Proofs for Parametric Schema Inference for Massive JSON Datasets

Mohamed-Amine Baazizi[1], Dario Colazzo[2], Giorgio Ghelli[3], and Carlo Sartiani[4]

[1]Sorbonne Université, CNRS, LIP6, 75005 Paris, France
[2]PSL Research University, CNRS, LAMSADE, 75016 Paris, France
[3]Dipartimento di Informatica, Università di Pisa, Italy
[4]DIMIE - Università della Basilicata

1 Proofs of the properties of *Reduce*

We present here the proofs of the main lemmas and theorems.

Property 2 (Stability of $\doteq$) For any $\doteq$-reduced types $\mathcal{T}_1$ and $\mathcal{T}_2$ and any two $\doteq$-reduced structural types $\mathcal{S}_1$ and $\mathcal{S}_2$, the following properties hold:

$$\mathcal{T}_1 \doteq \mathcal{T}_2 \;\Rightarrow\; \mathit{Reduce}(\mathcal{T}_1, \mathcal{T}_2, \doteq\,) \;\doteq\; \mathcal{T}_1 \;\doteq\; \mathcal{T}_2 \quad (1)$$
$$\mathcal{S}_1 \doteq \mathcal{S}_2 \;\Rightarrow\; \mathit{Fuse}(\mathcal{S}_1, \mathcal{S}_2, \doteq\,) \;\doteq\; \mathcal{S}_1 \;\doteq\; \mathcal{S}_2 \quad (2)$$

Proof. By mutual induction and by cases on the common kind of $\mathcal{S}_1$ and $\mathcal{S}_2$. Property (1): here we observe that every addend of $\circ\mathcal{T}_1$ has one $\doteq$-equivalent addend in $\circ\mathcal{T}_2$, by definition of $\doteq$, and only one, because the two types are $\doteq$-reduced. Hence, the result has one structural addend for each structural addend of $\circ\mathcal{T}_1$, and the two addends are $\doteq$-equivalent by induction. The other interesting case is the record type case of property (2). Here, by definition of $\doteq$, two record types are only fused when they have exactly the same keys and, for any key k in $\mathit{Keys}(\mathcal{R}_1)$, the types associated to k in $\mathcal{R}_1$ and $\mathcal{R}_2$ are $\doteq$ equivalent, hence, by (1), the type associated in the fused type is equivalent as well. The case for array types is immediate by (1), and the cases for the base types are immediate. ■

Corollary 1 (Lossless reduction)
For any $\doteq$*-reduced types* $\mathcal{T}_1$ *and* $\mathcal{T}_2$*:*

$$Reduce(\mathcal{T}_1, \mathcal{T}_2, \doteq) \quad \simeq \quad \mathcal{T}_1 + \mathcal{T}_2$$

Proof. The reduction process substitutes, inside $\mathcal{T}_1 + \mathcal{T}_2$, two equivalent addends $\mathcal{S}_1 \doteq \mathcal{S}_2$ with $Fuse(\mathcal{S}_1, \mathcal{S}_2, \doteq)$ which is, by Property 2 , syntactically congruent to each of them, hence is $\simeq$-equivalent to each of them, hence is $\simeq$-equivalent to their union. ■

We now introduce a bit of notation that will be used in all the proofs.

Notation 1.1 *For any SKER E, and any two E-reduced sets of structural types* $\mathcal{M}_1$ *and* $\mathcal{M}_2$*, and for any two sets* $\mathcal{F}_1$*,* $\mathcal{F}_2$ *of triples* $(k_i, \mathcal{T}_i, \mathsf{q}_i)$*, where each* $\mathcal{T}_i$ *is an E-reduced type, we define the following notation.*

$$\begin{array}{lcl}
\mathcal{M}_1 \setminus_E \mathcal{M}_2 & \triangleq & \{\!|\ \mathcal{S}_1 \in \mathcal{M}_1 \mid \nexists \mathcal{S}_2 \in \mathcal{M}_2.\ E(\mathcal{S}_1, \mathcal{S}_2)\ |\!\} \\
\mathcal{M}_1 \cap_E \mathcal{M}_2 & \triangleq & \{\!|\ \mathcal{S}_1 \in \mathcal{M}_1 \mid \exists \mathcal{S}_2 \in \mathcal{M}_2.\ E(\mathcal{S}_1, \mathcal{S}_2)\ |\!\} \\
\mathcal{M}_1 \bowtie_E \mathcal{M}_2 & \triangleq & \{\!|\ Fuse(\mathcal{S}_1, \mathcal{S}_2, E) \\
 & & \quad \mid \mathcal{S}_1 \in \mathcal{M}_1, \mathcal{S}_2 \in \mathcal{M}_2, E(\mathcal{S}_1, \mathcal{S}_2)\ |\!\} \\
\mathcal{F}_1 \setminus_{::} \mathcal{F}_2 & \triangleq & \{\!|\ (k_1, \mathcal{T}_1, \mathsf{q}_1) \in \mathcal{F}_1 \\
 & & \quad \mid \nexists (k_2, \mathcal{T}_2, \mathsf{q}_2) \in \mathcal{F}_2.\ k_1 = k_2\ |\!\} \\
\mathcal{F}_1 \cap_{::} \mathcal{F}_2 & \triangleq & \{\!|\ (k_1, \mathcal{T}_1, \mathsf{q}_1) \in \mathcal{F}_1 \\
 & & \quad \mid \exists (k_2, \mathcal{T}_2, \mathsf{q}_2) \in \mathcal{F}_2.\ k_1 = k_2\ |\!\} \\
?(\mathcal{F}) & \triangleq & \{\!|\ (k, \mathcal{T}, ?) \mid (k, \mathcal{T}, \mathsf{q}) \in \mathcal{F}\ |\!\} \\
\mathcal{F}_1 \bowtie_{::} \mathcal{F}_2 & \triangleq & \{\!|\ (k_1, Reduce(\mathcal{T}_1, \mathcal{T}_2, E), \mathsf{q}_1 \cdot \mathsf{q}_2) \\
 & & \quad \mid (k_1, \mathcal{T}_1, \mathsf{q}_1) \in \mathcal{F}_1, (k_1, \mathcal{T}_2, \mathsf{q}_2) \in \mathcal{F}_2\ |\!\}
\end{array}$$

These operators allow us to rewrite the definition of *Reduce* and *Fuse* as follows.

