowledge from the encompassing KB.

The primary tool for studying query answering within this setting is the so-called *chase*, an algorithm that iteratively saturates a given database under applications of existential rules (Beeri and Vardi 1984). The chase is useful in that it generates a *universal model* satisfying exactly those queries entailed by a KB, and thus, allows for the reduction of query entailment to query checking over the constructed universal model (Deutsch, Nash, and Remmel 2008). In this paper, we show how the chase corresponds to proof-search in an extension of Gentzen's sequent calculus, establishing a connection between a central tool in the theory of existential rules with the primary decidability tool in proof theory.

Sequent Calculi and Proof-Search. Since its introduction, Gentzen's sequent formalism (Gentzen 1935a; Gentzen 1935b) has become one of the preferred proof-theoretic frameworks for the creation and study of proof calculi. A sequent is an object of the form $\Gamma \vdash \Delta$ such that Γ and Δ are finite (multi)sets of logical formulae, and a sequent calculus is a set of inference rules that operate over such. Sequent systems, and generalizations thereof, have proved beneficial in establishing (meta)logical properties with a diverse number of applications, being used to write decision algorithms (Dyckhoff 1992; Slaney 1997), to calculate interpolants (Maehara 1960; Lyon et al. 2020), and have even been applied in knowledge intergation scenerios (Lyon and Gómez Álvarez 2022).

It is well-known that *geometric implications*, i.e. first-order formula of the form $\forall \boldsymbol{x}(\varphi \rightarrow \exists \boldsymbol{y}_1\psi_1 \lor \cdots \lor \exists \boldsymbol{y}_n\psi_n)$ with φ and ψ_i conjunctions of atoms, can be converted into an inference rules in a sequent calculus (Simpson 1994, p. 24). Since such formulae subsume the class of existential rules, we may leverage this insight to extend Gentzen's sequent calculus for first-order logic with such rules to carry out existential rule reasoning. When we do so, we find that sequent systems mimic existential rule reasoning and proof-search (described below) simulates the chase.

Proof-search is the central means by which decidability is obtained with a sequent calculus, and usually operates by applying the inference rules of a sequent calculus bottom-up on an input sequent with the goal of constructing a proof thereof. If a proof of the input is found, the input is confirmed to be valid, and if a proof of the input is not found, a counter-model can typically be extracted witnessing the invalidity of the input. We make the novel observation that counter-models extracted from proof-search (in the context of existential rules) are universal, being homomorphically

[1] Existential rules are also referred to as a *tuple-generating dependencies* (Abiteboul, Hull, and Vianu 1995), *conceptual graph rules* (Salvat and Mugnier 1996), Datalog$^{\pm}$ (Gottlob 2009), and $\forall\exists$-*rules* (Baget et al. 2011) in the literature.

equivalent to the universal model generated by the chase.

Contributions. Our contributions in this paper are as follows: (1) We establish a strong connection between tools in the domain of existential rules with that of proof theory; in particular, we show how to transform derivations with existential rules into sequent calculus proofs and vice versa. (2) We establish a correspondence between the chase and sequent-based proof-search, and (3) we recognize that proof-search, like the chase, generates universal models for knowledge bases, which is a novel, previously unknown insight regarding the capability of sequent systems.

Organization. The preliminaries are located in Section 2. In Section 3, we present the sequent calculus framework and write a proof-search algorithm that simulates the chase. Correspondences between existential rule reasoning and sequent-based reasoning are explicated in Section 4, and in Section 5, we conclude and discuss future research. We note that most proofs have been deferred to the appendix.

2 Preliminaries and Existential Rules

Formulae and Syntax. We let $\mathbf{C}$ and $\mathbf{V}$ be two disjoint denumerable sets of *constants* and *variables*. We use $a, b, c, \ldots$ to denote constants and $x, y, z, \ldots$ to denote variables. We define the set of *terms* to be $\mathbf{T} = \mathbf{C} \cup \mathbf{V}$, and we denote terms by t and annotated versions thereof. Moreover, we let $\mathbf{P} = \{p, q, r, \ldots\}$ be a denumerable set of *predicates* containing denumerably many predicates of each arity $n \in \mathbb{N}$, and use $ar(p) = n$ to denote that $p \in \mathbf{P}$ is of arity n. An *atom* is a formula of the form $p(t_1, \ldots, t_n)$ such that $t_1, \ldots, t_n \in \mathbf{T}$ and $ar(p) = n$. We will often write atoms as $p(\boldsymbol{t})$ with $\boldsymbol{t} = t_1, \ldots, t_n$. The *first-order language* $\mathcal{L}$ is defined via the following grammar in Backus–Naur form:

$$\varphi ::= p(\boldsymbol{t}) \mid \neg\varphi \mid \varphi \wedge \varphi \mid \exists x \varphi$$

such that $p \in \mathbf{P}$, $\boldsymbol{t} \in \mathbf{T}$, and $x \in \mathbf{V}$. We use φ, ψ, χ, ... to denote *formulae* from $\mathcal{L}$, and define $\varphi \vee \psi := \neg(\neg\varphi \wedge \neg\psi)$, $\varphi \rightarrow \psi := \neg\varphi \vee \psi$, and $\forall x \varphi := \neg\exists x \neg\varphi$. The occurrence of a variable is *free* in a formula φ when it does not occur within the scope of a quantifier. We let $\varphi(t/x)$ represent the formula obtained by substituting the term t for every free occurrence of the variable x in φ. We use Γ, Δ, Σ, ... to denote sets of formulae from $\mathcal{L}$, let $\mathbf{V}(\Gamma)$ denote the set of free variables in the formulae of Γ, and let $\mathbf{T}(\Gamma)$ denote the set of free variables and constants occurring in the formulae of Γ. We let $i \in [n]$ represent $1 \leq i \leq n$, and define a *ground atom* to be an atom $p(t_1, \ldots, t_n)$ such that for each $i \in [n]$, $t_i \in \mathbf{C}$. An *instance* $\mathcal{I}$ is defined to be a (potentially infinite) set of atoms, and a *database* $\mathcal{D}$ is defined to be a finite set of ground atoms. We let $\top$ be a special unary predicate and define $\mathcal{I}^\top = \mathcal{I} \cup \{\top(c) \mid c \in \mathbf{C}\}$. An instance $\mathcal{I}$ is referred to as an *interpretation iff* $\mathcal{I}^\top = \mathcal{I}$.

Substitutions. A *substitution* σ is defined to be a partial function over $\mathbf{T}$. A *homomorphism* from an instance $\mathcal{I}$ to an instance $\mathcal{J}$ is a substitution π from the terms of $\mathcal{I}$ to the terms of $\mathcal{J}$ such that (1) if $p(t_1, \ldots, t_n) \in \mathcal{I}$, then $p(\pi(t_1), \ldots, \pi(t_n)) \in \mathcal{J}$, and (2) $\pi(a) = a$, for each $a \in \mathbf{C}$. We say that an instance $\mathcal{I}$ *homomorphically maps* into an instance $\mathcal{J}$ *iff* a homomorphism exists from $\mathcal{I}$ to $\mathcal{J}$. Two instances $\mathcal{I}$ and $\mathcal{J}$ are defined to be *homomorphically equivalent*, written $\mathcal{I} \equiv \mathcal{J}$, *iff* each instance can be homomorphically mapped into the other. An $\mathcal{I}$-*assignment* is defined to be a substitution μ such that (1) $\mu(x) \in \mathbf{T}(\mathcal{I})$, for each $x \in \mathbf{V}$, and (2) $\mu(a) = a$, for each $a \in \mathbf{C}$. For an $\mathcal{I}$-assignment μ, we let $\mu(\varphi)$ denote the formula obtained by replacing each free variable of φ with its value under μ, and we let $\mu[\boldsymbol{t}/\boldsymbol{x}]$ be the same as μ, but where the variables $\boldsymbol{x}$ are respectively mapped to $\boldsymbol{t} \in \mathbf{T}$.

Models and Satisfaction. Given an interpretation $\mathcal{I}$ and an $\mathcal{I}$-assignment μ, we recursively define satisfaction $\models$ as:

(1) $\mathcal{I}, \mu \models p(t_1, \ldots, t_n)$ *iff* $p(\mu(t_1), \ldots, \mu(t_n)) \in \mathcal{I}$;
(2) $\mathcal{I}, \mu \models \neg\varphi$ *iff* $\mathcal{I}, \mu \not\models \varphi$;
(3) $\mathcal{I}, \mu \models \varphi \wedge \psi$ *iff* $\mathcal{I}, \mu \models \varphi$ and $\mathcal{I}, \mu \models \psi$;
(4) $\mathcal{I}, \mu \models \exists x \varphi$ *iff* $t \in \mathbf{T}(\mathcal{I})$ exists and $\mathcal{I}, \mu[t/x] \models \varphi$.

