Proofs for Parametric Schema Inference for Massive JSON Datasets

Mohamed-Amine Baazizi\textsuperscript{1}, Dario Colazzo\textsuperscript{2}, Giorgio Ghelli\textsuperscript{3}, and Carlo Sartiani\textsuperscript{4}

\textsuperscript{1}Sorbonne Université, CNRS, LIP6, 75005 Paris, France
\textsuperscript{2}PSL Research University, CNRS, LAMSADE, 75016 Paris, France
\textsuperscript{3}Dipartimento di Informatica, Università di Pisa, Italy
\textsuperscript{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 $\bowtie$) For any $\bowtie$-reduced types $T_1$ and $T_2$ and any two $\bowtie$-reduced structural types $S_1$ and $S_2$, the following properties hold:

\begin{align*}
T_1 \bowtie T_2 & \Rightarrow \text{Reduce}(T_1, T_2, \bowtie) \bowtie T_1 \bowtie T_2 \quad (1) \\
S_1 \bowtie S_2 & \Rightarrow \text{Fuse}(S_1, S_2, \bowtie) \bowtie S_1 \bowtie S_2 \quad (2)
\end{align*}

\textbf{Proof.} By mutual induction and by cases on the common kind of $S_1$ and $S_2$. Property (1): here we observe that every addend of $\bowtie$ has one $\bowtie$-equivalent addend in $\bowtie$, by definition of $\bowtie$, and only one, because the two types are $\bowtie$-reduced. Hence, the result has one structural addend for each structural addend of $\bowtie$, and the two addends are $\bowtie$-equivalent by induction. The other interesting case is the record type case of property (2). Here, by definition of $\bowtie$, two record types are only fused when they have exactly the same keys and, for any key $k$ in $\text{Keys}(R_1)$, the types associated to $k$ in $R_1$ and $R_2$ are $\bowtie$-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.

$\blacksquare$
Corollary 1 (Lossless reduction)
For any \( \equiv \)-reduced types \( T_1 \) and \( T_2 \):

\[
Reduce(T_1, T_2, \equiv) \simeq T_1 + T_2
\]

Proof. The reduction process substitutes, inside \( T_1 + T_2 \), two equivalent ad-
dends \( S_1 \equiv S_2 \) with \( Fuse(S_1, S_2, \equiv) \) 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 \( M_1 \) and \( M_2 \), and for any two sets \( F_1, F_2 \) of triples \((k_i, T_i, q_i)\), where
each \( T_i \) is an \( E \)-reduced type, we define the following notation.

\[
\begin{align*}
\mathcal{M}_1 \setminus_E \mathcal{M}_2 & \triangleq \{ S_1 \in \mathcal{M}_1 \mid \forall S_2 \in \mathcal{M}_2. E(S_1, S_2) \} \\
\mathcal{M}_1 \cap_E \mathcal{M}_2 & \triangleq \{ S_1 \in \mathcal{M}_1 \mid \exists S_2 \in \mathcal{M}_2. E(S_1, S_2) \} \\
\mathcal{M}_1 \bowtie_E \mathcal{M}_2 & \triangleq \{ Fuse(S_1, S_2, E) \mid S_1 \in \mathcal{M}_1, S_2 \in \mathcal{M}_2, E(S_1, S_2) \} \\
\mathcal{F}_1 \setminus:: \mathcal{F}_2 & \triangleq \{ (k_1, T_1, q_1) \in \mathcal{F}_1 \mid \exists (k_2, T_2, q_2) \in \mathcal{F}_2. k_1 = k_2 \} \\
\mathcal{F}_1 \cap:: \mathcal{F}_2 & \triangleq \{ (k_1, T_1, q_1) \in \mathcal{F}_1 \mid \exists (k_2, T_2, q_2) \in \mathcal{F}_2. k_1 = k_2 \} \\
?(\mathcal{F}) & \triangleq \{ (k, T, ?) \mid (k, T, q) \in \mathcal{F} \} \\
\mathcal{F}_1 \bowtie:: \mathcal{F}_2 & \triangleq \{ (k_1, \text{Reduce}(T_1, T_2, E), q_1 \cdot q_2) \mid (k_1, T_1, q_1) \in \mathcal{F}_1, (k_1, T_2, q_2) \in \mathcal{F}_2 \}
\end{align*}
\]

These operators allow us to rewrite the definition of \( Reduce \) and \( Fuse \) as
follows.

Lemma 1.2

\[
\begin{align*}
Reduce(T_1, T_2, E) & \triangleq \oplus (\circ_{T_1 \bowtie E} \circ T_2 \cup \circ_{T_1} \setminus_E T_2 \cup \circ T_2 \setminus_E T_1 ) \\
Fuse(R_1, R_2, E) & \triangleq \{ \circ_{R_1 \bowtie::} \circ R_2 \cup ?(\circ_{R_1} \setminus:: \circ R_2) \cup ?(\circ R_2 \setminus:: \circ R_1) \}
\end{align*}
\]
Lemma 1.3 For any SKER $E$, and any two $E$-reduced types $T_1$ and $T_2$, the sets $\circ T_1 \cap_E \circ T_2$, $\circ T_2 \cap_E \circ T_1$, and $\circ T_1 \vartriangleright_E \circ T_2$, are all $E$-distinct, and, for each pair of them, the $E$ relation defines a bijective function between the two.

