Appendix

Shy existential rules

This section is devoted to recall the formal definition of shy ontologies and their syntactic properties, as defined in Leone et al. (2012). For notational convenience and without loss of generality, we assume here that each pair of rules of an ontology share no variable. Let Σ be an ontology, α be a m-arity atom, $i \in \{1,\ldots,m\}$ be an index, $pred(\alpha) = a$, and X be an existential variable occurring in some rule of Σ. We say that position $a[i]$ is *invaded* by X if there exists a rule $\rho \in \Sigma$ such that $head(\rho) = \alpha$ and

(i) $\alpha[i] = X$; or
(ii) $\alpha[i]$ is a universal variable of ρ and all of its occurrences in $body(\rho)$ appear in positions invaded by X.

Let $\phi(\mathbf{X})$ be a conjunction of atoms, and let $X \in \mathbf{X}$. We say that X is *attacked* by a variable Y in $\phi(\mathbf{X})$ if all the positions where X appears are invaded by Y. On the other hand, we say that X is *protected* in $\phi(\mathbf{X})$, if it is attacked by no variable.

A rule ρ of an ontology Σ is called *shy* w.r.t. Σ if the following conditions are both satisfied:

(i) if a variable X occurs in more than one body atom, then X is protected in $body(\rho)$;
(ii) if two distinct variables are not protected in $body(\rho)$ but occur both in $head(\rho)$ and in two different body atoms, then they are not attacked by the same variable.

Finally, if each $\rho \in \Sigma$ is shy w.r.t. Σ, then call Σ a shy ontology.

Example 6.1
Consider the following rules

$$\begin{array}{rrcl} \rho_1 = & s(X_1) & \rightarrow & \exists Y_1 p(X_1,Y_1); \\ \rho_2 = & p(X_2,Y_2), u(Y_2) & \rightarrow & r(X_2,Y_2); \\ \rho_3 = & t(X_3) & \rightarrow & \exists Y_3 u(Y_3). \end{array}$$

Let $\Sigma = \{\rho_1,\rho_2,\rho_3\}$. Clearly, ρ_1 and ρ_3 are shy rules w.r.t. Σ, since they are also linear rules, namely rules with one single body atom, which cannot violate any of the two shy conditions. Moreover, rule ρ_2 is also shy w.r.t. Σ as the positions $p[2]$ and $u[1]$ are invaded by disjoint sets of existential variables. Indeed, $p[2]$ is invaded by the existential variable Y_1 of the first rule, and $u[1]$ is invaded by the existential variable Y_3 of the third rule. Therefore, Σ is a shy ontology.

Now, consider the further three existential rules

$$\begin{array}{rrcl} \rho_4 = & u(X_4) & \rightarrow & \exists Y_4 p(Y_4,X_4); \\ \rho_5 = & u(X_5) & \rightarrow & \exists Y_5 p(X_5,Y_5); \\ \rho_6 = & r(X_6,X_6) & \rightarrow & v(X_6). \end{array}$$

Let Σ' be the ontology $\Sigma \cup \{\rho_4\}$. It is easy to see that ρ_1, ρ_3 and ρ_4 are shy w.r.t. Σ'. However, ρ_2 is not shy w.r.t. Σ', as property (i) is not satisfied. Indeed, the variable Y_2 occurring in two body atoms in $body(\rho_2)$ is not protected, as the position $p[2]$ and $u[1]$ (the only positions in which Y_2 occurs) are invaded by the same existential variable, namely Y_3. Therefore, Σ' is not a shy ontology.

Let Σ'' be the ontology $\Sigma \cup \{\rho_5,\rho_6\}$. Again, ρ_1, ρ_3, ρ_5 and ρ_6 are trivially shy w.r.t. Σ''; and again ρ_2 is not shy w.r.t. Σ''. However, this time, ρ_2 is not shy because property (ii) is not satisfied. Indeed, the universal variables X_2 and Y_2, occurring in two different body atoms and in

$head(\rho_2)$, are not protected in $body(\rho_2)$, as the position $p[1]$ and $u[1]$ (in which occur X_2 and Y_2, respectively) are attacked by the same variable Y_3. Therefore, Σ'' is not a shy ontology. $\triangleleft$