Lemma 1.2

$$\begin{array}{l}
Reduce(\mathcal{T}_1, \mathcal{T}_2, E) \\
\doteq\ \oplus(\ \circ\mathcal{T}_1 \bowtie_E \circ\mathcal{T}_2\ \cup\ \circ\mathcal{T}_1 \setminus_E \circ\mathcal{T}_2\ \cup\ \circ\mathcal{T}_2 \setminus_E \circ\mathcal{T}_1\ ) \\
Fuse(\mathcal{R}_1, \mathcal{R}_2, E) \\
\doteq\ \underline{\{} \diamond\mathcal{R}_1 \bowtie_{::} \diamond\mathcal{R}_2 \cup ?(\diamond\mathcal{R}_1 \setminus_{::} \diamond\mathcal{R}_2) \cup ?(\diamond\mathcal{R}_2 \setminus_{::} \diamond\mathcal{R}_1) \underline{\}}
\end{array}$$

Lemma 1.3 *For any SKER E, and any two E-reduced types $\mathcal{T}_1$ and $\mathcal{T}_2$, the sets $\circ\mathcal{T}_1 \cap_E \circ\mathcal{T}_2$, $\circ\mathcal{T}_2 \cap_E \circ\mathcal{T}_1$, and $\circ\mathcal{T}_1 \bowtie_E \circ\mathcal{T}_2$, are all E-distinct, and, for each pair of them, the E relation defines a bijective function between the two.*

Proof. The sets $\circ\mathcal{T}_1 \cap_E \circ\mathcal{T}_2$ and $\circ\mathcal{T}_2 \cap_E \circ\mathcal{T}_1$ are E-distinct since each is a subset of a set that is E-distinct. The relation E defines an isomorphism between these two sets: every element of $\circ\mathcal{T}_1 \cap_E \circ\mathcal{T}_2$ E-corresponds to at least one element of $\circ\mathcal{T}_2 \cap_E \circ\mathcal{T}_1$ by construction, and it cannot E-correspond to two of them because, by transitivity, they would be E-equivalent, and the type $\mathcal{T}_2$ would then not be E-reduced. The same holds in the other direction, hence E defines a bijection, and it also defines a bijection between $\circ\mathcal{T}_1 \cap_E \circ\mathcal{T}_2$ and the following set of pairs, mapping every $\mathcal{S}_1$ to the only pair $(\mathcal{S}_1, \mathcal{S}_2)$ where $E(\mathcal{S}_1, \mathcal{S}_2)$:

$$\{\!|\ (\mathcal{S}_1, \mathcal{S}_2) \mid \mathcal{S}_1 \in \circ\mathcal{T}_1, \mathcal{S}_2 \in \circ\mathcal{T}_2, E(\mathcal{S}_1, \mathcal{S}_2)\ |\!\}$$

To every pair of this set, the element $\mathit{Fuse}(\mathcal{S}_1, \mathcal{S}_2, E)$ of $\circ\mathcal{T}_1 \bowtie_E \circ\mathcal{T}_2$ corresponds and vice versa. By stability, $\mathit{Fuse}(\mathcal{S}_1, \mathcal{S}_2, E)$ is E-equivalent to both $\mathcal{S}_1$ and $\mathcal{S}_2$, hence we can reason as in the previous case to prove, by transitivity, that no two distinct elements of $\circ\mathcal{T}_1 \bowtie_E \circ\mathcal{T}_2$ may be equivalent, hence it is E-reduced, and E is a bijection between it and both of $\circ\mathcal{T}_1 \cap_E \circ\mathcal{T}_2$ and $\circ\mathcal{T}_2 \cap_E \circ\mathcal{T}_1$. ∎

Proof of Lemmas 1 and 2 *The following properties hold.*

1. *For any two E-reduced types $\mathcal{T}_1$, $\mathcal{T}_2$,*
   $\mathit{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E)$ *is E-reduced*
2. *For any two E-reduced structural types $\mathcal{S}_1$, $\mathcal{S}_2$,*
   $\mathit{Fuse}(\mathcal{S}_1, \mathcal{S}_2, E)$ *is E-reduced*
3. *For any J, $\mathcal{S}$,*
   $\vdash^E J : \mathcal{S} \ \Rightarrow\ \mathcal{S}$ *is E-reduced*
4. *For any $J_1, \ldots, J_n$, $\mathcal{T}$,*
   $\vdash^E J_1, \ldots, J_n :^{\mathfrak{c}} \mathcal{T} \ \Rightarrow\ \mathcal{T}$ *is E-reduced*

Proof. The first two items are proved my mutual induction. The only interesting case is

$$\begin{array}{l} \mathit{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E) \\ \doteq\ \oplus(\ \circ\mathcal{T}_1 \bowtie_E \circ\mathcal{T}_2\ \cup\ \circ\mathcal{T}_1 \setminus_E \circ\mathcal{T}_2\ \cup\ \circ\mathcal{T}_2 \setminus_E \circ\mathcal{T}_1\ ) \end{array}$$

The set $\circ\mathcal{T}_1 \bowtie_E \circ\mathcal{T}_2$ is E-reduced by Lemma 1.3, and $\circ\mathcal{T}_1 \setminus_E \circ\mathcal{T}_2$ and $\circ\mathcal{T}_2 \setminus_E \circ\mathcal{T}_1$ are included in $\circ\mathcal{T}_1$ and $\circ\mathcal{T}_2$, which are E-reduced by hypothesis. We have hence just to prove that two structural types coming from two different sets among $\circ\mathcal{T}_1 \bowtie_E \circ\mathcal{T}_2$, $\circ\mathcal{T}_1 \setminus_E \circ\mathcal{T}_2$ and $\circ\mathcal{T}_2 \setminus_E \circ\mathcal{T}_1$ cannot be E-equivalent. If one of them comes from $\circ\mathcal{T}_1 \bowtie_E \circ\mathcal{T}_2$ and the other from $\circ\mathcal{T}_1 \setminus_E \circ\mathcal{T}_2$, they cannot be equivalent since the first is E-isomorphic to $\circ\mathcal{T}_1 \cap_E \circ\mathcal{T}_2$, and elements from $\circ\mathcal{T}_1 \setminus_E \circ\mathcal{T}_2$ cannot be equivalent to any element of $\circ\mathcal{T}_2$. The same holds for $\circ\mathcal{T}_1 \bowtie_E \circ\mathcal{T}_2$ and $\circ\mathcal{T}_2 \setminus_E \circ\mathcal{T}_1$. Finally, no element of $\circ\mathcal{T}_1 \setminus_E \circ\mathcal{T}_2$ may be equivalent to one element of $\circ\mathcal{T}_2 \setminus_E \circ\mathcal{T}_1$ since $\circ\mathcal{T}_1 \setminus_E \circ\mathcal{T}_2$ only contains types that are not equivalent to any element of $\circ\mathcal{T}_2$.