We say that $\mathcal{I}$ is a *model* of Γ and write $\mathcal{I} \models \Gamma$ *iff* for every $\varphi \in \Gamma$ and $\mathcal{I}$-assignment μ, we have $\mathcal{I}, \mu \models \varphi$. We define an instance $\mathcal{I}$ to be a *universal model* of Γ *iff* for any model $\mathcal{J}$ of Γ there exists a homomorphism from $\mathcal{I}$ to $\mathcal{J}$.

Existential Rules. An *existential rule* is a first-order formula $\rho = \forall \boldsymbol{x}\boldsymbol{y}\, \beta(\boldsymbol{x}, \boldsymbol{y}) \rightarrow \exists \boldsymbol{z}\, \alpha(\boldsymbol{y}, \boldsymbol{z})$ such that $\beta(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{body}(\rho)$ (called the body) and $\alpha(\boldsymbol{y}, \boldsymbol{z}) = \mathrm{head}(\rho)$ (called the head) are conjunctions of atoms over constants and the variables $\boldsymbol{x}, \boldsymbol{y}$ and $\boldsymbol{y}, \boldsymbol{z}$, respectively. We call a finite set $\mathcal{R}$ of existential rules a *rule set*. We define Γ to be $\mathcal{R}$-*valid iff* for every interpretation $\mathcal{I}$, if $\mathcal{I} \models \mathcal{R}$, then $\mathcal{I} \models \Gamma$.

Derivations and the Chase. We say that an existential rule ρ is *applicable* to an instance $\mathcal{I}$ *iff* there exists an $\mathcal{I}$-assignment μ such that $\mu(\beta(\boldsymbol{x}, \boldsymbol{y})) \subseteq \mathcal{I}$, and when this is the case, we say that $\tau = (\rho, \mu)$ is a *trigger* in $\mathcal{I}$. Given a trigger $\tau = (\rho, \mu)$ in $\mathcal{I}$ we define an *application* of the trigger τ to the instance $\mathcal{I}$ to be the instance $\tau(\mathcal{I}) = \mathcal{I} \cup \alpha(\mu(\boldsymbol{y}), \boldsymbol{z})$ where $\boldsymbol{z}$ is a tuple of fresh variables. We define a *chase derivation* $(\mathcal{I}_i, \tau_i)_{i \in [n]}$ to be a sequence $(\mathcal{I}_1, \tau_1), \ldots, (\mathcal{I}_n, \tau_n), (\mathcal{I}_{n+1}, \emptyset)$ such that for every $i \in [n]$, τ_i is a trigger in $\mathcal{I}_i$ and $\tau_i(\mathcal{I}_i) = \mathcal{I}_{i+1}$. For an instance $\mathcal{I}$ and a rule set $\mathcal{R}$, we define the *one-step chase* to be:

$$\mathbf{Ch}_1(\mathcal{I}, \mathcal{R}) = \bigcup_{\tau \text{ is a trigger in } \mathcal{I}} \tau(\mathcal{I}).$$

We let $\mathbf{Ch}_0(\mathcal{I}, \mathcal{R}) = \mathcal{I}$ as well as let $\mathbf{Ch}_{n+1}(\mathcal{I}, \mathcal{R}) = \mathbf{Ch}_1(\mathbf{Ch}_n(\mathcal{I}, \mathcal{R}), \mathcal{R})$. Finally, we define the *chase* to be $\mathbf{Ch}_\infty(\mathcal{I}, \mathcal{R}) = (\bigcup_{i \in \mathbb{N}} \mathbf{Ch}_i(\mathcal{I}, \mathcal{R}))^\top$, which serves as a universal model of $\mathcal{I} \cup \mathcal{R}$ (Deutsch, Nash, and Remmel 2008).[2]

Queries and Entailment. A *Boolean conjunctive query (or, BCQ)* is a formula $\exists \boldsymbol{x} q(\boldsymbol{x})$ such that $q(\boldsymbol{x})$ is a conjunction of atoms over the variables $\boldsymbol{x}$ and constants. We define a *knowledge base (or, KB)* to be an ordered pair $\mathcal{K} = (\mathcal{D}, \mathcal{R})$ with $\mathcal{D}$ a database and $\mathcal{R}$ a rule set, and let $\mathcal{I}$ be a *model* of $\mathcal{K}$, written $\mathcal{I} \models \mathcal{K}$, *iff* $\mathcal{I} \models \mathcal{D} \cup \mathcal{R}$. We write $\mathcal{K} \models \exists \boldsymbol{x} q(\boldsymbol{x})$ to mean that for every $\mathcal{I}$, if $\mathcal{I} \models \mathcal{K}$, then $\mathcal{I} \models \exists \boldsymbol{x} q(\boldsymbol{x})$. A chase derivation $(\mathcal{I}_i, \tau_i)_{i \in [n]}$ *witnesses* $(\mathcal{D}, \mathcal{R}) \models \exists \boldsymbol{x} q(\boldsymbol{x})$ *iff* $\mathcal{I}_1 = \mathcal{D}$, only rules from $\mathcal{R}$ are applied, and there exists

[2]We use a *restricted* variant of the chase; cf. (Fagin et al. 2005).

$$\frac{}{\Gamma, p(\boldsymbol{t}) \vdash p(\boldsymbol{t}), \Delta}(id) \qquad \frac{\Gamma \vdash \varphi, \Delta}{\Gamma, \neg\varphi \vdash \Delta}(\neg_L) \qquad \frac{\Gamma, \varphi \vdash \Delta}{\Gamma \vdash \neg\varphi, \Delta}(\neg_R) \qquad \frac{\Gamma, \varphi, \psi \vdash \Delta}{\Gamma, \varphi \wedge \psi \vdash \Delta}(\wedge_L)$$

$$\frac{\Gamma \vdash \varphi, \Delta \qquad \Gamma \vdash \psi, \Delta}{\Gamma \vdash \varphi \wedge \psi, \Delta}(\wedge_R) \qquad \frac{\Gamma, \varphi(y/x) \vdash \Delta}{\Gamma, \exists x\varphi \vdash \Delta}(\exists_L)\ y \text{ fresh} \qquad \frac{\Gamma \vdash \exists x\varphi, \varphi(t/x), \Delta}{\Gamma \vdash \exists x\varphi, \Delta}(\exists_R)\ t \in \mathbf{T}$$

Figure 1: The sequent calculus G3 for first-order logic.

an $\mathcal{I}_{n+1}$-assignment μ such that $\mu(q(\boldsymbol{x})) \subseteq \mathcal{I}_{n+1}$.

3 Sequent Systems and Proof-Search

We define a *sequent* to be an object of the form $\Gamma \vdash \Delta$ such that Γ and Δ are *finite* sets of formulae from $\mathcal{L}$. Typically, multisets are used in sequents rather than sets, however, we are permitted to use sets in the setting of classical logic; cf. (Kleene 1952). For a sequent $\Gamma \vdash \Delta$, we call Γ the *antecedent* and Δ the *consequent*. We define the *formula interpretation* of a sequent to be $f(\Gamma \vdash \Delta) = \bigwedge \Gamma \rightarrow \bigvee \Delta$.

The sequent calculus G3 (Kleene 1952) for first-order logic is defined to be the set of inference rules presented in Figure 1. It consists of the *initial rule* (id) along with *logical rules* that introduce complex logical formulae in either the antecedent or consequent of a sequent. The $(\exists_L)$ rule is subject to a side condition, stating that the rule is applicable only if y is *fresh*, i.e. y does not occur in the surrounding context Γ, Δ. The $(\exists_R)$ rule allows for the bottom-up instantiation of an existentially quantified formula with a term t. An *application* of a rule is obtained by instantiating the rule with formulae from $\mathcal{L}$. We call an application of rule *top-down* (*bottom-up*) whenever the conclusion (premises) is (are) obtained from the premises (conclusion).

