



**HAL**  
open science

# Proofs for Parametric Schema Inference for Massive JSON Datasets

Mohamed-Amine Baazizi, Dario Colazzo, Giorgio Ghelli, Carlo Sartiani

► **To cite this version:**

Mohamed-Amine Baazizi, Dario Colazzo, Giorgio Ghelli, Carlo Sartiani. Proofs for Parametric Schema Inference for Massive JSON Datasets. 2018. hal-01960464

**HAL Id: hal-01960464**

**<https://hal.science/hal-01960464>**

Preprint submitted on 19 Dec 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Proofs for Parametric Schema Inference for Massive JSON Datasets

Mohamed-Amine Baazizi<sup>1</sup>, Dario Colazzo<sup>2</sup>, Giorgio Ghelli<sup>3</sup>, and Carlo Sartiani<sup>4</sup>

<sup>1</sup>Sorbonne Université, CNRS, LIP6, 75005 Paris, France

<sup>2</sup>PSL Research University, CNRS, LAMSADE, 75016 Paris, France

<sup>3</sup>Dipartimento di Informatica, Università di Pisa, Italy

<sup>4</sup>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:

$$\mathcal{T}_1 \doteq \mathcal{T}_2 \Rightarrow \text{Reduce}(\mathcal{T}_1, \mathcal{T}_2, \doteq) \doteq \mathcal{T}_1 \doteq \mathcal{T}_2 \quad (1)$$

$$\mathcal{S}_1 \doteq \mathcal{S}_2 \Rightarrow \text{Fuse}(\mathcal{S}_1, \mathcal{S}_2, \doteq) \doteq \mathcal{S}_1 \doteq \mathcal{S}_2 \quad (2)$$

**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  $\text{Keys}(\mathcal{R}_1)$ , 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  $\dot{=}$ -reduced types  $\mathcal{T}_1$  and  $\mathcal{T}_2$ :

$$\text{Reduce}(\mathcal{T}_1, \mathcal{T}_2, \dot{=}) \simeq \mathcal{T}_1 + \mathcal{T}_2$$

**Proof.** The reduction process substitutes, inside  $\mathcal{T}_1 + \mathcal{T}_2$ , two equivalent addends  $\mathcal{S}_1 \dot{=} \mathcal{S}_2$  with  $\text{Fuse}(\mathcal{S}_1, \mathcal{S}_2, \dot{=})$  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  $(k_i, \mathcal{T}_i, \mathbf{q}_i)$ , where each  $\mathcal{T}_i$  is an  $E$ -reduced type, we define the following notation.

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

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

**Lemma 1.2**

$$\begin{aligned} &\text{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E) \\ &\dot{=} \oplus( \circ\mathcal{T}_1 \bowtie_E \circ\mathcal{T}_2 \cup \circ\mathcal{T}_1 \setminus_E \circ\mathcal{T}_2 \cup \circ\mathcal{T}_2 \setminus_E \circ\mathcal{T}_1 ) \\ &\text{Fuse}(\mathcal{R}_1, \mathcal{R}_2, E) \\ &\dot{=} \{ \diamond\mathcal{R}_1 \bowtie_{::} \diamond\mathcal{R}_2 \cup ?(\diamond\mathcal{R}_1 \setminus_{::} \diamond\mathcal{R}_2) \cup ?(\diamond\mathcal{R}_2 \setminus_{::} \diamond\mathcal{R}_1) \} \end{aligned}$$

**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  $Fuse(\mathcal{S}_1, \mathcal{S}_2, E)$  of  $\circ\mathcal{T}_1 \bowtie_E \circ\mathcal{T}_2$  corresponds and vice versa. By stability,  $Fuse(\mathcal{S}_1, \mathcal{S}_2, E)$  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  $\circ\mathcal{T}_2 \cap_E \circ\mathcal{T}_1$ .  $\blacksquare$

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

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

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

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

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  $\vdash^E J : \mathcal{T}$  are actually produced by a  $Reduce(\mathcal{T}_1, \mathcal{T}_2, E)$  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  $\mathcal{T}_2$ :*