Properties (3) and (4) follow immediately, since all the union types that are produced by the judgments for $\vdash^E J : \mathcal{S}$ and $\vdash^E J :^{\mathfrak{c}} \mathcal{T}$ are actually produced by a $\mathit{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E)$ operation applied to arguments that are E-reduced by induction hypothesis.

■

We can now prove the inclusion theorem.

Theorem 3 (Inclusion)

For any SKER E and for any two E-reduced types $\mathcal{T}_1$ and $\mathcal{T}_2$:

$$\mathcal{T}_1 + \mathcal{T}_2 \leq \mathit{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E)$$

For any two E-reduced structural types $\mathcal{S}_1$ and $\mathcal{S}_2$:

$$E(\mathcal{S}_1, \mathcal{S}_2) \;\Rightarrow\; \mathcal{S}_1 + \mathcal{S}_2 \leq \mathit{Fuse}(\mathcal{S}_1, \mathcal{S}_2, E)$$

Proof. By mutual induction.

We want to prove that:

$$\begin{array}{l} \mathcal{T}_1 + \mathcal{T}_2 \\ \leq \; \oplus(\; \circ\mathcal{T}_1 \bowtie_E \circ\mathcal{T}_2 \;\cup\; \circ\mathcal{T}_1 \setminus_E \circ\mathcal{T}_2 \;\cup\; \circ\mathcal{T}_2 \setminus_E \circ\mathcal{T}_1 \;) \end{array}$$

That is:

$$\begin{array}{l} \oplus(\circ(\mathcal{T}_1 + \mathcal{T}_2)) \\ \leq \; \oplus(\; \circ\mathcal{T}_1 \bowtie_E \circ\mathcal{T}_2 \;\cup\; \circ\mathcal{T}_1 \setminus_E \circ\mathcal{T}_2 \;\cup\; \circ\mathcal{T}_2 \setminus_E \circ\mathcal{T}_1 \;) \end{array}$$

That is:

$$\begin{array}{l} \mathcal{S} \in (\circ(\mathcal{T}_1 + \mathcal{T}_2)) \Rightarrow \\ [\![\, \mathcal{S} \,]\!] \;\subseteq\; \bigcup_{\mathcal{S}' \in (\circ\mathcal{T}_1 \bowtie_E \circ\mathcal{T}_2 \;\cup\; \circ\mathcal{T}_1 \setminus_E \circ\mathcal{T}_2 \;\cup\; \circ\mathcal{T}_2 \setminus_E \circ\mathcal{T}_1)} [\![\, \mathcal{S}' \,]\!] \end{array}$$

The set $\circ(\mathcal{T}_1+\mathcal{T}_2)$ can be decomposed as follows.

$$\begin{aligned}\circ(\mathcal{T}_1+\mathcal{T}_2) &= (\circ\mathcal{T}_1 \cap_E \circ\mathcal{T}_2) \;\cup\; (\circ\mathcal{T}_1 \setminus_E \circ\mathcal{T}_2) \\ &\quad \cup (\circ\mathcal{T}_2 \cap_E \circ\mathcal{T}_1) \;\cup\; (\circ\mathcal{T}_2 \setminus_E \circ\mathcal{T}_1)\end{aligned}$$

If $\mathcal{S} \in \circ\mathcal{T}_1 \cap_E \circ\mathcal{T}_2$, then there exists $\mathcal{S}_2 \in \circ\mathcal{T}_2$ with $E(\mathcal{S},\mathcal{S}_2)$ such that $\mathit{Fuse}(\mathcal{S},\mathcal{S}_2,E)$ belongs to $\circ\mathcal{T}_1 \bowtie_E \circ\mathcal{T}_2$, and, by induction, we know that:

$$E(\mathcal{S},\mathcal{S}_2) \;\Rightarrow\; [\![\, \mathcal{S} \,]\!] \subseteq [\![\, \mathcal{S}+\mathcal{S}_2 \,]\!] \subseteq [\![\, \mathit{Fuse}(\mathcal{S},\mathcal{S}_2,E) \,]\!]$$

The case for $\mathcal{S} \in \circ\mathcal{T}_2 \cap_E \circ\mathcal{T}_1$ is analogous. The other two cases, $\mathcal{S} \in \circ\mathcal{T}_1 \setminus_E \circ\mathcal{T}_2$ and $\mathcal{S} \in \circ\mathcal{T}_2 \setminus_E \circ\mathcal{T}_1$, are trivial.

We move now to the proof of

$$E(\mathcal{S}_1,\mathcal{S}_2) \;\Rightarrow\; \mathcal{S}_1+\mathcal{S}_2 \leq \mathit{Fuse}(\mathcal{S}_1,\mathcal{S}_2,E)$$

by cases on the common kind of $\mathcal{S}_1$ and $\mathcal{S}_2$.

If they belong to an atomic kind, the thesis is immediate.

If they are of array type, then we have $\mathcal{S}_1 = [\mathcal{T}_1]$ and $\mathcal{S}_2 = [\mathcal{T}_2]$. We want to prove:

$$\begin{aligned}[\![\, [\mathcal{T}_1] \,]\!] \cup [\![\, [\mathcal{T}_2] \,]\!] &\subseteq [\![\, \mathit{Fuse}([\mathcal{T}_1],[\mathcal{T}_2],E) \,]\!] \\ &= [\![\, [\mathit{Reduce}(\mathcal{T}_1,\mathcal{T}_2,E)] \,]\!]\end{aligned}$$

That is,

$$[\![\, [\mathcal{T}_1] \,]\!] \subseteq [\![\, [\mathit{Reduce}(\mathcal{T}_1,\mathcal{T}_2,E)] \,]\!]$$

and

$$[\![\, [\mathcal{T}_2] \,]\!] \subseteq [\![\, [\mathit{Reduce}(\mathcal{T}_1,\mathcal{T}_2,E)] \,]\!].$$

Let us prove the first. Assume that $\langle\!\langle V_1,\ldots,V_n \rangle\!\rangle \in [\![\, [\mathcal{T}_1] \,]\!]$. This implies that, for any i, we have that $V_i \in [\![\, \mathcal{T}_1 \,]\!]$.

By induction, $[\![\, \mathcal{T}_1 \,]\!] \subseteq [\![\, \mathit{Reduce}(\mathcal{T}_1,\mathcal{T}_2,E) \,]\!]$, hence, for any i, we have that $V_i \in [\![\, \mathit{Reduce}(\mathcal{T}_1,\mathcal{T}_2,E) \,]\!]$, hence $\langle\!\langle V_1,\ldots,V_n \rangle\!\rangle \in [\![\, [\mathit{Reduce}(\mathcal{T}_1,\mathcal{T}_2,E)] \,]\!]$.