Essentially, during every possible chase step, condition (i) guarantees that each variable that occurs in more than one body atom is always mapped into a constant. Although this is the key property behind shy, we now explain the role played by condition (ii) and its importance. To this aim, we exploit again Σ'', as introduced in the previous example, and we reveal why this second condition, in a sense, turns into the first one. Indeed, the rule ρ_6 bypasses the propagation of the same null in ρ_2 via different variables. However, one can observe that the rules ρ_2 and ρ_6 imply the rule $\rho_6' : p(X_6,Y_6), u(X_6) \rightarrow v(X_6)$, which of course does not satisfy condition (i). Actually, it is not difficult to see that every ontology can be rewritten (independently from D and q) into an en equivalent one (w.r.t. query answering) where all the rules satisfy condition (i). As an example, consider the following rule ρ

$$p(X_1,Y_1), r(Y_1,Z_1), u(Z_1,Y_1) \quad \rightarrow \quad \exists W_1\, t(X_1,Z_1,W_1),$$

and assume that it belongs to some ontology Σ and that it is not shy w.r.t. Σ because it violates condition (i) only. Let us now construct Σ' as $\Sigma \setminus \{\rho\}$ plus the following two rules:

$$\begin{array}{rcl} p(X_1,Y_1), r(Y_1',Z_1), u(Z_1',Y_1'') & \rightarrow & aux_\rho(X_1,Y_1,Y_1',Z_1,Z_1',Y_1''); \\ aux_\rho(X_1,Y_1,Y_1,Z_1,Z_1,Y_1) & \rightarrow & \exists W_1\, t(X_1,Z_1,W_1). \end{array}$$

Both the new rules satisfy now condition (i) w.r.t. Σ'. Moreover, it is not difficult to see that, for every database D and for every UBCQ q, it holds that $D \cup \Sigma \models q$ if and only if $D \cup \Sigma' \models q$. However, since ρ does not satisfy condition (i), this immediately implies that the first new rule does not satisfy condition (ii).

The syntactic properties of shy make the class quite expressive since it strictly contains both linear and datalog. Moreover, these properties are easy recognizable and guarantee efficient answering to conjunctive queries, as experimentally shown in Leone et al. (2012). In fact, ontology-based query answering over shy ontologies preservers the same data and combined complexity of OBQA over datalog, namely PTIME-complete and EXPTIME-complete, respectively.

Formal Proofs

Proof of Proposition 3.2

We prove that $\mathfrak{R}(chase(D^c,\Sigma^c)) = chase(D,\Sigma)$ by induction on the chase step. Let $I_0 = D \subset I_1 \subset \ldots \subset I_m \subset \ldots$ be a chase procedure of D and Σ; and let $I_0^c = D^c \subset I_1^c \subset \ldots \subset I_m^c \subset \ldots$ be a chase procedure of D^c and Σ^c.

Clearly, the base case follows, since, by definition of the canonical rewriting of D, $\mathfrak{R}(D^c) = D$.

Then, assume that $\mathfrak{R}(I_m^c) = I_m$. We have to prove that $\mathfrak{R}(I_{m+1}^c) = I_{m+1}$. By definition of chase step, there exist a rule $\rho \in \Sigma$ and a homomorphism h from $body(\rho)$ to I_m, such that $\langle \rho, h\rangle(I_m) = I_{m+1}$. That is, $I_{m+1} = I_m \cup \{h(head(\rho))\}$. By construction of a canonical rule, there exists a safe substitution ς w.r.t. ρ, such that $\varsigma(\rho)^c$ is a canonical rule and, by inductive hypothesis, there exists a homomorphism h^c from $body(\varsigma(\rho)^c)$ to I_m^c. Consider the following homomorphism $(h^c)' = (h \setminus h|_{\mathbf{X}}) \cup h^c|_{\mathbf{X}} \supseteq h^c|_{\mathbf{X}}$. Therefore, $I_{m+1}^c = I_m^c \cup \{(h^c)'(head(\langle \rho, \varsigma\rangle))\}$. Moreover,

$$\begin{array}{rll} \mathfrak{R}(I_{m+1}^c) & = \mathfrak{R}(I_m^c \cup \{(h^c)'(head(\varsigma(\rho)^c))\}) & = \\ & = \mathfrak{R}(I_m^c) \cup \mathfrak{R}(\{(h^c)'(head(\varsigma(\rho)^c))\}) & = \\ & = I_m \cup \{h'(\mathfrak{R}(head(\varsigma(\rho)^c)))\} & = \\ & = I_m \cup \{h'(head(\rho))\} & = I_{m+1}. \end{array}$$