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

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

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

**Proof.** By mutual induction.

We want to prove that:

$$\begin{aligned} & \mathcal{T}_1 + \mathcal{T}_2 \\ & \leq \oplus(\circ\mathcal{T}_1 \bowtie_E \circ\mathcal{T}_2 \cup \circ\mathcal{T}_1 \setminus_E \circ\mathcal{T}_2 \cup \circ\mathcal{T}_2 \setminus_E \circ\mathcal{T}_1) \end{aligned}$$

That is:

$$\begin{aligned} & \oplus(\circ(\mathcal{T}_1 + \mathcal{T}_2)) \\ & \leq \oplus(\circ\mathcal{T}_1 \bowtie_E \circ\mathcal{T}_2 \cup \circ\mathcal{T}_1 \setminus_E \circ\mathcal{T}_2 \cup \circ\mathcal{T}_2 \setminus_E \circ\mathcal{T}_1) \end{aligned}$$

That is:

$$\begin{aligned} & \mathcal{S} \in (\circ(\mathcal{T}_1 + \mathcal{T}_2)) \Rightarrow \\ & \llbracket \mathcal{S} \rrbracket \subseteq \bigcup_{\mathcal{S}' \in (\circ\mathcal{T}_1 \bowtie_E \circ\mathcal{T}_2 \cup \circ\mathcal{T}_1 \setminus_E \circ\mathcal{T}_2 \cup \circ\mathcal{T}_2 \setminus_E \circ\mathcal{T}_1)} \llbracket \mathcal{S}' \rrbracket \end{aligned}$$

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  $Fuse(\mathcal{S}, \mathcal{S}_2, E)$  belongs to  $\circ\mathcal{T}_1 \bowtie_E \circ\mathcal{T}_2$ , and, by induction, we know that:

$$E(\mathcal{S}, \mathcal{S}_2) \Rightarrow \llbracket \mathcal{S} \rrbracket \subseteq \llbracket \mathcal{S} + \mathcal{S}_2 \rrbracket \subseteq \llbracket Fuse(\mathcal{S}, \mathcal{S}_2, E) \rrbracket$$

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

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

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:

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

That is,

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

and

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

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

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

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

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

We want to prove:

$$\llbracket \{\diamond\mathcal{S}_1\} \rrbracket \cup \llbracket \{\diamond\mathcal{S}_2\} \rrbracket \subseteq \llbracket Fuse(\{\diamond\mathcal{S}_1\}, \{\diamond\mathcal{S}_2\}, E) \rrbracket$$

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

$$\llbracket \{\diamond\mathcal{S}_1\} \rrbracket \subseteq \llbracket Fuse(\{\diamond\mathcal{S}_1\}, \{\diamond\mathcal{S}_2\}, E) \rrbracket$$

We rewrite it as follows:

$$\begin{aligned} & \llbracket \{ \underline{\mathcal{S}}_1 \} \rrbracket \\ & \subseteq \llbracket \{ (\diamond \mathcal{S}_1 \bowtie \diamond \mathcal{S}_2) \cup ?(\diamond \mathcal{S}_1 \setminus \diamond \mathcal{S}_2) \cup ?(\diamond \mathcal{S}_2 \setminus \diamond \mathcal{S}_1) \} \rrbracket \end{aligned}$$

Consider a record  $\mathcal{V} \in \llbracket \{ \underline{\mathcal{S}}_1 \} \rrbracket$ . By definition,

$$\mathcal{V} = \{ (k_1, \mathcal{V}_1), \dots, (k_n, \mathcal{V}_n) \}$$

such that:

1. for any  $i \in 1..n$ ,  $\exists \mathcal{T}_i, \mathbf{q}_i$  such that  $(k_i, \mathcal{T}_i, \mathbf{q}_i)$  belongs to  $\diamond \mathcal{S}_1$ , and  $\mathcal{V}_i \in \llbracket \mathcal{T}_i \rrbracket$
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

$$\{ \underline{(\diamond \mathcal{S}_1 \bowtie \diamond \mathcal{S}_2) \cup ?(\diamond \mathcal{S}_1 \setminus \diamond \mathcal{S}_2) \cup ?(\diamond \mathcal{S}_2 \setminus \diamond \mathcal{S}_1)} \}$$