The inclusion $[\![\, [\mathcal{T}_2] \,]\!] \subseteq [\![\, [\mathit{Reduce}(\mathcal{T}_1,\mathcal{T}_2,E)] \,]\!]$ can be proved in the same way.

The last case is that of record types, that is, $\mathcal{S}_1 = \underline{\{}\, \diamond\mathcal{S}_1 \,\underline{\}}$ and $\mathcal{S}_2 = \underline{\{}\, \diamond\mathcal{S}_2 \,\underline{\}}$.

We want to prove:

$$[\![\, \underline{\{}\, \diamond\mathcal{S}_1 \,\underline{\}} \,]\!] \cup [\![\, \underline{\{}\, \diamond\mathcal{S}_2 \,\underline{\}} \,]\!] \;\subseteq\; [\![\, \mathit{Fuse}(\underline{\{}\, \diamond\mathcal{S}_1 \,\underline{\}}, \underline{\{}\, \diamond\mathcal{S}_2 \,\underline{\}}, E) \,]\!]$$

We prove the case for $\mathcal{S}_1$, the one for $\mathcal{S}_2$ being analogous.

$$[\![\, \underline{\{}\, \diamond\mathcal{S}_1 \,\underline{\}} \,]\!] \;\subseteq\; [\![\, \mathit{Fuse}(\underline{\{}\, \diamond\mathcal{S}_1 \,\underline{\}}, \underline{\{}\, \diamond\mathcal{S}_2 \,\underline{\}}, E) \,]\!]$$

We rewrite it as follows:

$$\begin{array}{l} [\![\underline{\{} \mathcal{S}_1 \underline{\}}]\!] \\ \subseteq \; [\![\underline{\{} (\diamond\mathcal{S}_1 \bowtie_{::} \diamond\mathcal{S}_2) \cup ?(\diamond\mathcal{S}_1 \setminus_{::} \circ\mathcal{S}_2) \cup ?(\diamond\mathcal{S}_2 \setminus_{::} \circ\mathcal{S}_1) \; \underline{\}}]\!] \end{array}$$

Consider a record $\mathcal{V} \in [\![\underline{\{} \mathcal{S}_1 \underline{\}}]\!]$. By definition,

$$\mathcal{V} = \{\!| \; (k_1, \mathcal{V}_1), \ldots, (k_n, \mathcal{V}_n) \; |\!\}$$

such that:

1. for any $i \in 1...n$, $\exists \mathcal{T}_i, \mathsf{q}_i$ such that $(k_i, \mathcal{T}_i, \mathsf{q}_i)$ belongs to $\diamond\mathcal{S}_1$, and $\mathcal{V}_i \in [\![\mathcal{T}_i]\!]$
2. for any $(k_j, \mathcal{T}_j, !) \in \diamond\mathcal{S}_1$, a pair $(k_j, \mathcal{V}_j)$ is in $\mathcal{V}$.

We want to prove the same properties for $\mathcal{V}$ with respect to

$$\underline{\{} \; (\diamond\mathcal{S}_1 \bowtie_{::} \diamond\mathcal{S}_2) \cup ?(\diamond\mathcal{S}_1 \setminus_{::} \circ\mathcal{S}_2) \cup ?(\diamond\mathcal{S}_2 \setminus_{::} \circ\mathcal{S}_1) \, \underline{\}}$$

We first prove the first property. Assume that the pair $(k_i, \mathcal{V}_i)$ belongs to $\mathcal{V}$. By (1) above, we have a triple $(k_i, \mathcal{T}_i, \mathsf{q}_i)$ in $\diamond\mathcal{S}_1$ with $\mathcal{V}_i \in [\![\mathcal{T}_i]\!]$. If a matching k exists in $\mathcal{S}_2$, then we have a triple $(k_i, \mathit{Reduce}(\mathcal{T}_i, \mathcal{T}_2, E), \_)$ in $\diamond\mathcal{S}_1 \bowtie_{::} \diamond\mathcal{S}_2$. By induction, $[\![\mathcal{T}_i]\!] \subseteq [\![\mathit{Reduce}(\mathcal{T}_i, \mathcal{T}_2, E)]\!]$, hence $\mathcal{V}_i \in [\![\mathit{Reduce}(\mathcal{T}_i, \mathcal{T}_2, E)]\!]$, as required. If no matching k exists in $\mathcal{S}_2$, then we have a triple $(k_i, \mathcal{T}_i, ?)$ in $\diamond\mathcal{S}_1 \setminus_{::} \diamond\mathcal{S}_2$, and $\mathcal{V}_i \in [\![\mathcal{T}_i]\!]$ holds by hypothesis.

For the second property, every triple $(k_j, \mathcal{T}_j, !)$ in

$$(\diamond\mathcal{S}_1 \bowtie_{::} \diamond\mathcal{S}_2) \cup ?(\diamond\mathcal{S}_1 \setminus_{::} \circ\mathcal{S}_2) \cup ?(\diamond\mathcal{S}_2 \setminus_{::} \circ\mathcal{S}_1)$$

comes from the $\diamond\mathcal{S}_1 \bowtie_{::} \diamond\mathcal{S}_2$ component and, by definition of $\mathsf{q}_1 \cdot \mathsf{q}_2$, it corresponds to a triple $(k_j, \_, !)$ in $\diamond\mathcal{S}_1$, hence $\mathcal{V}$ contains a field with the key k_j by hypothesis.

■

We can now prove that the $\mathit{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E)$ operator enjoys the commutativity and associativity properties that enable an efficient distributed map-reduce implementation.

Theorem 4 (Commutativity)

1. *Given two E-reduced types $\mathcal{T}_1$, $\mathcal{T}_2$, we have:*

$$\mathit{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E) \; \doteq \; \mathit{Reduce}(\mathcal{T}_2, \mathcal{T}_1, E)$$

2. *Given two structural E-reduced types $\mathcal{S}_1$ and $\mathcal{S}_2$ we have:*

$$E(\mathcal{S}_1, \mathcal{S}_2) \;\Rightarrow\; \mathit{Fuse}(\mathcal{S}_1, \mathcal{S}_2, E) \;\doteq\; \mathit{Fuse}(\mathcal{S}_2, \mathcal{S}_1, E)$$

Proof. Immediate, since the definition is symmetric, modulo order, and E enjoys symmetry. ■

We need a simple lemma before proving the main theorem.