It is well-known that every *geometric implication*, which is a formula of the form $\forall \boldsymbol{x}(\varphi \rightarrow \exists \boldsymbol{y}_1\psi_1 \vee \cdots \vee \exists \boldsymbol{y}_n\psi_n)$ with φ and ψ_i conjunctions of atoms, can be converted into an inference rule; see (Simpson 1994, p. 24) for a discussion. We leverage this insight to transform existential rules (which are special instances of geometric implications) into inference rules that can be added to the sequent calculus G3. For an existential rule $\rho = \forall \boldsymbol{x}\boldsymbol{y}\beta(\boldsymbol{x}, \boldsymbol{y}) \rightarrow \exists \boldsymbol{z}\alpha(\boldsymbol{y}, \boldsymbol{z})$, we define its corresponding *sequent rule* $s(\rho)$ to be:

$$\frac{\Gamma, \beta(\boldsymbol{x}, \boldsymbol{y}), \alpha(\boldsymbol{y}, \boldsymbol{z}) \vdash \Delta}{\Gamma, \beta(\boldsymbol{x}, \boldsymbol{y}) \vdash \Delta}\ s(\rho)\ \boldsymbol{z} \text{ fresh}$$

Note that we take the body $\beta(\boldsymbol{x}, \boldsymbol{y})$ and head $\alpha(\boldsymbol{y}, \boldsymbol{z})$ to be sets of atoms, rather than conjunctions of atoms, and we note that $\boldsymbol{x}, \boldsymbol{y}$ may be instantiated with terms in rule applications. Also, $s(\rho)$ is subject to the side condition that the rule is applicable only if all variables $\boldsymbol{z}$ are fresh. We define the sequent calculus $\mathsf{G3}(\mathcal{R}) = \mathsf{G3} \cup \{s(\rho) \mid \rho \in \mathcal{R}\}$. We define a *derivation* to be any sequence of applications of rules in $\mathsf{G3}(\mathcal{R})$ to arbitrary sequents, define an *$\mathcal{R}$-derivation* to be a derivation that only applies $s(\rho)$ rules, and define a *proof* to be a derivation starting from applications of the (id) rule. An example of a proof is shown on the left side of Figure 3.

Theorem 1 (Soundness and Completeness)**.** *$f(\Gamma \vdash \Delta)$ is $\mathcal{R}$-valid iff there exists a proof of $\Gamma \vdash \Delta$ in $\mathsf{G3}(\mathcal{R})$.*

We now define a proof-search algorithm that decides (under certain conditions) if a BCQ is entailed by a knowledge base. The algorithm `Prove` (shown in Figure 2) takes a sequent of the form $\mathcal{D} \vdash \exists \boldsymbol{x}q(\boldsymbol{x})$ as input and bottom-up applies inference rules from $\mathsf{G3}(\mathcal{R})$ with the goal of constructing a proof thereof. Either, `Prove` returns a proof witnessing that $(\mathcal{D}, \mathcal{R}) \models \exists \boldsymbol{x}q(\boldsymbol{x})$, or a counter-model to this claim can be extracted from failed proof search. The functionality of this algorithm depends on certain *saturation conditions*, defined in Definition 2 below, and which determine when a rule from $\mathsf{G3}(\mathcal{R})$ is bottom-up applicable. Due to the shape of the input $\mathcal{D} \vdash \exists \boldsymbol{x}q(\boldsymbol{x})$, only (id), $(\wedge_R)$, $(\exists_R)$, and $s(\rho)$ rules are applicable during proof search. Moreover, we let $\prec$ be an arbitrary cyclic order over $\mathcal{R} = \{\rho_1, \ldots, \rho_n\}$, that is, $\rho_1 \prec \rho_2 \cdots \rho_{n-1} \prec \rho_n \prec \rho_1$. We use $\prec$ to ensure the *fair application* of $s(\rho)$ rules during proof-search, meaning that no bottom-up rule application is delayed indefinitely.

Algorithm: `Prove`
Input: A sequent $\Gamma \vdash \Delta$.
Output: A Boolean `True` and `False`.

If $\Gamma \vdash \Delta$ is saturated, **Return** `False`;
If there exists a $p(\boldsymbol{t}) \in \Gamma \cap \Delta$, **Return** `True`;
If $\varphi \wedge \psi \in \Delta$, but $\varphi, \psi \notin \Delta$,
  Set $\Delta_1 := \varphi, \Delta$ and $\Delta_2 := \psi, \Delta$;
  If `Prove`$(\Gamma \vdash \Delta_i) =$ `False` for some $i \in \{1, 2\}$,
    Return `False`;
  Else
    Return `True`;
If $\exists x\varphi \in \Delta$ and $t \in \mathbf{T}(\Gamma)$, but $\varphi(t/x) \notin \Delta$,
  Set $\Delta := \varphi(t/x), \Delta$;
  Return `Prove`$(\Gamma \vdash \Delta)$
Let $\rho = \forall \boldsymbol{x}\boldsymbol{y}\ \beta(\boldsymbol{x}, \boldsymbol{y}) \rightarrow \exists \boldsymbol{z}\ \alpha(\boldsymbol{y}, \boldsymbol{z})$ be the next rule according to $\prec$ (if no rule has yet been picked, choose one in $\mathcal{R}$);
If a Γ-assignment μ exists such that $\mu(\beta(\boldsymbol{x}, \boldsymbol{y})) \subseteq \Gamma$, but no terms $\boldsymbol{t} \in \mathbf{T}(\Gamma)$ exist such that $\mu[\boldsymbol{t}/\boldsymbol{z}](\alpha(\boldsymbol{y}, \boldsymbol{z})) \subseteq \Gamma$;
  Set $\Gamma := \alpha(\mu(\boldsymbol{y}), \boldsymbol{z}), \Gamma$ with $\boldsymbol{z}$ fresh;
  Return `Prove`$(\Gamma \vdash \Delta)$;

Figure 2: The proof-search algorithm `Prove`.

Definition 2 (Saturation)**.** Let $\Gamma \vdash \Delta$ be a sequent. We say that $\Gamma \vdash \Delta$ is *saturated iff* it satisfies the following:

id**.** if $p(\boldsymbol{t}) \in \Gamma$, then $p(\boldsymbol{t}) \notin \Delta$;

$\wedge_R$**.** If $\varphi \wedge \psi \in \Delta$, then either $\varphi \in \Delta$ or $\psi \in \Delta$;

$\exists_R$**.** If $\exists x\varphi \in \Delta$, then for every $t \in \mathbf{T}(\Gamma)$, $\varphi(t/x) \in \Delta$;

er**.** For each $\rho \in \mathcal{R}$, if a Γ-assignment μ exists such that $\mu(\text{body}(\rho)) \subseteq \Gamma$, then there exist $\boldsymbol{t} \in \mathbf{T}(\Gamma)$ such that $\mu[\boldsymbol{t}/\boldsymbol{z}](\text{head}(\rho)) \subseteq \Gamma$ holds.

Theorem 3. *Let $\mathcal{R}$ be a rule set, $\mathcal{D}$ be a database, and $\exists \boldsymbol{x}q(\boldsymbol{x})$ be a BCQ. Then,*

1. *If* `Prove`$(\mathcal{D} \vdash \exists \boldsymbol{x}q(\boldsymbol{x})) =$ `True`*, then a proof in $\mathsf{G3}(\mathcal{R})$ can be constructed witnessing that $(\mathcal{D}, \mathcal{R}) \models \exists \boldsymbol{x}q(\boldsymbol{x})$;*
2. *If* `Prove`$(\mathcal{D} \vdash \exists \boldsymbol{x}q(\boldsymbol{x})) \neq$ `True`*, then a universal model can be constructed witnessing that $(\mathcal{D}, \mathcal{R}) \not\models \exists \boldsymbol{x}q(\boldsymbol{x})$.*

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{\Gamma \vdash \exists x(\mathtt{A}(x,a) \wedge \mathtt{F}(x)), \mathtt{A}(c,a)}\,(id) \qquad \dfrac{}{\Gamma \vdash \exists x(\mathtt{A}(x,a) \wedge \mathtt{F}(x)), \mathtt{F}(c)}\,(id)}{\Gamma \vdash \exists x(\mathtt{A}(x,a) \wedge \mathtt{F}(x)), \mathtt{A}(c,a) \wedge \mathtt{F}(c)}\,(\wedge_R)}{\mathtt{M}(b,a), \mathtt{A}(b,a), \mathtt{F}(b), \mathtt{M}(c,b), \mathtt{A}(c,b), \mathtt{F}(c), \mathtt{A}(c,a) \vdash \exists x(\mathtt{A}(x,a) \wedge \mathtt{F}(x))}\,(\exists_R)}{\mathtt{M}(b,a), \mathtt{A}(b,a), \mathtt{F}(b), \mathtt{M}(c,b), \mathtt{A}(c,b), \mathtt{F}(c) \vdash \exists x(\mathtt{A}(x,a) \wedge \mathtt{F}(x))}\,s(\rho_2)}{\mathtt{M}(b,a), \mathtt{A}(b,a), \mathtt{F}(b), \mathtt{M}(c,b) \vdash \exists x(\mathtt{A}(x,a) \wedge \mathtt{F}(x))}\,s(\rho_1)}{\mathtt{M}(b,a), \mathtt{M}(c,b) \vdash \exists x(\mathtt{A}(x,a) \wedge \mathtt{F}(x))}\,s(\rho_1)$$