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  $(k_i, \mathcal{T}_i, \mathbf{q}_i)$  in  $\diamond \mathcal{S}_1$  with  $\mathcal{V}_i \in \llbracket \mathcal{T}_i \rrbracket$ . If a matching  $k$  exists in  $\mathcal{S}_2$ , then we have a triple  $(k_i, \text{Reduce}(\mathcal{T}_i, \mathcal{T}_2, E), -)$  in  $\diamond \mathcal{S}_1 \bowtie \diamond \mathcal{S}_2$ . By induction,  $\llbracket \mathcal{T}_i \rrbracket \subseteq \llbracket \text{Reduce}(\mathcal{T}_i, \mathcal{T}_2, E) \rrbracket$ , hence  $\mathcal{V}_i \in \llbracket \text{Reduce}(\mathcal{T}_i, \mathcal{T}_2, E) \rrbracket$ , 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  $\mathcal{V}_i \in \llbracket \mathcal{T}_i \rrbracket$  holds by hypothesis.

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

$$(\diamond \mathcal{S}_1 \bowtie \diamond \mathcal{S}_2) \cup ?(\diamond \mathcal{S}_1 \setminus \diamond \mathcal{S}_2) \cup ?(\diamond \mathcal{S}_2 \setminus \diamond \mathcal{S}_1)$$

comes from the  $\diamond \mathcal{S}_1 \bowtie \diamond \mathcal{S}_2$  component and, by definition of  $\mathbf{q}_1 \cdot \mathbf{q}_2$ , 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  $\text{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E)$  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:

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

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

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

**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  $(k_i, \mathcal{T}_i, \mathbf{q}_i)$ , where each  $\mathcal{T}_i$  is an  $E$ -reduced type, the following equalities hold.*

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

**Proof.** By definition of  $\bowtie_E$ :

$$\begin{aligned} &(\mathcal{M}_1 \cup \mathcal{M}_2) \bowtie_E \mathcal{M} \\ &= \{ \{ Fuse(\mathcal{S}, \mathcal{S}', E) \mid \mathcal{S} \in \mathcal{M}_1 \cup \mathcal{M}_2, \mathcal{S}' \in \mathcal{M}, E(\mathcal{S}, \mathcal{S}') \} \} \\ &= \{ \{ Fuse(\mathcal{S}, \mathcal{S}', E) \mid \mathcal{S} \in \mathcal{M}_1, \mathcal{S}' \in \mathcal{M}, E(\mathcal{S}, \mathcal{S}') \} \} \\ &\quad \cup \{ \{ Fuse(\mathcal{S}, \mathcal{S}', E) \mid \mathcal{S} \in \mathcal{M}_2, \mathcal{S}' \in \mathcal{M}, E(\mathcal{S}, \mathcal{S}') \} \} \\ &= (\mathcal{M}_1 \bowtie_E \mathcal{M}) \cup (\mathcal{M}_2 \bowtie_E \mathcal{M}) \end{aligned}$$

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

$$\begin{aligned} &(\mathcal{F}_1 \cup \mathcal{F}_2) \bowtie_{::} \mathcal{F} \\ &= \{ \{ (k, Reduce(\mathcal{T}, \mathcal{T}', E), \mathbf{q} \cdot \mathbf{q}') \} \\ &\quad \mid (k, \mathcal{T}, \mathbf{q}) \in (\mathcal{F}_1 \cup \mathcal{F}_2), (k, \mathcal{T}', \mathbf{q}') \in \mathcal{F} \} \\ &= \{ \{ (k, Reduce(\mathcal{T}, \mathcal{T}', E), \mathbf{q} \cdot \mathbf{q}') \} \\ &\quad \mid (k, \mathcal{T}, \mathbf{q}) \in \mathcal{F}_1, (k, \mathcal{T}', \mathbf{q}') \in \mathcal{F} \} \\ &\quad \cup \{ \{ (k, Reduce(\mathcal{T}, \mathcal{T}', E), \mathbf{q} \cdot \mathbf{q}') \} \\ &\quad \mid (k, \mathcal{T}, \mathbf{q}) \in \cup \mathcal{F}_2, (k, \mathcal{T}', \mathbf{q}') \in \mathcal{F} \} \\ &= (\mathcal{F}_1 \bowtie_{::} \mathcal{F}) \cup (\mathcal{F}_2 \bowtie_{::} \mathcal{F}) \end{aligned}$$