Lemma 1.4 (Distributivity of join over set union) *For any SKER E, for any E-reduced sets of structural types $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}$, and for any sets $\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}$ of triples $(k_i, \mathcal{T}_i, \mathtt{q}_i)$, where each $\mathcal{T}_i$ is an E-reduced type, the following equalities hold.*

$$\begin{array}{lcl}
(\mathcal{M}_1 \cup \mathcal{M}_2) \bowtie_E \mathcal{M} & = & (\mathcal{M}_1 \bowtie_E \mathcal{M}) \cup (\mathcal{M}_2 \bowtie_E \mathcal{M}) \\
(\mathcal{F}_1 \cup \mathcal{F}_2) \bowtie_{::} \mathcal{F} & = & (\mathcal{F}_1 \bowtie_{::} \mathcal{F}) \cup (\mathcal{F}_2 \bowtie_{::} \mathcal{F}) \\
\mathcal{M} \bowtie_E (\mathcal{M}_1 \cup \mathcal{M}_2) & = & (\mathcal{M} \bowtie_E \mathcal{M}_1) \cup (\mathcal{M} \bowtie_E \mathcal{M}_2) \\
\mathcal{F} \bowtie_{::} (\mathcal{F}_1 \cup \mathcal{F}_2) & = & (\mathcal{F} \bowtie_{::} \mathcal{F}_1) \cup (\mathcal{F} \bowtie_{::} \mathcal{F}_2)
\end{array}$$

Proof. By definition of $\bowtie_E$:

$$\begin{array}{l}
(\mathcal{M}_1 \cup \mathcal{M}_2) \bowtie_E \mathcal{M} \\
= \{\!\!\{\, \mathit{Fuse}(\mathcal{S}, \mathcal{S}', E) \mid \mathcal{S} \in \mathcal{M}_1 \cup \mathcal{M}_2,\ \mathcal{S}' \in \mathcal{M}, E(\mathcal{S}, \mathcal{S}') \,\}\!\!\} \\
= \{\!\!\{\, \mathit{Fuse}(\mathcal{S}, \mathcal{S}', E) \mid \mathcal{S} \in \mathcal{M}_1,\ \mathcal{S}' \in \mathcal{M}, E(\mathcal{S}, \mathcal{S}') \,\}\!\!\} \\
\quad \cup \{\!\!\{\, \mathit{Fuse}(\mathcal{S}, \mathcal{S}', E) \mid \mathcal{S} \in \mathcal{M}_2,\ \mathcal{S}' \in \mathcal{M}, E(\mathcal{S}, \mathcal{S}') \,\}\!\!\} \\
= (\mathcal{M}_1 \bowtie_E \mathcal{M}) \cup (\mathcal{M}_2 \bowtie_E \mathcal{M})
\end{array}$$

By definition of $\bowtie_{::}$:

$$\begin{array}{l}
(\mathcal{F}_1 \cup \mathcal{F}_2) \bowtie_{::} \mathcal{F} \\
= \{\!\!\{\, (k, \mathit{Reduce}(\mathcal{T}, \mathcal{T}', E), \mathtt{q} \cdot \mathtt{q}') \\
\qquad \mid (k, \mathcal{T}, \mathtt{q}) \in (\mathcal{F}_1 \cup \mathcal{F}_2), (k, \mathcal{T}', \mathtt{q}') \in \mathcal{F} \,\}\!\!\} \\
= \{\!\!\{\, (k, \mathit{Reduce}(\mathcal{T}, \mathcal{T}', E), \mathtt{q} \cdot \mathtt{q}') \\
\qquad \mid (k, \mathcal{T}, \mathtt{q}) \in \mathcal{F}_1, (k, \mathcal{T}', \mathtt{q}') \in \mathcal{F} \,\}\!\!\} \\
\quad \cup \{\!\!\{\, (k, \mathit{Reduce}(\mathcal{T}, \mathcal{T}', E), \mathtt{q} \cdot \mathtt{q}') \\
\qquad\quad \mid (k, \mathcal{T}, \mathtt{q}) \in \cup\mathcal{F}_2, (k, \mathcal{T}', \mathtt{q}') \in \mathcal{F} \,\}\!\!\} \\
= (\mathcal{F}_1 \bowtie_{::} \mathcal{F}) \cup (\mathcal{F}_2 \bowtie_{::} \mathcal{F})
\end{array}$$

The last two cases are analogous. ■

Theorem 4 (Associativity)

The following two properties hold, for any stable KER E.

1. *Given three E-reduced types $\mathcal{T}_1$, $\mathcal{T}_2$ and $\mathcal{T}_3$, we have*

$$\begin{array}{l} \mathit{Reduce}(\mathit{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E), \mathcal{T}_3, E) \\ \doteq\ \mathit{Reduce}(\mathcal{T}_1, \mathit{Reduce}(\mathcal{T}_2, \mathcal{T}_3, E), E) \end{array}$$

2. *Given three E-reduced structural types $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_3$ that are mutually E-equivalent, we have*

$$\begin{array}{l} \mathit{Fuse}(\mathit{Fuse}(\mathcal{S}_1, \mathcal{S}_2, E), \mathcal{S}_3, E) \\ \doteq\ \mathit{Fuse}(\mathcal{S}_1, \mathit{Fuse}(\mathcal{S}_2, \mathcal{S}_3, E), E) \end{array}$$

Proof. We proof (1) and (2) by mutual induction.

We first partition each of $\circ\mathcal{T}_1$, $\circ\mathcal{T}_2$ and $\circ\mathcal{T}_3$ in four parts, that correspond to four possible combinations of $\_ \cap_E \_$ and $\_ \setminus_E \_$, as follows.

$$\begin{array}{lcll} M_1^{23} & = & \{\!\mid \mathcal{S}_1 \in \circ\mathcal{T}_1 \mid & \exists \mathcal{S}_2 \in \circ\mathcal{T}_2.\ E(\mathcal{S}_1, \mathcal{S}_2), \\ & & & \exists \mathcal{S}_3 \in \circ\mathcal{T}_3.\ E(\mathcal{S}_1, \mathcal{S}_3)\ \mid\!\} \\ M_1^{2\not3} & = & \{\!\mid \mathcal{S}_1 \in \circ\mathcal{T}_1 \mid & \exists \mathcal{S}_2 \in \circ\mathcal{T}_2.\ E(\mathcal{S}_1, \mathcal{S}_2), \\ & & & \nexists \mathcal{S}_3 \in \circ\mathcal{T}_3.\ E(\mathcal{S}_1, \mathcal{S}_3)\ \mid\!\} \\ M_1^{\not23} & = & \{\!\mid \mathcal{S}_1 \in \circ\mathcal{T}_1 \mid & \nexists \mathcal{S}_2 \in \circ\mathcal{T}_2.\ E(\mathcal{S}_1, \mathcal{S}_2), \\ & & & \exists \mathcal{S}_3 \in \circ\mathcal{T}_3.\ E(\mathcal{S}_1, \mathcal{S}_3)\ \mid\!\} \\ M_1^{\not2\not3} & = & \{\!\mid \mathcal{S}_1 \in \circ\mathcal{T}_1 \mid & \nexists \mathcal{S}_2 \in \circ\mathcal{T}_2.\ E(\mathcal{S}_1, \mathcal{S}_2), \\ & & & \nexists \mathcal{S}_3 \in \circ\mathcal{T}_3.\ E(\mathcal{S}_1, \mathcal{S}_3)\ \mid\!\} \end{array}$$