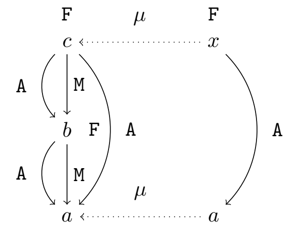

Figure 3: Above left is a proof in $\mathsf{G3}(\mathcal{R})$ witnessing that $\mathcal{K} \models \exists x(\mathtt{A}(x,a) \wedge \mathtt{F}(x))$, where $\mathcal{K} = (\mathcal{D}, \mathcal{R})$ is as defined in Example 11 and $\Gamma = \mathtt{M}(b,a), \mathtt{A}(b,a), \mathtt{F}(b), \mathtt{M}(c,b), \mathtt{A}(c,b), \mathtt{F}(c), \mathtt{A}(c,a)$. Above right is an illustration showing that the BCQ $\exists x(\mathtt{A}(x,a) \wedge \mathtt{F}(x))$ (to the right) can be mapped into the chase $\mathbf{Ch}_\infty(\mathcal{D}, \mathcal{R})$ (to the left) via the $\mathbf{Ch}_\infty(\mathcal{D}, \mathcal{R})$ -assignment μ (dotted arrows).

We refer to the universal model of $(\mathcal{D}, \mathcal{R})$ stated in the second claim of Theorem 3 as the *witnessing counter-model*.

4 Simulations and Equivalences

We present a sequence of results which culminate in the establishment of two main theorems: (1) Theorem 9, which confirms that chase derivations are mutually transformable with certain proofs in $\mathsf{G3}(\mathcal{R})$, and (2) Theorem 10, which confirms the equivalence of `Prove` and the chase. We end the section by providing an example illustrating the latter correspondence between proofs and the chase.

Observation 4. *Let $\mathcal{R}$ be a rule set. If $\rho \in \mathcal{R}$, then any application of $(\wedge_R)$ and $(\exists_R)$ permute above $s(\rho)$.*

Proof. It is straightforward to confirm the permutation of such rules as the $s(\rho)$ rules operate on the antecedent of a sequent, and $(\wedge_R)$ and $(\exists_R)$ operate on the consequent. □

Observation 5. *If $\mathcal{I}$ is an instance, then only $s(\rho)$ rules of $\mathsf{G3}(\mathcal{R})$ can be bottom-up applied to $\mathcal{I} \vdash \emptyset$. Moreover, such an application yields a sequent $\mathcal{I}' \vdash \emptyset$ with $\mathcal{I}'$ an instance.*

Observation 6. *The inference shown below left is a correct application of $s(\rho)$ iff the inference shown below right is:*

$$\frac{\Gamma' \vdash \emptyset}{\Gamma \vdash \emptyset}\, s(\rho) \qquad \frac{\Gamma' \vdash \Delta}{\Gamma \vdash \Delta}\, s(\rho)$$

Observation 7. *Let $\mathcal{I}$ and $\mathcal{I}'$ be instances with $\tau = (\rho, \mu)$ a trigger on $\mathcal{I}$. Then, $(\mathcal{I}, \tau), (\mathcal{I}', \emptyset)$ is a chase derivation iff the following is a correct application of $s(\rho)$:*

$$\frac{\mathcal{I}' \vdash \emptyset}{\mathcal{I} \vdash \emptyset}\, s(\rho)$$

Lemma 8. *For every rule set $\mathcal{R}$, $n \in \mathbb{N}$, and instances $\mathcal{I}_1, \ldots, \mathcal{I}_n$ there exists a chase derivation $(\mathcal{I}_i, \tau_i)_{i \in [n-1]}$ iff there exists an $\mathcal{R}$-derivation of $\mathcal{I}_1 \vdash \emptyset$ from $\mathcal{I}_n \vdash \emptyset$.*

In the proof of the following theorem, one shows that every chase derivation can be transformed into a proof in $\mathsf{G3}(\mathcal{R})$ and vice-versa, showing how existential rule reasoning and proofs in $\mathsf{G3}(\mathcal{R})$ simulate one another.

Theorem 9. *Let $\mathcal{R}$ be a rule set. A chase derivation $(\mathcal{I}_i, \tau_i)_{i \in [n]}$ witnessing $(\mathcal{D}, \mathcal{R}) \models \exists \boldsymbol{x} q(\boldsymbol{x})$ exists iff a proof in $\mathsf{G3}(\mathcal{R})$ of $\mathcal{D} \vdash \exists \boldsymbol{x} q(\boldsymbol{x})$ exists.*

Leveraging Theorems 3 and 9, it is straightforward to prove the first claim of the theorem below. The second claim is immediate as $\mathcal{I}$ and $\mathbf{Ch}_\infty(\mathcal{D}, \mathcal{R})$ are universal models. We note that the following theorem expresses a correspondence between proof-search and the chase.

Theorem 10. *Let $\mathcal{R}$ be a rule set, $\mathcal{D}$ be a database, and $\exists \boldsymbol{x} q(\boldsymbol{x})$ be a BCQ. Then,*

1. $\mathtt{Prove}(\mathcal{D} \vdash \exists \boldsymbol{x} q(\boldsymbol{x})) = \mathtt{True}$ *iff there is an $n \in \mathbb{N}$ such that $\mathbf{Ch}_n(\mathcal{D}, \mathcal{R}) \models \exists \boldsymbol{x} q(\boldsymbol{x})$ iff $\mathbf{Ch}_\infty(\mathcal{D}, \mathcal{R}) \models \exists \boldsymbol{x} q(\boldsymbol{x})$;*
2. *If $\mathtt{Prove}(\mathcal{D} \vdash \exists \boldsymbol{x} q(\boldsymbol{x})) \neq \mathtt{True}$, then $\mathcal{I} \equiv \mathbf{Ch}_\infty(\mathcal{D}, \mathcal{R})$ with $\mathcal{I}$ the witnessing counter-model.*

Example 11. We provide an example demonstrating the relationship between a proof and the chase. We read $\mathtt{F}(x)$ as 'x is female', $\mathtt{M}(x,y)$ as 'x is the mother of y' and $\mathtt{A}(x,y)$ as 'x is the ancestor of y'. We let $\mathcal{K} = (\mathcal{D}, \mathcal{R})$ be a knowledge base such that $\mathcal{D} = \{\mathtt{M}(b,a), \mathtt{M}(c,b)\}$, $\mathcal{R} = \{\rho_1, \rho_2\}$, and

$\rho_1 = \forall xy(\mathtt{M}(x,y) \rightarrow \mathtt{A}(x,y) \wedge \mathtt{F}(x))$;
$\rho_2 = \forall xy(\mathtt{A}(x,y) \wedge \mathtt{A}(y,z) \rightarrow \mathtt{A}(x,z))$.

In Figure 3, $\mathcal{K} \models \exists x(\mathtt{A}(x,a) \wedge \mathtt{F}(x))$ is witnessed and verified by the proof shown left. The graph shown right demonstrates that the BCQ $\exists x(\mathtt{A}(x,a) \wedge \mathtt{F}(x))$ (to the right) can be mapped into the chase $\mathbf{Ch}_\infty(\mathcal{D}, \mathcal{R})$ (to the left) via a $\mathbf{Ch}_\infty(\mathcal{D}, \mathcal{R})$-assignment μ (depicted as dotted arrows). (NB. We have omitted the points $\{\top(c) \mid c \in \mathbf{C}\}$ in the picture of $\mathbf{Ch}_\infty(\mathcal{D}, \mathcal{R})$ for simplicity.)

5 Concluding Remarks

We have formally established an equivalence between existential rule reasoning and sequent calculus proofs, effectively showing that proof-search simulates the chase. This work is meaningful as it uncovers and connects two central reasoning tasks and tools in the domain of existential rules and proof theory. Moreover, we have found that the counter-models extracted from failed proof-search are universal, implying their homomorphic equivalence to the chase—a previously unrecognized observation.