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

$$\begin{aligned} & \text{Reduce}(\text{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E), \mathcal{T}_3, E) \\ & \doteq \text{Reduce}(\mathcal{T}_1, \text{Reduce}(\mathcal{T}_2, \mathcal{T}_3, E), E) \end{aligned}$$

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

$$\begin{aligned} & \text{Fuse}(\text{Fuse}(\mathcal{S}_1, \mathcal{S}_2, E), \mathcal{S}_3, E) \\ & \doteq \text{Fuse}(\mathcal{S}_1, \text{Fuse}(\mathcal{S}_2, \mathcal{S}_3, E), E) \end{aligned}$$

**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.

$$\begin{aligned} M_1^{23} &= \{ \mathcal{S}_1 \in \circ\mathcal{T}_1 \mid \exists \mathcal{S}_2 \in \circ\mathcal{T}_2. E(\mathcal{S}_1, \mathcal{S}_2), \\ & \quad \exists \mathcal{S}_3 \in \circ\mathcal{T}_3. E(\mathcal{S}_1, \mathcal{S}_3) \} \\ M_1^{2\bar{3}} &= \{ \mathcal{S}_1 \in \circ\mathcal{T}_1 \mid \exists \mathcal{S}_2 \in \circ\mathcal{T}_2. E(\mathcal{S}_1, \mathcal{S}_2), \\ & \quad \bar{\exists} \mathcal{S}_3 \in \circ\mathcal{T}_3. E(\mathcal{S}_1, \mathcal{S}_3) \} \\ M_1^{\bar{2}3} &= \{ \mathcal{S}_1 \in \circ\mathcal{T}_1 \mid \bar{\exists} \mathcal{S}_2 \in \circ\mathcal{T}_2. E(\mathcal{S}_1, \mathcal{S}_2), \\ & \quad \exists \mathcal{S}_3 \in \circ\mathcal{T}_3. E(\mathcal{S}_1, \mathcal{S}_3) \} \\ M_1^{\bar{2}\bar{3}} &= \{ \mathcal{S}_1 \in \circ\mathcal{T}_1 \mid \bar{\exists} \mathcal{S}_2 \in \circ\mathcal{T}_2. E(\mathcal{S}_1, \mathcal{S}_2), \\ & \quad \bar{\exists} \mathcal{S}_3 \in \circ\mathcal{T}_3. E(\mathcal{S}_1, \mathcal{S}_3) \} \end{aligned}$$

The partitions  $\{ M_2^{13}, M_2^{1\bar{3}}, M_2^{\bar{1}3}, M_2^{\bar{1}\bar{3}} \}$  of  $\circ\mathcal{T}_2$  and  $\{ M_3^{12}, M_3^{1\bar{2}}, M_3^{\bar{1}2}, M_3^{\bar{1}\bar{2}} \}$  of  $\circ\mathcal{T}_3$  are defined in the same way. Now we can decompose  $\circ\text{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E)$  as follows. In all of our computations we will make use of distributivity of join over set union (Lemma 1.4).

$$\begin{aligned} \circ\text{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E) &= ((M_1^{23} \cup M_1^{2\bar{3}}) \bowtie_E (M_2^{13} \cup M_2^{1\bar{3}})) \\ & \quad \cup M_1^{\bar{2}3} \cup M_1^{\bar{2}\bar{3}} \cup M_2^{\bar{1}3} \cup M_2^{\bar{1}\bar{3}} \\ &= ((M_1^{23} \bowtie_E M_2^{13}) \cup (M_1^{2\bar{3}} \bowtie_E M_2^{1\bar{3}})) \\ & \quad \cup M_1^{\bar{2}3} \cup M_1^{\bar{2}\bar{3}} \cup M_2^{\bar{1}3} \cup M_2^{\bar{1}\bar{3}} \end{aligned}$$