The partitions $\{\!\mid M_2^{13}, M_2^{1\not3}, M_2^{\not13}, M_2^{\not1\not3} \mid\!\}$ of $\circ\mathcal{T}_2$ and $\{\!\mid M_3^{12}, M_3^{1\not2}, M_3^{\not12}, M_3^{\not1\not2} \mid\!\}$ of $\circ\mathcal{T}_3$ are defined in the same way. Now we can decompose $\circ\mathit{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E)$ as follows. In all of our computations we will make use of distributivity of join over set union (Lemma 1.4).

$$\begin{array}{rcl} \circ\mathit{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E) & = & ((M_1^{23} \cup M_1^{2\not3}) \bowtie_E (M_2^{13} \cup M_2^{1\not3})) \\ & & \cup M_1^{\not23} \cup M_1^{\not2\not3} \cup M_2^{\not13} \cup M_2^{\not1\not3} \\ & = & ((M_1^{23} \bowtie_E M_2^{13}) \cup (M_1^{2\not3} \bowtie_E M_2^{1\not3})) \\ & & \cup M_1^{\not23} \cup M_1^{\not2\not3} \cup M_2^{\not13} \cup M_2^{\not1\not3} \end{array}$$

Now we compute $\circ\mathit{Reduce}(\mathit{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E), \mathcal{T}_3, E)$. The first two lines join the components of $\circ\mathit{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E)$ that match some component of $\circ\mathcal{T}_3$ with the corresponding component of $\circ\mathcal{T}_3$, while the last line lists all the non-matching components of $\circ\mathit{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E)$ and $\circ\mathcal{T}_3$.

$$\begin{array}{l} \circ\mathit{Reduce}(\mathit{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E), \mathcal{T}_3, E) = \\ \qquad ((M_1^{23} \bowtie_E M_2^{13}) \bowtie_E M_3^{12}) \\ \qquad \cup\ (M_1^{\not23} \bowtie_E M_3^{1\not2}) \cup (M_2^{\not13} \bowtie_E M_3^{\not12}) \\ \qquad \cup\ (M_1^{2\not3} \bowtie_E M_2^{1\not3}) \cup M_1^{\not2\not3} \cup M_2^{\not1\not3} \cup\ M_3^{\not1\not2} \end{array}$$

By reordering the components, we have the following equation for $\circ Reduce(Reduce(\mathcal{T}_1, \mathcal{T}_2, E), \mathcal{T}_3, E)$.

$$
\begin{aligned}
&\circ Reduce(Reduce(\mathcal{T}_1, \mathcal{T}_2, E), \mathcal{T}_3, E) = \\
&((M_1^{23} \bowtie_E M_2^{13}) \bowtie_E M_3^{12}) \\
&\cup(M_1^{\cancel{23}} \bowtie_E M_2^{\cancel{13}}) \cup (M_1^{\cancel{23}} \bowtie_E M_3^{\cancel{12}}) \cup (M_2^{\cancel{13}} \bowtie_E M_3^{\cancel{12}}) \\
&\cup M_1^{\cancel{23}} \cup M_2^{\cancel{13}} \cup M_3^{\cancel{12}}
\end{aligned}
$$

The same computation for $\circ Reduce(\mathcal{T}_1, Reduce(\mathcal{T}_2, \mathcal{T}_3, E), E)$ yields the same result with the only exception of the first term.

$$
\begin{aligned}
&\circ Reduce(\mathcal{T}_1, Reduce(\mathcal{T}_2, \mathcal{T}_3, E), E) = \\
&(M_1^{23} \bowtie_E (M_2^{13} \bowtie_E M_3^{12})) \\
&\cup(M_1^{\cancel{23}} \bowtie_E M_2^{\cancel{13}}) \cup (M_1^{\cancel{23}} \bowtie_E M_3^{\cancel{12}}) \cup \ (M_2^{\cancel{13}} \bowtie_E M_3^{\cancel{12}}) \\
&\cup M_1^{\cancel{23}} \cup M_2^{\cancel{13}} \cup M_3^{\cancel{12}}
\end{aligned}
$$

Hence, we only have to prove that

$$((M_1^{23} \bowtie_E M_2^{13}) \bowtie_E M_3^{12}) = (M_1^{23} \bowtie_E (M_2^{13} \bowtie_E M_3^{12}))$$

By definition, we have the following equalities.

$$
\begin{aligned}
&((M_1^{23} \bowtie_E M_2^{13}) \bowtie_E M_3^{12}) \\
&= \{\!|\ Fuse(\mathcal{S}_1, \mathcal{S}_2, E) \\
&\qquad\qquad \mid \mathcal{S}_1 \in M_1^{23}, \mathcal{S}_2 \in M_2^{13},\ E(\mathcal{S}_1, \mathcal{S}_2)\ |\!\} \bowtie_E M_3^{12} \\
&= \{\!|\ Fuse(Fuse(\mathcal{S}_1, \mathcal{S}_2, E), \mathcal{S}_3, E) \\
&\qquad\qquad \mid \mathcal{S}_1 \in M_1^{23}, \mathcal{S}_2 \in M_2^{13}, \mathcal{S}_3 \in M_3^{12}, \\
&\qquad\qquad\ \ E(\mathcal{S}_1, \mathcal{S}_2), E(Fuse(\mathcal{S}_1, \mathcal{S}_2, E), \mathcal{S}_3)\ |\!\} \\
&(M_1^{23} \bowtie_E (M_2^{13} \bowtie_E M_3^{12})) \\
&= \{\!|\ Fuse(\mathcal{S}_1, Fuse(\mathcal{S}_2, \mathcal{S}_3, E), E) \\
&\qquad\qquad \mid \mathcal{S}_1 \in M_1^{23}, \mathcal{S}_2 \in M_2^{13}, \mathcal{S}_3 \in M_3^{12}, \\
&\qquad\qquad\ \ E(\mathcal{S}_2, \mathcal{S}_3), E(\mathcal{S}_1, Fuse(\mathcal{S}_2, \mathcal{S}_3, E))\ |\!\}
\end{aligned}
$$