For future work, we aim to examine the relationship between the *disjunctive chase* (Bourhis et al. 2016) and proof-search in sequent calculi with disjunctive inference rules. It may additionally be worthwhile to investigate if our sequent systems can be adapted to facilitate reasoning with non-classical variants or extensions of existential rules. For example, we could merge our sequent calculi with those of (Lyon and Gómez Álvarez 2022) for *standpoint logic*—a modal logic used in knowledge integration to reason with diverse and potentially conflicting knowledge sources (Gómez Álvarez and Rudolph 2021). Finally, as this paper presents a sequent calculus for querying with existential rules, we plan to further explore its utility; e.g. by identifying admissible rules or applying loop checking techniques to uncover new classes of existential rules with decidable query entailment.

Acknowledgments

Work supported by the European Research Council (ERC) Consolidator Grant 771779 (DeciGUT).

References

Abiteboul, S.; Hull, R.; and Vianu, V. 1995. *Foundations of Databases*. Addison-Wesley.

Baget, J.-F.; Leclère, M.; Mugnier, M.-L.; and Salvat, E. 2009. Extending decidable cases for rules with existential variables. In Boutilier, C., ed., *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI'09)*, 677–682. Morgan Kaufmann.

Baget, J.-F.; Leclère, M.; Mugnier, M.-L.; and Salvat, E. 2011. On rules with existential variables: Walking the decidability line. *Artificial Intelligence* 175(9):1620–1654.

Beeri, C., and Vardi, M. Y. 1984. A proof procedure for data dependencies. *Journal of the ACM* 31(4):718–741.

Bourhis, P.; Manna, M.; Morak, M.; and Pieris, A. 2016. Guarded-based disjunctive tuple-generating dependencies. *ACM Trans. Database Syst.* 41(4).

Deutsch, A.; Nash, A.; and Remmel, J. B. 2008. The chase revisited. In Lenzerini, M., and Lembo, D., eds., *Proceedings of the 27th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS'08)*, 149–158. ACM.

Dyckhoff, R. 1992. Contraction-free sequent calculi for intuitionistic logic. *The Journal of Symbolic Logic* 57(3):795–807.

Fagin, R.; Kolaitis, P. G.; Miller, R. J.; and Popa, L. 2005. Data exchange: semantics and query answering. *Theoretical Computer Science* 336(1):89–124. Database Theory.

Feller, T.; Lyon, T. S.; Ostropolski-Nalewaja, P.; and Rudolph, S. 2023. Finite-cliquewidth sets of existential rules: Toward a general criterion for decidable yet highly expressive querying. In *Proceedings of the 26th International Conference on Database Theory (ICDT 2023)*, LIPIcs. Schloss Dagstuhl - Leibniz-Zentrum für Informatik. To appear. Preprint available via https://arxiv.org/abs/2209.02464.

Gentzen, G. 1935a. Untersuchungen über das logische schließen. i. *Mathematische zeitschrift* 39(1):176–210.

Gentzen, G. 1935b. Untersuchungen über das logische schließen. ii. *Mathematische Zeitschrift* 39(1):405–431.

Gómez Álvarez, L., and Rudolph, S. 2021. Standpoint logic: Multi-perspective knowledge representation. In Neuhaus, F., and Brodaric, B., eds., *Proceedings of the 12th International Conference on Formal Ontology in Information Systems*, volume 344 of *FAIA*, 3–17. IOS Press.

Gottlob, G. 2009. Datalog+/-: A unified approach to ontologies and integrity constraints. In Antonellis, V. D.; Castano, S.; Catania, B.; and Guerrini, G., eds., *Proceedings of the 17th Italian Symposium on Advanced Database Systems, (SEBD'09)*, 5–6. Edizioni Seneca.

Kleene, S. C. 1952. *Introduction to Metamathematics*. American Elsevier Publishing Company, INC. - New York.

Kripke, S. A. 1959. A completeness theorem in modal logic. *Journal of Symbolic Logic* 24(1):1–14.

Lyon, T. S., and Gómez Álvarez, L. 2022. Automating Reasoning with Standpoint Logic via Nested Sequents. In *Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning*, 257–266.

Lyon, T., and Karge, J. 2022. Uniform and modular sequent systems for description logics. In *Proceedings of the 35th International Workshop on Description Logics (DL 2022) co-located with Federated Logic Conference (FLoC 2022), Haifa, Israel, August 7th to 10th, 2022*.

Lyon, T.; Tiu, A.; Goré, R.; and Clouston, R. 2020. Syntactic interpolation for tense logics and bi-intuitionistic logic via nested sequents. In Fernández, M., and Muscholl, A., eds., *28th EACSL Annual Conference on Computer Science Logic, CSL 2020, January 13-16, 2020, Barcelona, Spain*, volume 152 of *LIPIcs*, 28:1–28:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

Maehara, S. 1960. On the interpolation theorem of craig. *Sûgaku* 12(4):235–237.

Salvat, E., and Mugnier, M.-L. 1996. Sound and complete forward and backward chainings of graph rules. In Eklund, P. W.; Ellis, G.; and Mann, G., eds., *Proceedings of the 4th International Conference on Conceptual Structures (ICCS'96)*, volume 1115 of *LNCS*, 248–262. Springer.

Simpson, A. K. 1994. *The proof theory and semantics of intuitionistic modal logic*. Ph.D. Dissertation, University of Edinburgh. College of Science and Engineering. School of Informatics.

Slaney, J. K. 1997. Minlog: A minimal logic theorem prover. In McCune, W., ed., *Automated Deduction - CADE-14, 14th International Conference on Automated Deduction, Townsville, North Queensland, Australia, July 13-17, 1997, Proceedings*, volume 1249 of *Lecture Notes in Computer Science*, 268–271. Springer.

A Proofs for Section 3

Lemma 12 (Soundness)**.** *If there exists a proof of* $\Gamma \vdash \Delta$ *in* $\mathsf{G3}(\mathcal{R})$*, then* $f(\Gamma \vdash \Delta)$ *is* $\mathcal{R}$*-valid.*

Proof. We prove the result by induction on the number of inferences in the proof of $\Gamma \vdash \Delta$.

Base case. If $\Gamma \vdash \Delta$ is an instance of (id), then $p(\boldsymbol{t}) \in \Gamma \cap \Delta$. Let $\Gamma = \Gamma', p(\boldsymbol{t})$ and $\Delta = \Delta', p(\boldsymbol{t})$. Suppose $\mathcal{I}$ is an interpretation and μ is an $\mathcal{I}$-assignment such that $\mathcal{I}, \mu \models \bigwedge \Gamma' \land p(\boldsymbol{t})$. Then, $\mu(p(\boldsymbol{t})) \in \mathcal{I}$, which shows that $\mathcal{I}, \mu \models \bigvee \Delta' \lor p(\boldsymbol{t})$. Hence, $\mathcal{I} \models \bigwedge \Gamma \rightarrow \bigvee \Delta$ for any interpretation $\mathcal{I}$, including any interpretation $\mathcal{I}$ such that $\mathcal{I} \models \mathcal{R}$. This shows that $f(\Gamma \vdash \Delta)$ is $\mathcal{R}$-valid.

Inductive step. We consider the final inference in the given proof and show that if the conclusion is not $\mathcal{R}$-valid, then the premise is not $\mathcal{R}$-valid, establishing that the conclusion is $\mathcal{R}$-valid as the premise is $\mathcal{R}$-valid by the inductive hypothesis. We argue the $(\exists_L)$, $(\exists_R)$, and $s(\rho)$ cases as the remaining cases are straightforward.

$(\exists_L)$. Suppose an interpretation $\mathcal{I}$ and $\mathcal{I}$-assignment μ exist such that $\mathcal{I} \models \mathcal{R}$ and $\mathcal{I}, \mu \not\models f(\Gamma, \exists x\varphi \vdash \Delta)$. Then, there exists a term $t \in \mathbf{T}(\mathcal{I})$ such that $\mathcal{I}, \mu[t/x] \models \varphi$. It follows that $\mathcal{I}, \mu[t/y] \models \varphi(y/x)$, showing that $\mathcal{I}, \mu[t/y] \not\models f(\Gamma, \varphi(y/x) \vdash \Delta)$, which concludes the proof of the case.