Proof. The sets $\circ T_1 \cap_E \circ T_2$ and $\circ T_2 \cap_E \circ T_1$ 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 $\circ T_1 \cap_E \circ T_2$ $E$-corresponds to at least one element of $\circ T_2 \cap_E \circ T_1$ by construction, and it cannot $E$-correspond to two of them because, by transitivity, they would be $E$-equivalent, and the type $T_2$ 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 $\circ T_1 \cap_E \circ T_2$ and the following set of pairs, mapping every $S_1$ to the only pair $(S_1, S_2)$:

$$\{ (S_1, S_2) \mid S_1 \in \circ T_1, S_2 \in \circ T_2, E(S_1, S_2) \}$$

To every pair of this set, the element $\text{Fuse}(S_1, S_2, E)$ of $\circ T_1 \vartriangleright_E \circ T_2$ corresponds and vice versa. By stability, $\text{Fuse}(S_1, S_2, E)$ is $E$-equivalent to both $S_1$ and $S_2$, hence we can reason as in the previous case to prove, by transitivity, that no two distinct elements of $\circ T_1 \vartriangleright_E \circ T_2$ may be equivalent, hence it is $E$-reduced, and $E$ is a bijection between it and both of $\circ T_1 \cap_E \circ T_2$ and $\circ T_2 \cap_E \circ T_1$.

Proof of Lemmas 1 and 2 The following properties hold.

1. For any two $E$-reduced types $T_1$, $T_2$, $\text{Reduce}(T_1, T_2, E)$ is $E$-reduced

2. For any two $E$-reduced structural types $S_1$, $S_2$, $\text{Fuse}(S_1, S_2, E)$ is $E$-reduced

3. For any $J$, $S$, $\vdash^E J : S \Rightarrow S$ is $E$-reduced

4. For any $J_1, \ldots, J_n$, $\mathcal{T}$, $\vdash^E J_1, \ldots, J_n : \mathcal{T} \Rightarrow \mathcal{T}$ is $E$-reduced

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

$$\text{Reduce}(T_1, T_2, E) \equiv \oplus \left( \circ T_1 \vartriangleright_E \circ T_2 \cup \circ T_1 \setminus_E \circ T_2 \cup \circ T_2 \setminus_E \circ T_1 \right)$$
The set $\circ T_1 \bowtie_E \circ T_2$ is $E$-reduced by Lemma 1.3, and $\circ T_1 \setminus E \circ T_2$ and $\circ T_2 \setminus E \circ T_1$ are included in $\circ T_1$ and $\circ T_2$, which are $E$-reduced by hypothesis. We have hence just to prove that two structural types coming from two different sets among $\circ T_1 \bowtie_E \circ T_2$, $\circ T_1 \setminus E \circ T_2$ and $\circ T_2 \setminus E \circ T_1$ cannot be $E$-equivalent. If one of them comes from $\circ T_1 \bowtie_E \circ T_2$ and the other from $\circ T_1 \setminus E \circ T_2$, they cannot be equivalent since the first is $E$-isomorphic to $\circ T_1 \cap E \circ T_2$, and elements from $\circ T_1 \setminus E \circ T_2$ cannot be equivalent to any element of $\circ T_2$. The same holds for $\circ T_1 \bowtie_E \circ T_2$ and $\circ T_2 \setminus E \circ T_1$. Finally, no element of $\circ T_1 \setminus E \circ T_2$ may be equivalent to one element of $\circ T_2 \setminus E \circ T_1$ since $\circ T_1 \setminus E \circ T_2$ only contains types that are not equivalent to any element of $\circ T_2$.

Properties (3) and (4) follow immediately, since all the union types that are produced by the judgments for $\vdash^E J : S$ and $\vdash^E J :: T$ are actually produced by a $\text{Reduce}(T_1, T_2, E)$ 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 $T_1$ and $T_2$:

$$T_1 + T_2 \leq \text{Reduce}(T_1, T_2, E)$$

For any two $E$-reduced structural types $S_1$ and $S_2$:

$$E(S_1, S_2) \Rightarrow S_1 + S_2 \leq \text{Fuse}(S_1, S_2, E)$$

**Proof.** By mutual induction.

We want to prove that:

$$T_1 + T_2 \leq \bigoplus(\circ T_1 \bowtie_E \circ T_2 \cup \circ T_1 \setminus E \circ T_2 \cup \circ T_2 \setminus E \circ T_1)$$

That is:

$$\bigoplus(\circ (T_1 + T_2))$$

$$\leq \bigoplus(\circ T_1 \bowtie_E \circ T_2 \cup \circ T_1 \setminus E \circ T_2 \cup \circ T_2 \setminus E \circ T_1)$$