By stability, both

$$E(\mathcal{S}_1, \mathcal{S}_2) \wedge E(Fuse(\mathcal{S}_1, \mathcal{S}_2, E), \mathcal{S}_3)$$

and

$$E(\mathcal{S}_2, \mathcal{S}_3) \wedge E(\mathcal{S}_1, Fuse(\mathcal{S}_2, \mathcal{S}_3, E))$$

can be rewritten as

$$E(\mathcal{S}_1, \mathcal{S}_2) \wedge E(\mathcal{S}_2, \mathcal{S}_3),$$

while $\mathit{Fuse}(\mathit{Fuse}(\mathcal{S}_1, \mathcal{S}_2, E), \mathcal{S}_3, E)$ is equivalent to

$$\mathit{Fuse}(\mathcal{S}_1, \mathit{Fuse}(\mathcal{S}_2, \mathcal{S}_3, E), E)$$

by induction, hence we conclude.

(2) Observe that $\mathcal{S}_1$, $\mathcal{S}_2$, and $\mathcal{S}_3$ have the same kind, by the hypothesis that they are mutually E-equivalent. We prove (2) by cases on their kind.

If they have an atomic kind, the thesis follows by definition of *Reduce*.

If they are of array type, then we have $\mathcal{S}_1 = [\mathcal{T}_1]$, $\mathcal{S}_2 = [\mathcal{T}_2]$, and $\mathcal{S}_3 = [\mathcal{T}_3]$, for some $\mathcal{T}_1$, $\mathcal{T}_2$, and $\mathcal{T}_3$, and we have:

$$\begin{array}{l} \mathit{Fuse}(\mathit{Fuse}([\mathcal{T}_1], [\mathcal{T}_2], E), [\mathcal{T}_3], E) \\ \qquad \doteq \mathit{Fuse}([\mathit{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E)], [\mathcal{T}_3], E) \\ \qquad \doteq [\mathit{Reduce}(\mathit{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E), \mathcal{T}_3, E)] \\ \mathit{Fuse}([\mathcal{T}_1], \mathit{Fuse}([\mathcal{T}_2], [\mathcal{T}_3], E), E) \\ \qquad \doteq \mathit{Fuse}([\mathcal{T}_1], [\mathit{Reduce}(\mathcal{T}_2, \mathcal{T}_3, E)], E) \\ \qquad \doteq [\mathit{Reduce}(\mathcal{T}_1, \mathit{Reduce}(\mathcal{T}_2, \mathcal{T}_3, E), E)] \end{array}$$

The thesis follows by case (1) and mutual induction.

The last case is that of record types, that is, $\mathcal{S}_1 = \underline{\{}\, \diamond\mathcal{S}_1 \,\underline{\}}$, $\mathcal{S}_2 = \underline{\{}\, \diamond\mathcal{S}_2 \,\underline{\}}$, and $\mathcal{S}_3 = \underline{\{}\, \diamond\mathcal{S}_3 \,\underline{\}}$.

We will follow the same structure as in the proof of the first case, that of $\mathit{Reduce}(\mathit{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E), \mathcal{T}_3, E)$.

As in the first case, we partition $\diamond\mathcal{S}_1$ in four parts F_1^{23}, $F_1^{2\not3}$, $F_1^{\not2 3}$, $F_1^{\not2\not3}$, according to the existence of a matching field in $\diamond\mathcal{S}_2$ and of a matching field in $\diamond\mathcal{S}_3$.

$$\begin{array}{rcl} F_1^{23} & = & (\diamond\mathcal{S}_1 \cap_{::} \diamond\mathcal{S}_2) \cap_{::} \diamond\mathcal{S}_3 \\ F_1^{2\not3} & = & (\diamond\mathcal{S}_1 \cap_{::} \diamond\mathcal{S}_2) \setminus_{::} \diamond\mathcal{S}_3 \\ F_1^{\not2 3} & = & (\diamond\mathcal{S}_1 \setminus_{::} \diamond\mathcal{S}_2) \cap_{::} \diamond\mathcal{S}_3 \\ F_1^{\not2\not3} & = & (\diamond\mathcal{S}_1 \setminus_{::} \diamond\mathcal{S}_2) \setminus_{::} \diamond\mathcal{S}_3 \end{array}$$

Now we can decompose $\diamond\mathit{Fuse}(\mathcal{S}_1, \mathcal{S}_2, E)$ as follows.

$$\begin{array}{rcl} \diamond\mathit{Fuse}(\mathcal{S}_1, \mathcal{S}_2, E) & = & ((M_1^{23} \cup M_1^{2\not3}) \bowtie_E (M_2^{13} \cup M_2^{1\not3})) \\ & & \cup M_1^{\not2 3} \cup M_1^{\not2\not3} \cup M_2^{\not1\not3} \cup M_2^{\not1 3} \\ & = & ((M_1^{23} \bowtie_E M_2^{13}) \cup (M_1^{2\not3} \bowtie_E M_2^{1\not3})) \\ & & \cup M_1^{\not2 3} \cup M_1^{\not2\not3} \cup M_2^{\not1\not3} \cup M_2^{\not1 3} \end{array}$$

Now we compute $\diamond\mathit{Fuse}(\mathit{Fuse}(\mathcal{S}_1, \mathcal{S}_2, E), \mathcal{S}_3, E)$. The first two lines join the components of $\diamond\mathit{Fuse}(\mathcal{S}_1, \mathcal{S}_2, E)$ that match some component of $\diamond\mathcal{S}_3$ with the corresponding component of $\diamond\mathcal{S}_3$, while the last line lists all the non-matching components of $\diamond \mathit{Fuse}(\mathcal{S}_1, \mathcal{S}_2, E)$ and $\diamond\mathcal{S}_3$.

$$
\begin{aligned}
\diamond \mathit{Fuse}(\mathit{Fuse}(\mathcal{S}_1, \mathcal{S}_2, E), \mathcal{S}_3, E) = {} & \\
& ((F_1^{23} \bowtie_{::} F_2^{13}) \bowtie_{::} F_3^{12}) \\
& \cup\ (F_1^{\not{2}3} \bowtie_{::} F_3^{1\not{2}}) \cup (F_2^{\not{1}3} \bowtie_{::} F_3^{\not{1}2}) \\
& \cup\ (F_1^{2\not{3}} \bowtie_{::} F_2^{1\not{3}}) \cup F_1^{\not{2}\not{3}} \cup F_2^{\not{1}\not{3}} \cup\ F_3^{\not{1}\not{2}}
\end{aligned}
$$