Now we compute  $\circ\text{Reduce}(\text{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E), \mathcal{T}_3, E)$ . The first two lines join the components of  $\circ\text{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E)$  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  $\circ\text{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E)$  and  $\circ\mathcal{T}_3$ .

$$\begin{aligned} \circ\text{Reduce}(\text{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E), \mathcal{T}_3, E) &= \\ & ((M_1^{23} \bowtie_E M_2^{13}) \bowtie_E M_3^{12}) \\ & \cup (M_1^{2\bar{3}} \bowtie_E M_3^{1\bar{2}}) \cup (M_2^{\bar{1}3} \bowtie_E M_3^{\bar{1}2}) \\ & \cup (M_1^{\bar{2}3} \bowtie_E M_2^{\bar{1}3}) \cup M_1^{\bar{2}\bar{3}} \cup M_2^{\bar{1}\bar{3}} \cup M_3^{\bar{1}\bar{2}} \end{aligned}$$

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

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

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

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

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.

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

By stability, both

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

and

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

can be rewritten as

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

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

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

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:

$$\begin{aligned} & Fuse(Fuse([\mathcal{T}_1], [\mathcal{T}_2], E), [\mathcal{T}_3], E) \\ & \doteq Fuse([Reduce(\mathcal{T}_1, \mathcal{T}_2, E)], [\mathcal{T}_3], E) \\ & \doteq [Reduce(Reduce(\mathcal{T}_1, \mathcal{T}_2, E), \mathcal{T}_3, E)] \\ & Fuse([\mathcal{T}_1], Fuse([\mathcal{T}_2], [\mathcal{T}_3], E), E) \\ & \doteq Fuse([\mathcal{T}_1], [Reduce(\mathcal{T}_2, \mathcal{T}_3, E)], E) \\ & \doteq [Reduce(\mathcal{T}_1, Reduce(\mathcal{T}_2, \mathcal{T}_3, E), E)] \end{aligned}$$

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

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

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

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

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

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

$$\begin{aligned} \diamond Fuse(\mathcal{S}_1, \mathcal{S}_2, E) &= ((M_1^{23} \cup M_1^{2\bar{3}}) \bowtie_E (M_2^{13} \cup M_2^{1\bar{3}})) \\ &\quad \cup M_1^{2\bar{3}} \cup M_1^{\bar{2}3} \cup M_2^{1\bar{3}} \cup M_2^{\bar{1}3} \\ &= ((M_1^{23} \bowtie_E M_2^{13}) \cup (M_1^{2\bar{3}} \bowtie_E M_2^{\bar{1}3})) \\ &\quad \cup M_1^{\bar{2}3} \cup M_1^{2\bar{3}} \cup M_2^{1\bar{3}} \cup M_2^{\bar{1}3} \end{aligned}$$

Now we compute  $\diamond Fuse(Fuse(\mathcal{S}_1, \mathcal{S}_2, E), \mathcal{S}_3, E)$ . The first two lines join the components of  $\diamond Fuse(\mathcal{S}_1, \mathcal{S}_2, E)$  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  $\diamond\text{Fuse}(\mathcal{S}_1, \mathcal{S}_2, E)$  and  $\diamond\mathcal{S}_3$ .

$$\begin{aligned} \diamond\text{Fuse}(\text{Fuse}(\mathcal{S}_1, \mathcal{S}_2, E), \mathcal{S}_3, E) = & \\ & ((F_1^{23} \bowtie F_2^{13}) \bowtie F_3^{12}) \\ & \cup (F_1^{23} \bowtie F_3^{12}) \cup (F_2^{13} \bowtie F_3^{12}) \\ & \cup (F_1^{23} \bowtie F_2^{13}) \cup F_1^{23} \cup F_2^{13} \cup F_3^{12} \end{aligned}$$

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

$$\begin{aligned} \diamond\text{Fuse}(\text{Fuse}(\mathcal{S}_1, \mathcal{S}_2, E), \mathcal{S}_3, E) = & \\ & ((F_1^{23} \bowtie F_2^{13}) \bowtie F_3^{12}) \\ & \cup (F_1^{23} \bowtie F_2^{13}) \cup (F_1^{23} \bowtie F_3^{12}) \cup (F_2^{13} \bowtie F_3^{12}) \\ & \cup F_1^{23} \cup F_2^{13} \cup F_3^{12} \end{aligned}$$