$(\exists_R)$. Suppose an interpretation $\mathcal{I}$ and $\mathcal{I}$-assignment μ exist such that $\mathcal{I} \models \mathcal{R}$ and $\mathcal{I}, \mu \not\models f(\Gamma \vdash \exists x\varphi, \Delta)$. It follows that $\mathcal{I}, \mu \not\models \exists x\varphi$, meaning that for every term $t \in \mathbf{T}(\mathcal{I})$, $\mathcal{I}, \mu \not\models \varphi(t/x)$. As $\mathcal{I}$ is an interpretation, $\mathbf{T}(\mathcal{I}) = \mathbf{T}$, showing that $\mathcal{I}, \mu \not\models \varphi(t/x)$. This implies that $\mathcal{I}, \mu \not\models f(\Gamma \vdash \exists x\varphi, \varphi(t/x), \Delta)$, i.e. the premise is not $\mathcal{R}$-valid.

$s(\rho)$. Let $\rho = \forall \boldsymbol{xy}\beta(\boldsymbol{x}, \boldsymbol{y}) \rightarrow \exists \boldsymbol{z}\alpha(\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{R}$. Suppose an interpretation $\mathcal{I}$ and $\mathcal{I}$-assignment μ exist such that $\mathcal{I} \models \mathcal{R}$ and $\mathcal{I}, \mu \not\models f(\Gamma, \beta(\boldsymbol{x}, \boldsymbol{y}) \vdash \Delta)$. It follows that we have $\mathcal{I}, \mu \models \beta(\boldsymbol{x}, \boldsymbol{y})$, implying $\mathcal{I}, \mu \models \exists \boldsymbol{z}\alpha(\boldsymbol{y}, \boldsymbol{z})$ as $\mathcal{I} \models \mathcal{R}$. Therefore, there exist $\boldsymbol{t} \in \mathbf{T}(\mathcal{I})$ such that $\mathcal{I}, \mu[\boldsymbol{t}/\boldsymbol{z}] \models \alpha(\boldsymbol{y}, \boldsymbol{z})$, showing that $\mathcal{I}, \mu \not\models f(\Gamma, \beta(\boldsymbol{x}, \boldsymbol{y}), \alpha(\boldsymbol{y}, \boldsymbol{z}) \vdash \Delta)$, i.e. the premise is not $\mathcal{R}$-valid. □

Lemma 13 (Completeness)**.** *If $f(\Gamma \vdash \Delta)$ is $\mathcal{R}$-valid, there exists a proof of $\Gamma \vdash \Delta$ in $\mathsf{G3}(\mathcal{R})$.*

Proof. Similar to the proof-search procedure `Prove`, we take $\Gamma \vdash \Delta$ as input and apply rules from $\mathsf{G3}(\mathcal{R})$ bottom-up with the goal of constructing a proof thereof. We bottom-up apply the rules from $\mathsf{G3}(\mathcal{R})$ in a roundabout fashion, applying (id), $(\neg_L)$, $(\neg_R)$, $(\wedge_L)$, $(\wedge_R)$, $(\exists_L)$, $(\exists_R)$, $s(\rho)$, and then circling back to (id). When we consider a rule, we bottom-up apply it in all possible ways to the derivation being constructed. Furthermore, if a path in the derivation exists such that it is not an instance of (id) and none of the other rules are bottom-up applicable to the sequent at the top of the path, then we copy it above itself. This strategy of completeness is common in the proof theory literature; e.g. see (Kripke 1959; Lyon and Karge 2022).

Let us suppose that no proof of $\Gamma \vdash \Delta$ exists in $\mathsf{G3}(\mathcal{R})$. Then, the above described process will not terminate, implying that an infinite derivation of $\Gamma \vdash \Delta$ will be constructed, which has the shape of an infinite tree with finite branching. Therefore, by König's lemma an infinite path

$$\mathcal{P} = (\Gamma_0 \vdash \Delta_0), (\Gamma_1 \vdash \Delta_1), \ldots, (\Gamma_n \vdash \Delta_n), \ldots$$

of sequents exists such that $\Gamma_0 = \Gamma$ and $\Delta_0 = \Delta$. We define:

$$\mathbf{\Gamma} = \bigcup_{i \in \mathbb{N}} \Gamma_i \qquad \mathbf{\Delta} = \bigcup_{i \in \mathbb{N}} \Delta_i$$

Using $\mathbf{\Gamma}$, we define an interpretation $\mathcal{I}$ accordingly:

$$\mathcal{I} = \{p(\boldsymbol{t}) \mid p(\boldsymbol{t}) \in \mathbf{\Gamma}\}^{\top}$$

It is straightforward to argue that $\mathcal{I} \models \mathcal{R}$ and is similar to the argument given in Theorem 3 below. Let us now define an $\mathcal{I}$-assignment μ such that $\mu(t) = t$ for each term $t \in \mathbf{T}$. We now show by simultaneous induction on the number of logical connectives in φ that (1) if $\varphi \in \mathbf{\Gamma}$, then $\mathcal{I}, \mu \models \varphi$, and (2) if $\varphi \in \mathbf{\Delta}$, then $\mathcal{I}, \mu \not\models \varphi$. We only show the atomic, negation, and existential cases below for claim (1) as the remaining cases are simple or similar.

$p(\boldsymbol{t}) \in \mathbf{\Gamma}$**.** If $p(\boldsymbol{t}) \in \mathbf{\Gamma}$, then $p(\boldsymbol{t}) \in \mathcal{I}$ by definition, showing that $\mu(p(\boldsymbol{t})) \in \mathcal{I}$, which implies that $\mathcal{I}, \mu \models p(\boldsymbol{t})$.

$\neg\psi \in \mathbf{\Gamma}$**.** If $\neg\psi \in \mathbf{\Gamma}$, then $\psi \in \mathbf{\Delta}$ as the $(\neg_L)$ rule will eventually be bottom-up applied in the procedure described above, so by IH and claim (2), $\mathcal{I}, \mu \not\models \psi$, which shows that $\mathcal{I}, \mu \models \neg\psi$.

$\exists x\psi \in \mathbf{\Gamma}$**.** If $\exists x\psi \in \mathbf{\Gamma}$, then at some stage $(\exists_L)$ will be applied bottom-up ensuring that $\psi(y/x) \in \mathbf{\Gamma}$. By IH, $\mathcal{I}, \mu[y/y] \models \psi(y/x)$, and we know $y \in \mathbf{T}(\mathcal{I})$, showing that $\mathcal{I}, \mu \models \exists x\psi$.

Since $\Gamma \subseteq \mathbf{\Gamma}$ and $\Delta \subseteq \mathbf{\Delta}$, we have that $\mathcal{I}, \mu \models \bigwedge \Gamma$ and $\mathcal{I}, \mu \not\models \bigvee \Delta$, implying that $\mathcal{I}, \mu \not\models \bigwedge \Gamma \rightarrow \bigvee \Delta$, which demonstrates that $f(\Gamma \vdash \Delta)$ is not $\mathcal{R}$-valid. □

Theorem 1 (Soundness and Completeness)**.** *$f(\Gamma \vdash \Delta)$ is $\mathcal{R}$-valid iff there exists a proof of $\Gamma \vdash \Delta$ in $\mathsf{G3}(\mathcal{R})$.*

Proof. Follows from Lemma 12 and Lemma 13 above. □

Theorem 3. *Let $\mathcal{R}$ be a rule set, $\mathcal{D}$ be a database, and $\exists \boldsymbol{x}q(\boldsymbol{x})$ be a BCQ. Then,*

1. *If $\texttt{Prove}(\mathcal{D} \vdash \exists \boldsymbol{x}q(\boldsymbol{x})) = \texttt{True}$, then a proof in $\mathsf{G3}(\mathcal{R})$ can be constructed witnessing that $(\mathcal{D}, \mathcal{R}) \models \exists \boldsymbol{x}q(\boldsymbol{x})$;*
2. *If $\texttt{Prove}(\mathcal{D} \vdash \exists \boldsymbol{x}q(\boldsymbol{x})) \neq \texttt{True}$, then a universal model can be constructed witnessing that $(\mathcal{D}, \mathcal{R}) \not\models \exists \boldsymbol{x}q(\boldsymbol{x})$.*

Proof. The first claim is immediate since if $\texttt{Prove}(\mathcal{D} \vdash \exists \boldsymbol{x}q(\boldsymbol{x})) = \texttt{True}$, then `Prove` constructs a proof of $\mathcal{D} \vdash \exists \boldsymbol{x}q(\boldsymbol{x})$ as every recursive call of `Prove` corresponds to a bottom-up application of $(\wedge_R)$, $(\exists_R)$, or $s(\rho)$. This implies that $(\mathcal{D}, \mathcal{R}) \models \exists \boldsymbol{x}q(\boldsymbol{x})$ as $\mathsf{G3}(\mathcal{R})$ is sound. Let us therefore argue that the second claim holds.