By reordering the components, we have the following equation for $\diamond \mathit{Fuse}(\mathit{Fuse}(\mathcal{S}_1, \mathcal{S}_2, E), \mathcal{S}_3, E)$.

$$
\begin{aligned}
& \diamond \mathit{Fuse}(\mathit{Fuse}(\mathcal{S}_1, \mathcal{S}_2, E), \mathcal{S}_3, E) = \\
& ((F_1^{23} \bowtie_{::} F_2^{13}) \bowtie_{::} F_3^{12}) \\
& \cup(F_1^{2\not{3}} \bowtie_{::} F_2^{1\not{3}}) \cup (F_1^{\not{2}3} \bowtie_{::} F_3^{1\not{2}}) \cup (F_2^{\not{1}3} \bowtie_{::} F_3^{\not{1}2}) \\
& \cup F_1^{\not{2}\not{3}} \cup F_2^{\not{1}\not{3}} \cup F_3^{\not{1}\not{2}}
\end{aligned}
$$

The same computation for $\diamond \mathit{Fuse}(\mathcal{S}_1, \mathit{Fuse}(\mathcal{S}_2, \mathcal{S}_3, E), E)$ yields the same result with the only exception of the first term.

$$
\begin{aligned}
& \diamond \mathit{Fuse}(\mathcal{S}_1, \mathit{Fuse}(\mathcal{S}_2, \mathcal{S}_3, E), E) = \\
& (F_1^{23} \bowtie_{::} (F_2^{13} \bowtie_{::} F_3^{12})) \\
& \cup(F_1^{2\not{3}} \bowtie_{::} F_2^{1\not{3}}) \cup (F_1^{\not{2}3} \bowtie_{::} F_3^{1\not{2}}) \cup\ (F_2^{\not{1}3} \bowtie_{::} F_3^{\not{1}2}) \\
& \cup F_1^{\not{2}\not{3}} \cup F_2^{\not{1}\not{3}} \cup F_3^{\not{1}\not{2}}
\end{aligned}
$$

Hence, we only have to prove that

$$
((F_1^{23} \bowtie_{::} F_2^{13}) \bowtie_{::} F_3^{12}) = (F_1^{23} \bowtie_{::} (F_2^{13} \bowtie_{::} F_3^{12}))
$$

By definition, we have the following equalities.

$$
\begin{aligned}
& ((F_1^{23} \bowtie_{::} F_2^{13}) \bowtie_{::} F_3^{12}) \\
& = \{\!|\ (k, \mathit{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E), \mathsf{q}_1 \cdot \mathsf{q}_2) \\
& \qquad \mid (k, \mathcal{T}_1, \mathsf{q}_1) \in F_1^{23}, (k, \mathcal{T}_2, \mathsf{q}_2) \in F_2^{13}\ |\!\} \bowtie_{::} F_3^{12} \\
& = \{\!|\ (k, \mathit{Reduce}(\mathit{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E), \mathcal{T}_3, E), (\mathsf{q}_1 \cdot \mathsf{q}_2) \cdot \mathsf{q}_3) \\
& \qquad \mid (k, \mathcal{T}_1, \mathsf{q}_1) \in F_1^{23}, (k, \mathcal{T}_2, \mathsf{q}_2) \in F_2^{13}, \\
& \qquad\quad (k, \mathcal{T}_3, \mathsf{q}_3) \in F_3^{12}\ |\!\} \\
& (F_1^{23} \bowtie_{::} (F_2^{13} \bowtie_{::} F_3^{12})) \\
& = \{\!|\ (k, \mathit{Reduce}(\mathcal{T}_1, \mathit{Reduce}(\mathcal{T}_2, \mathcal{T}_3, E), E), \mathsf{q}_1 \cdot (\mathsf{q}_2 \cdot \mathsf{q}_3)) \\
& \qquad \mid (k, \mathcal{T}_1, \mathsf{q}_1) \in F_1^{23}, (k, \mathcal{T}_2, \mathsf{q}_2) \in F_2^{13}, \\
& \qquad\quad (k, \mathcal{T}_3, \mathsf{q}_3) \in F_3^{12}\ |\!\}
\end{aligned}
$$

By induction $\mathit{Reduce}(\mathit{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E), \mathcal{T}_3, E)$ is equivalent to $\mathit{Reduce}(\mathcal{T}_1, \mathit{Reduce}(\mathcal{T}_2, \mathcal{T}_3, E), E)$, associativity of $\mathsf{q}' \cdot \mathsf{q}''$ is immediate, hence we conclude.

■

Theorem 5

For any SKER E, for any JSON expressions $J, J_1, \ldots, J_n$*:*

$$\begin{array}{lcl} \vdash^E J : \mathcal{S} & \Rightarrow & [\![J]\!] \in [\![\mathcal{S}]\!] \\ \vdash^E J_1, \ldots, J_n :^{\mathfrak{c}} \mathcal{T} & \Rightarrow & \{\!\!\{ [\![J_1]\!], \ldots, [\![J_n]\!] \}\!\!\} \subseteq [\![\mathcal{T}]\!] \end{array}$$

Proof. We prove it by mutual induction on the size of the inference proof and by cases on the last applied rule. The base rules are trivial. The cases for the record and array rules are an immediate consequence of the semantics of records and arrays. The empty collection rule is trivial and the singleton rule follows immediately by induction. For the crucial (TypeCollection) rule, we know by induction that

$$\begin{array}{l} \{\!\!\{ [\![J_1]\!], \ldots, [\![J_i]\!] \}\!\!\} \subseteq [\![\mathcal{T}_1]\!] \\ \{\!\!\{ [\![J_{i+1}]\!], \ldots, [\![J_n]\!] \}\!\!\} \subseteq [\![\mathcal{T}_2]\!] \end{array}$$

By Theorem 2,

$$\mathcal{T}_1 \leq \mathit{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E) \text{ and } \mathcal{T}_2 \leq \mathit{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E)$$

Hence, by transitivity, we have that

$$\begin{array}{l} \{\!\!\{ [\![J_1]\!], \ldots, [\![J_i]\!] \}\!\!\} \subseteq [\![\mathit{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E)]\!] \\ \{\!\!\{ [\![J_{i+1}]\!], \ldots, [\![J_n]\!] \}\!\!\} \subseteq [\![\mathit{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E)]\!] \end{array}$$

hence

$$\{\!\!\{ [\![J_1]\!], \ldots, [\![J_n]\!] \}\!\!\} \subseteq [\![\mathit{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E)]\!].$$

■