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

$$\begin{aligned} \diamond\text{Fuse}(\mathcal{S}_1, \text{Fuse}(\mathcal{S}_2, \mathcal{S}_3, E), E) = & \\ & (F_1^{23} \bowtie (F_2^{13} \bowtie F_3^{12})) \\ & \cup (F_1^{23} \bowtie F_2^{13}) \cup (F_1^{23} \bowtie F_3^{12}) \cup (F_2^{13} \bowtie F_3^{12}) \\ & \cup F_1^{23} \cup F_2^{13} \cup F_3^{12} \end{aligned}$$

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.

$$\begin{aligned} & ((F_1^{23} \bowtie F_2^{13}) \bowtie F_3^{12}) \\ & = \{ \! \{ (k, \text{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E), \mathbf{q}_1 \cdot \mathbf{q}_2) \\ & \quad | (k, \mathcal{T}_1, \mathbf{q}_1) \in F_1^{23}, (k, \mathcal{T}_2, \mathbf{q}_2) \in F_2^{13} \} \! \} \bowtie F_3^{12} \\ & = \{ \! \{ (k, \text{Reduce}(\text{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E), \mathcal{T}_3, E), (\mathbf{q}_1 \cdot \mathbf{q}_2) \cdot \mathbf{q}_3) \\ & \quad | (k, \mathcal{T}_1, \mathbf{q}_1) \in F_1^{23}, (k, \mathcal{T}_2, \mathbf{q}_2) \in F_2^{13}, \\ & \quad (k, \mathcal{T}_3, \mathbf{q}_3) \in F_3^{12} \} \! \} \\ & (F_1^{23} \bowtie (F_2^{13} \bowtie F_3^{12})) \\ & = \{ \! \{ (k, \text{Reduce}(\mathcal{T}_1, \text{Reduce}(\mathcal{T}_2, \mathcal{T}_3, E), E), \mathbf{q}_1 \cdot (\mathbf{q}_2 \cdot \mathbf{q}_3)) \\ & \quad | (k, \mathcal{T}_1, \mathbf{q}_1) \in F_1^{23}, (k, \mathcal{T}_2, \mathbf{q}_2) \in F_2^{13}, \\ & \quad (k, \mathcal{T}_3, \mathbf{q}_3) \in F_3^{12} \} \! \} \end{aligned}$$

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

**Theorem 5**

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

$$\begin{aligned} \vdash^E J : \mathcal{S} &\Rightarrow \llbracket J \rrbracket \in \llbracket \mathcal{S} \rrbracket \\ \vdash^E J_1, \dots, J_n :^c \mathcal{T} &\Rightarrow \{ \llbracket J_1 \rrbracket, \dots, \llbracket J_n \rrbracket \} \subseteq \llbracket \mathcal{T} \rrbracket \end{aligned}$$

**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

$$\begin{aligned} \{ \llbracket J_1 \rrbracket, \dots, \llbracket J_i \rrbracket \} &\subseteq \llbracket \mathcal{T}_1 \rrbracket \\ \{ \llbracket J_{i+1} \rrbracket, \dots, \llbracket J_n \rrbracket \} &\subseteq \llbracket \mathcal{T}_2 \rrbracket \end{aligned}$$

By Theorem 2,

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

Hence, by transitivity, we have that

$$\begin{aligned} \{ \llbracket J_1 \rrbracket, \dots, \llbracket J_i \rrbracket \} &\subseteq \llbracket \text{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E) \rrbracket \\ \{ \llbracket J_{i+1} \rrbracket, \dots, \llbracket J_n \rrbracket \} &\subseteq \llbracket \text{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E) \rrbracket \end{aligned}$$

hence

$$\{ \llbracket J_1 \rrbracket, \dots, \llbracket J_n \rrbracket \} \subseteq \llbracket \text{Reduce}(\mathcal{T}_1, \mathcal{T}_2, E) \rrbracket.$$

■