Finally, let q^c be the canonical rewriting of the UBCQ $q = \exists \mathbf{Y}_1 \psi_1(\mathbf{Y}_1) \vee \ldots \vee \exists \mathbf{Y}_k \psi_k(\mathbf{Y}_k)$. For each $j \in \{1,\ldots,k\}$, consider the safe substitution ς_j mapping each variable of $\psi_j(\mathbf{Y}_j)$ in a different null. Therefore, there exists a conjunction of atoms, say $\psi_j^c(\mathbf{Y}_j) = \varsigma_j(\psi_j(\mathbf{Y}_j))^c$ in q^c, such that $\mathfrak{R}(\psi_j^c(\mathbf{Y}_j)) = \psi_j(\mathbf{Y}_j)$, for each $j \in \{1,\ldots,k\}$. Hence, $q \subseteq \mathfrak{R}(q^c)$. Moreover, it is easy to see that, each other safe substitution ς' w.r.t. some ψ_j, produces a conjunction of atoms, $\varsigma'(\psi_j(\mathbf{Y}_j))^c$ such that $\mathfrak{R}(\varsigma'(\psi_j(\mathbf{Y}_j))^c)$ is contained in $\mathfrak{R}(\varsigma_j(\psi_j(\mathbf{Y}_j))^c)$. Therefore, $\mathfrak{R}(q^c) \subseteq q$. Thus, $\mathfrak{R}(q^c) = q$. □

Proof of Theorem 3.1

We know that, for each database D, ontology Σ and UBCQ q, it holds that $D \cup \Sigma \models q$ if and only if $chase(D,\Sigma) \models q$ (Fagin et al. 2005). Therefore, also $D^c \cup \Sigma^c \models q^c$ if and only if $chase(D^c,\Sigma^c) \models q^c$. Moreover, by Proposition 3.2, we have that $\mathfrak{R}(chase(D^c,\Sigma^c)) = chase(D,\Sigma)$ and $\mathfrak{R}(q^c) \equiv q$. Hence, remain to prove that $\mathfrak{R}(chase(D^c,\Sigma^c)) \models \mathfrak{R}(q^c)$ if and only if $chase(D^c,\Sigma^c) \models q^c$.

We prove the "if" part, given that the "only if" part can be obtained retracing the chain of the following implications. Suppose that $chase(D^c,\Sigma^c) \models q^c$. Therefore, there is a homomorphism h from at least one disjunct of q^c, say $\varsigma_j(\psi_j(\mathbf{Y}_j))^c$ (where ς_j is a canonical substitution), to $chase(D^c,\Sigma^c)$, that is $h(\varsigma_j(\psi_j(\mathbf{Y}_j))^c) \subseteq chase(D^c,\Sigma^c)$. Therefore, $\mathfrak{R}(h(\varsigma_j(\psi_j(\mathbf{Y}_j))^c)) \subseteq \mathfrak{R}(chase(D^c,\Sigma^c))$. Moreover, note that $\mathfrak{R}(h(\varsigma_j(\psi_j(\mathbf{Y}_j))^c)) = h(\mathfrak{R}(\varsigma_j(\psi_j(\mathbf{Y}_j))^c))$. Hence, $h(\mathfrak{R}(\varsigma_j(\psi_j(\mathbf{Y}_j))^c)) \subseteq \mathfrak{R}(chase(D^c,\Sigma^c))$. Thus, h is also a homomorphism from a disjunct of $\mathfrak{R}(q^c)$ to $\mathfrak{R}(chase(D^c,\Sigma^c))$, that is $\mathfrak{R}(chase(D^c,\Sigma^c)) \models \mathfrak{R}(q^c)$. □

Proof of Proposition 4.1

Let Σ be a shy ontology. Note that, for each rule $\rho \in \Sigma$, there exists a rule $\varsigma(\rho)^c \in \Sigma^c$ such that $\varsigma(X^i) = n_i$ for each variable X^i occurring in ρ. It is easy to see that a such ς is a safe substitution. We denote by $\bar{\Sigma}^c$ the set of all and anly this kind of rules in Σ^c. Note that, if Σ^c is a shy ontology, then $\bar{\Sigma}^c \subseteq \Sigma^c$ is also a shy ontology.