We assume w.l.o.g. that `Prove` does not terminate and show how to extract a counter-model witnessing that $(\mathcal{D}, \mathcal{R}) \not\models \exists \boldsymbol{x}q(\boldsymbol{x})$. Since `Prove` does not terminate, it generates an infinite derivation in the form of a tree with $\mathcal{D} \vdash \exists \boldsymbol{x}q(\boldsymbol{x})$ root and which is finite branching (as $\mathsf{G3}(\mathcal{R})$ only consists of unary and binary rules). Hence, by König's lemma there exists an infinite path

$$\mathcal{P} = (\Gamma_0 \vdash \Delta_0), (\Gamma_1 \vdash \Delta_1), \ldots, (\Gamma_n \vdash \Delta_n), \ldots$$

of sequents in the infinite derivation such that $\Gamma_0 = \mathcal{D}$ and $\Delta_0 = \exists \boldsymbol{x}q(\boldsymbol{x})$. We use this path to construct an interpretation $\mathcal{I}$ such that $\mathcal{I} \models \mathcal{D}$ and $\mathcal{I} \not\models \exists \boldsymbol{x}q(\boldsymbol{x})$. Let us now define:

$$\mathcal{I} = (\bigcup_{i \in \mathbb{N}} \Gamma_i)^{\top} \quad \Delta = \bigcup_{i \in \mathbb{N}} \Delta_i$$

We now argue (1) $\mathcal{I} \models (\mathcal{D}, \mathcal{R})$, and (2) if $\varphi \in \Delta$, then $\mathcal{I} \not\models \varphi$. We argue claim (1) first: Since $\mathcal{D} = \Gamma_0 \subseteq \mathcal{I}$, we know that $\mathcal{I}, \mu \models \mathcal{D}$ for any $\mathcal{I}$-assignment μ as all $\mathcal{I}$-assignments map constants in the same way and $\mathcal{D}$ contains only constants; hence, $\mathcal{I} \models \mathcal{D}$. Let us now argue that $\mathcal{I} \models \mathcal{R}$ as well. Let μ be an arbitrary $\mathcal{I}$-assignment and $\rho \in \mathcal{R}$ with $\rho = \forall \boldsymbol{xy}\ \beta(\boldsymbol{x}, \boldsymbol{y}) \rightarrow \exists \boldsymbol{z}\ \alpha(\boldsymbol{y}, \boldsymbol{z})$. Suppose that $\mathcal{I}, \mu \models \beta(\boldsymbol{x}, \boldsymbol{y})$. It follows that for some $\Gamma_i \vdash \Delta_i$ in $\mathcal{P}$, $\mu(\beta(\boldsymbol{x}, \boldsymbol{y})) \subseteq \Gamma_i$. If $\Gamma_i, \mu \not\models \exists \boldsymbol{z}\ \alpha(\boldsymbol{y}, \boldsymbol{z})$, then due to the cyclic order $\prec$ (ensuring fairness) imposed during proof-search, eventually ρ will be considered and $s(\rho)$ applied bottom-up. This ensures that $\alpha(\mu(\boldsymbol{y}), \boldsymbol{z}) \in \mathcal{I}$ with $\boldsymbol{z}$ fresh in the application of $s(\rho)$. Hence, $\mathcal{I}, \mu \models \rho$, showing that $\mathcal{I} \models \rho$ as μ was assumed arbitrary, establishing claim (1).

Let us now argue (2) by induction on the number of logical operators in φ. Note that due to the shape of the input $\mathcal{D} \vdash \exists \boldsymbol{x} q(\boldsymbol{x})$ and the rules applied during proof-search, only atomic formulae, conjunctions, and existentials will occur in Δ. We define $\mu(t) = t$ for each $t \in \mathbf{T}$.

$p(\boldsymbol{t}) \in \Delta$. If $p(\boldsymbol{t}) \in \mathcal{I}$, then there exists some i such that $p(\boldsymbol{t}) \in \Gamma_i \cap \Delta_i$, meaning that `Prove` would terminate and return `True` contrary to our assumption. Thus, $p(\boldsymbol{t}) \notin \mathcal{I}$, implying that $\mathcal{I}, \mu \not\models p(\boldsymbol{t})$.

$\psi \wedge \chi \in \Delta$. If $\psi \wedge \chi \in \Delta$, then there exists a minimal i such that $\psi \wedge \chi \in \Delta_i$. Therefore, by the conjunction step of `Prove`, we know that at some stage $j \geq i$ either $\psi \in \Delta_j$ or $\chi \in \Delta_j$. By IH, either $\psi \in \Delta$ or $\chi \in \Delta$, meaning either $\mathcal{I}, \mu \not\models \psi$ or $\mathcal{I}, \mu \not\models \chi$, showing that $\mathcal{I}, \mu \not\models \psi \wedge \chi$ regardless of which case holds.

$\exists x \psi \in \Delta$. Let $t \in \mathbf{T}(\mathcal{I})$. Since $\exists x \psi \in \Delta$, there exists some minimal i, $\exists x \psi \in \Delta_i$. By the existential step of `Prove`, we know that $\psi(t/x) \in \Delta_j$ for some $j \geq i$. By IH, $\mathcal{I} \not\models \psi(t/x)$, and since t was chosen arbitrarily, we have that for all $t \in \mathbf{T}(\mathcal{I})$, $\mathcal{I}, \mu \not\models \psi(t/x)$, showing that $\mathcal{I}, \mu \not\models \exists x \psi$.

This concludes the proof of claim (2). As a consequence, since $\exists \boldsymbol{x} q(\boldsymbol{x}) \in \Delta$, it follows that $\mathcal{I}, \mu \not\models \exists \boldsymbol{x} q(\boldsymbol{x})$. This fact, in conjunction with claim (1), establishes that $(\mathcal{D}, \mathcal{R}) \not\models \exists \boldsymbol{x} q(\boldsymbol{x})$. Last, we argue that $\mathcal{I}$ is a universal model for $(\mathcal{D}, \mathcal{R})$. Let $\mathcal{J}$ be any model of $(\mathcal{D}, \mathcal{R})$. We first define a sequence $s' = \langle \Gamma'_i \rangle_{i \in \mathbb{N}}$ relative to the sequence $s = \langle \Gamma_i \rangle_{i \in \mathbb{N}}$ of antecedents from the path $\mathcal{P}$ accordingly: (1) $\Gamma'_0 = \Gamma_0 = \mathcal{D}$, (2) $\Gamma'_{n+1} = \min\{\Gamma_k \mid \Gamma_k \neq \Gamma'_n, n \leq k\}$.

We inductively build an increasing sequence of homomorphisms $\langle h_i : \Gamma'_i \to \mathcal{J} \rangle_{i \in \mathbb{N}}$. By the definition of s', we have that $\Gamma'_0 = \Gamma_0 = \mathcal{D}$. As $\mathcal{D} \subseteq \mathcal{J}$, we simply define h_0 to be an identity on the domain of $\mathcal{D}$. Now, assume that h_i is a homomorphism from $\Gamma'_i \to \mathcal{J}$. We will build a homomorphism $h_{i+1} : \Gamma'_{i+1} \to \mathcal{J}$. Observe, Γ'_{i+1} can only be obtained from Γ'_i by an application of $s(\rho)$ with $\rho = \forall \boldsymbol{x}\boldsymbol{y}\ \beta(\boldsymbol{x}, \boldsymbol{y}) \to \exists \boldsymbol{z}\ \alpha(\boldsymbol{y}, \boldsymbol{z})$. Thus, from Observation 7, a trigger $\tau = (\rho, \mu)$ exists in Γ'_i. Let $\mu(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{t}, \boldsymbol{t}'$ where $\boldsymbol{x}, \boldsymbol{y}$ is the tuple of universally quantified variables in $\mathrm{body}(\rho)$. Note that there exists a trigger (ρ, μ') in $\mathcal{J}$ such that $\mu'(\boldsymbol{x}, \boldsymbol{y}) = h_i(\boldsymbol{t}, \boldsymbol{t}')$. However, as $\mathcal{J}$ is a model of $(\mathcal{D}, \mathcal{R})$ we know that there exists a $\mathcal{J}$-assignment μ'' mapping $\mathrm{head}(\rho)$ into $\mathcal{J}$ that agrees on $\boldsymbol{y}$ with μ'. As Γ'_{i+1} is created by applying the trigger τ to Γ'_i, we know that there exists a Γ'_i-assignment μ''' that maps $\mathrm{head}(\rho)$ to Γ'_{i+1} which agrees with μ on $\boldsymbol{y}$. From those facts we can see that h_{i+1} can be constructed from h_i by ensuring that $h_{i+1} \circ \mu'''$ is identical to μ'' when restricted to $\boldsymbol{z}$. Note that $h_{i+1} \circ \mu'''$ already agrees on $\boldsymbol{x}$ and $\boldsymbol{y}$ with μ''. Finally, as the homomorphisms h_i form an ascending sequence, we take the limit $\bigcup_{i \in \mathbb{N}} h_i$ as the homomorphism from $\mathcal{I}$ to $\mathcal{J}$. ☐