By contradiction, suppose that $\bar{\Sigma}^c$ is not a shy ontology.

First, suppose that there exists a rule $\varsigma(\rho)^c \in \bar{\Sigma}^c$ such that there exists a variable, say X, occurring in more than one body atom and X is not protected in $body(\varsigma(\rho)^c)$. Therefore, for each existential variable Y, there exists an atom $\beta \in body(\varsigma(\rho)^c)$ and some position $pred(\beta)[i]$ in which X occurs, and $pred(\beta)[i]$ is not invaded by Y. Consider the unpacked rule $\mathfrak{R}(\varsigma(\rho)^c) = \rho \in \Sigma$. Therefore, by construction, for each existential variable Y, there exists $\alpha \in body(\rho)$ and some position $pred(\alpha)[j]$ in which X occurs, and $pred(\alpha)[j]$ is not invaded by Y. Hence, X occurs in more than one body atom of ρ and X is not protected in $body(\rho)$. So that, ρ is not a shy rule, and, thus, Σ is not a shy ontology.

Then, suppose that there exists a rule $\varsigma(\rho)^c \in \bar{\Sigma}^c$ such that there are two distinct universal variables, say X and Y, that are not protected in $body(\varsigma(\rho)^c)$; occur in $head(\varsigma(\rho)^c)$; occur in two different body atoms; and they are attacked by the same variable. Therefore, there exists an existential variable Z such that X and Y occur only in invaded position by Z. Consider again the unpacked rule $\mathfrak{R}(\varsigma(\rho)^c) = \rho \in \Sigma$. Then, by the unpacking function, X and Y are not protected in $body(\rho)$, and they occur in $head(\rho)$, in two different body atoms, and only in invaded position by Z. Thus, they are attacked by the same variable. Therefore, also in this case, ρ is not a shy rule. Hence, Σ is not a shy ontology. □

Proof of Proposition 4.2

Let M be a finite model of $D \cup \Sigma$. Clearly, if M is a well-supported finite model of $D \cup \Sigma$, we are done. Therefore, suppose that M is not a well-supported finite model of $D \cup \Sigma$. Let $\Omega_1 = (\alpha_1, \ldots, \alpha_m)$ be an ordering of the atoms of M. Hence, by assumption, there exists $\alpha \in M$ that is not a well-supported atom w.r.t. Ω_1. Let α_{j_1} be the first atom in the ordering Ω_1 that is not well-supported. And consider a new ordering $\Omega_2 = (\alpha_1, \ldots, \alpha_{j_1-1}, \alpha_{j_1+1}, \ldots, \alpha_m, \alpha_{j_1})$, where α_{j_1} is shifted from the position j_1 to the position n. As $M \notin wsfmods(D, \Sigma)$, then Ω_2 is not a well-supported ordering of M. Moreover, the first $j_1 - 1$ atoms are well-supported w.r.t. Ω_2. Therefore, let α_{j_2} be the first atom in the ordering Ω_2 that is not well-supported. Again, we consider a new ordering, say Ω_3, where α_{j_2} is shifted from position $j_2 - 1$ to the position n. Iteratively, we build a sequence $\Omega_1, \Omega_2, \ldots, \Omega_m, \ldots$ of orderings that are not well-supported. Note that, as the number of different orderings is finite, there exist at least two orderings in the sequence that are the same. Therefore, let Ω_{m_1} and Ω_{m_2} be the first two orderings of the sequence, with $m_2 > m_1$, such that $\Omega_{m_1} = \Omega_{m_2}$ (i.e., Ω_{m_1} and Ω_{m_2} are the same ordering). Consider the subset $A \subseteq M$ containing the first $n - (m_2 - m_1)$ elements in Ω_{m_1}, and the set B of the last $m_2 - m_1$ atoms in Ω_{m_1}. By construction, A is a well-supported instance. Moreover, each $\beta \in B$ is not well-supported by A, as $\Omega_{m_2} = \Omega_{m_1}$. That is, there is no rule ρ in Σ and no homomorphism h such that $h(body(\rho)) \subseteq A$ and $h(head(\rho)) = \{\beta\}$. Hence, as M is a model, whenever $A \models body(\rho)$, there exists an atom α in A, such that $\alpha \models head(\rho)$. Therefore, A is a model.

To complete the proof, let M be a finite minimal model of $D \cup \Sigma$. As just proved, there exists a well-supported finite model $M' \subseteq M$. By minimality of M, the model M' must be equal to M. Therefore, M is a well-supported finite model. □

Proof of Theorem 4.3

We have to prove that for each $M \in wsfmods(D^c, \Sigma^c_a)$, there exist $M' \in wsfmods(D^c, \Sigma^c)$ and a homomorphism h' such that $h'(M') \subseteq M$. Indeed, by hypothesis, there exists a homomorphism h such that $h(q) \subseteq M'$, and so $(h' \circ h)(q) \subseteq M$.

Let $M \in wsfmods(D^c, \Sigma^c_a)$, and let $(\alpha_1, \ldots, \alpha_m)$ be a well-supported ordering of M, and let $(\langle\alpha_1\rangle, \ldots, \langle\alpha_m\rangle)$ be a propagation ordering of $(\alpha_1, \ldots, \alpha_m)$. If there exists a join rule $\rho \in \Sigma^c$ satisfied by M with a null or a constant t in the join variables, then we consider the set of join atoms in the body of ρ w.r.t. the term t, say $A \subseteq M$. First, we substitute a term t of some $\alpha \in A$ in position l, with the corresponding term $\langle t, j, k\rangle$ of $\langle\alpha\rangle$, that can be considered as a fresh null. This new atom is denoted by α', so that $\alpha'[l] = \langle t, j, k\rangle$. Then, for each $\alpha_i \in M$ such that $\langle\alpha_i\rangle[l] = \langle t, j, k\rangle$, for some position l, we set $\alpha'_i[l] = \langle t, j, k\rangle$. Otherwise, $\alpha'_i[l] = \alpha_i[l]$. In this way, we build an instance $M' = \{\alpha' : \alpha \in M\}$ of Σ, and a homomorphism h' such that $h'(\langle t, j, k\rangle) = t$, for each introduced fresh null $\langle t, j, k\rangle$ to substitute t. By construction, it holds that $h'(\alpha') = \alpha$, so that $h'(M') = M$. Note that, by construction, M' is a well-supported finite instance of $D^c \cup \Sigma^c$.

Therefore, it remains to prove that M' is a model of $D^c \cup \Sigma^c$. By contradiction, suppose that M' is not a model. Hence, there exists a rule $\rho \in \Sigma^c$ such that $M' \models body(\rho)$, and $M' \not\models head(\rho)$. We distinguish two cases.

(i) First, suppose that ρ is not a join rule. Then, there exists a safe substitution $\hat{\varsigma}$, mapping each variable in the atoms of ρ into a different null, so that $\hat{\varsigma}(\rho)^c \in \Sigma^c_a$, as it is not a harmless rule of Shy. By hypothesis, $M' \models body(\rho)$, so that there exists a homomorphism h'' such that $h''(body(\rho)) \subseteq M'$. Therefore, $h'(h''(body(\rho))) \subseteq h'(M') = M$, and so $M \models body(\rho)$. Hence, also $M \models body(\hat{\varsigma}(\rho)^c)$. As M is a model of Σ^c_a, then $M \models head(\hat{\varsigma}(\rho)^c)$. Therefore, there exists a homomorphism h''' such that $h'''(head(\hat{\varsigma}(\rho)^c)) = \alpha_j$, for some $j \in \{1, \ldots, m\}$. Hence,

$\alpha_j \in M$. Therefore, $\alpha'_j \in M'$. Moreover, $\alpha'_j \models \mathit{head}(\rho)$, as $h'(\alpha'_j) = \alpha_j \models \mathit{head}(\rho)$. Therefore, $M' \models \mathit{head}(\rho)$.

(ii) Now, suppose that ρ is a join rule. Since, by hypothesis, $M' \models \mathit{body}(\rho)$, then, the join variables in the body of ρ are instantiated by the same null, as $D^c \cup \Sigma^c$ is a constant-free logical theory. However, by construction of M, it is not possible that the same term comes from an instantiation of two different existential variables, since we replaced each such instantiation with a fresh null in at least one joined term.

Therefore, M' is a well-supported finite model of $D^c \cup \Sigma^c$. □