B Proofs for Section 4

Lemma 8. *For every rule set $\mathcal{R}$, $n \in \mathbb{N}$, and instances $\mathcal{I}_1, \ldots, \mathcal{I}_n$ there exists a chase derivation $(\mathcal{I}_i, \tau_i)_{i \in [n-1]}$ iff there exists an $\mathcal{R}$-derivation of $\mathcal{I}_1 \vdash \emptyset$ from $\mathcal{I}_n \vdash \emptyset$.*

Proof. ($\Rightarrow$) Follows from Observation 7. ($\Leftarrow$) Let Π be a bottom-up $\mathcal{R}$-derivation of $\mathcal{I}_1 \vdash \emptyset$ from $\mathcal{I}_n \vdash \emptyset$. By the definition of an $\mathcal{R}$-derivation, only $s(\rho)$ rules are applied in Π. We may transform Π into the chase derivation $(\mathcal{I}_i, \tau_i)_{i \in [n-1]}$ by Observation 7. ☐

Theorem 9. *Let $\mathcal{R}$ be a rule set. A chase derivation $(\mathcal{I}_i, \tau_i)_{i \in [n]}$ witnessing $(\mathcal{D}, \mathcal{R}) \models \exists \boldsymbol{x} q(\boldsymbol{x})$ exists iff a proof in $\mathsf{G3}(\mathcal{R})$ of $\mathcal{D} \vdash \exists \boldsymbol{x} q(\boldsymbol{x})$ exists.*

Proof. ($\Rightarrow$) By assumption, an $\mathcal{I}_{n+1}$-assignment μ exists such that $\mathcal{I}_{n+1}, \mu \models q(\boldsymbol{x})$. It follows that $\mathcal{I}_{n+1} \models \exists \boldsymbol{x} q(\boldsymbol{x})$, which implies that a proof Π of $\mathcal{I}_{n+1} \vdash \exists \boldsymbol{x} q(\boldsymbol{x})$ can be constructed in $\mathsf{G3}(\mathcal{R})$ by completeness (Theorem 1). By our assumption and Lemma 8, we obtain a derivation Π' of $\mathcal{D} \vdash \emptyset$ from $\mathcal{I}_{n+1} \vdash \emptyset$. Moreover, by Observation 5 we know that Π' uses only $s(\rho)$ rules, and by repeated application of Observation 6, we may transform Π' into a new derivation Π'' of $\mathcal{D} \vdash \exists \boldsymbol{x} q(\boldsymbol{x})$ from $\mathcal{I}_{n+1} \vdash \exists \boldsymbol{x} q(\boldsymbol{x})$. By affixing Π atop Π'' we obtain a proof in $\mathsf{G3}(\mathcal{R})$ of $\mathcal{D} \vdash \exists \boldsymbol{x} q(\boldsymbol{x})$.

($\Leftarrow$) Suppose Π is a proof of $\mathcal{D} \vdash \exists \boldsymbol{x} q(\boldsymbol{x})$ in $\mathsf{G3}(\mathcal{R})$. Observe that Π can only use $s(\rho)$, $(\wedge_R)$, and $(\exists_R)$ rules as $\mathcal{D}$ is an instance and $\exists \boldsymbol{x} q(\boldsymbol{x})$ is a BCQ. By Observation 4, we may transform Π into a new proof which consists of two fragments: a top proof Π' of $\mathcal{I} \vdash \exists \boldsymbol{x} q(\boldsymbol{x})$ consisting only of $(\wedge_R)$ and $(\exists_R)$ applications, and a bottom derivation Π'' of $\mathcal{D} \vdash \exists \boldsymbol{x} q(\boldsymbol{x})$ from $\mathcal{I} \vdash \exists \boldsymbol{x} q(\boldsymbol{x})$ consisting only of $s(\rho)$ applications. By the soundness of $\mathsf{G3}(\mathcal{R})$ (Theorem 1), there exists an $\mathcal{I}$-assignment μ such that $\mu(q(\boldsymbol{x})) \subseteq \mathcal{I}$. By Observation 6 and Lemma 8, we can transform Π'' into a chase derivation $(\mathcal{I}_i, \tau_i)_{i \in [n]}$ with $\mathcal{I}_{n+1} = \mathcal{I}$. As $\mu(q(\boldsymbol{x})) \subseteq \mathcal{I}$, this chase derivation witnesses $(\mathcal{D}, \mathcal{R}) \models \exists \boldsymbol{x} q(\boldsymbol{x})$. ☐

Theorem 10. *Let $\mathcal{R}$ be a rule set, $\mathcal{D}$ be a database, and $\exists \boldsymbol{x} q(\boldsymbol{x})$ be a BCQ. Then,*

1. `Prove`$(\mathcal{D} \vdash \exists \boldsymbol{x} q(\boldsymbol{x})) =$ `True` *iff there is an $n \in \mathbb{N}$ such that $\mathbf{Ch}_n(\mathcal{D}, \mathcal{R}) \models \exists \boldsymbol{x} q(\boldsymbol{x})$ iff $\mathbf{Ch}_\infty(\mathcal{D}, \mathcal{R}) \models \exists \boldsymbol{x} q(\boldsymbol{x})$;*
2. *If* `Prove`$(\mathcal{D} \vdash \exists \boldsymbol{x} q(\boldsymbol{x})) \neq$ `True` *, then $\mathcal{I} \equiv \mathbf{Ch}_\infty(\mathcal{D}, \mathcal{R})$ with $\mathcal{I}$ the witnessing counter-model.*

Proof. We note that claim (2) is straightforward as both $\mathcal{I}$ and $\mathbf{Ch}_\infty(\mathcal{D}, \mathcal{R})$ are universal models of $(\mathcal{D}, \mathcal{R})$. Regarding (1), we argue the first equivalence as the second equivalence is trivial.

($\Rightarrow$) By assumption, a proof Π of $\mathcal{D} \vdash \exists \boldsymbol{x} q(\boldsymbol{x})$ exists in $\mathsf{G3}(\mathcal{R})$. By Theorem 9, a chase derivation $(\mathcal{I}_i, \tau_i)_{i \in [n]}$ witnessing $(\mathcal{D}, \mathcal{R}) \models \exists \boldsymbol{x} q(\boldsymbol{x})$ can be constructed from Π. At some step $n \in \mathbb{N}$ of the chase an instance $\mathcal{I}$ will be generated such that $\mathcal{I}_{n+1} \subseteq \mathcal{I}$, showing that $\mathbf{Ch}_n(\mathcal{D}, \mathcal{R}) \models \exists \boldsymbol{x} q(\boldsymbol{x})$.

($\Leftarrow$) If $\mathbf{Ch}_n(\mathcal{D}, \mathcal{R}) \models \exists \boldsymbol{x} q(\boldsymbol{x})$, then $\mathbf{Ch}_n(\mathcal{D}, \mathcal{R})$ can be linearized into a chase derivation $(\mathcal{I}_i, \tau_i)_{i \in [n]}$ witnessing $(\mathcal{D}, \mathcal{R}) \models \exists \boldsymbol{x} q(\boldsymbol{x})$, from which a proof Π of $\mathcal{D} \vdash \exists \boldsymbol{x} q(\boldsymbol{x})$ can be constructed in $\mathsf{G3}(\mathcal{R})$ by Theorem 9. By Theorem 3, $\texttt{Prove}(\mathcal{D} \vdash \exists \boldsymbol{x} q(\boldsymbol{x})) = \texttt{True}$. $\square$