That is:

$$S \in (\circ (T_1 + T_2)) \Rightarrow [S] \subseteq \bigcup_{S' \in (\circ T_1 \bowtie_E \circ T_2 \cup \circ T_1 \setminus E \circ T_2 \cup \circ T_2 \setminus E \circ T_1)} [S']$$
The set ◦(T₁ + T₂) can be decomposed as follows.

\[ ◦(T₁ + T₂) = ( ◦T₁ ∩_E ◦T₂) \cup ( ◦T₁ \setminus_E ◦T₂) \]
\[ \cup ( ◦T₂ ∩_E ◦T₁) \cup ( ◦T₂ \setminus_E ◦T₁) \]

If \( S \in ◦T₁ \cap_E ◦T₂ \), then there exists \( S₂ \in ◦T₂ \) with \( E(S, S₂) \) such that \( Fuse(S, S₂, E) \) belongs to \( ◦T₁ \setminus_E ◦T₂ \), and, by induction, we know that:

\[ E(S, S₂) \Rightarrow \llbracket S \rrbracket \subseteq \llbracket Fuse(S, S₂, E) \rrbracket \]

The case for \( S \in ◦T₂ \cap_E ◦T₁ \) is analogous. The other two cases, \( S \in ◦T₁ \setminus_E ◦T₂ \) and \( S \in ◦T₂ \setminus_E ◦T₁ \), are trivial.

We move now to the proof of

\[ E(S₁, S₂) \Rightarrow S₁ + S₂ \leq Fuse(S₁, S₂, E) \]

by cases on the common kind of \( S₁ \) and \( S₂ \).

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

If they are of array type, then we have \( S₁ = [T₁] \) and \( S₂ = [T₂] \). We want to prove:

\[ \llbracket [T₁] \rrbracket \cup [T₂] \subseteq [ Fuse([T₁], [T₂], E) ] \]

That is,

\[ [T₁] \subseteq [ Reduce([T₁], [T₂], E) ] \]

and

\[ [T₂] \subseteq [ Reduce([T₁], [T₂], E) ] \]

Let us prove the first. Assume that \( \llbracket V₁, \ldots, Vₙ \rrbracket \in \llbracket [T₁] \rrbracket \). This implies that, for any \( i \), we have that \( Vᵢ \in [T₁] \).

By induction, \( [T₁] \subseteq [ Reduce([T₁], [T₂], E) ] \), hence, for any \( i \), we have that \( Vᵢ \in [ Reduce([T₁], [T₂], E) ] \), hence \( \llbracket V₁, \ldots, Vₙ \rrbracket \in [ Reduce([T₁], [T₂], E) ] \).

The inclusion \( [T₂] \subseteq [ Reduce([T₁], [T₂], E) ] \) can be proved in the same way.

The last case is that of record types, that is, \( S₁ = \{ \circ S₁ \} \) and \( S₂ = \{ \circ S₂ \} \).

We want to prove:

\[ \llbracket \{ \circ S₁ \} \rrbracket \cup \llbracket \{ \circ S₂ \} \rrbracket \subseteq [ Fuse(\{ \circ S₁ \}, \{ \circ S₂ \}, E) ] \]

We prove the case for \( S₁ \), the one for \( S₂ \) being analogous.

\[ [ \{ \circ S₁ \} ] \subseteq [ Fuse(\{ \circ S₁ \}, \{ \circ S₂ \}, E) ] \]
We rewrite it as follows:
\[
\llbracket \{ S_1 \} \rrbracket \\
\subseteq \llbracket \{ (\diamond S_1 \subseteq \diamond S_2) \cup ?(\diamond S_1 \setminus \diamond S_2) \cup ?(\diamond S_2 \setminus \diamond S_1) \} \rrbracket 
\]

Consider a record \( V \in \llbracket \{ S_1 \} \rrbracket \). By definition,
\[
V = \{ (k_1, \nu_1), \ldots, (k_n, \nu_n) \}
\]
such that:

1. for any \( i \in 1...n \), \( \exists T_i, q_i \) such that \( (k_i, T_i, q_i) \) belongs to \( \diamond S_1 \), and \( \nu_i \in \llbracket T_i \rrbracket \)

2. for any \((k_j, T_j, !)\) \( \in \diamond S_1 \), a pair \((k_j, \nu_j)\) is in \( V \).

We want to prove the same properties for \( V \) with respect to
\[
\{ (\diamond S_1 \subseteq \diamond S_2) \cup ?(\diamond S_1 \setminus \diamond S_2) \cup ?(\diamond S_2 \setminus \diamond S_1) \}
\]

We first prove the first property. Assume that the pair \((k_i, \nu_i)\) belongs to \( V \). By (1) above, we have a triple \((k_i, T_i, q_i)\) in \( \diamond S_1 \) with \( \nu_i \in \llbracket T_i \rrbracket \). If a matching \( k \) exists in \( S_2 \), then we have a triple \((k_i, \text{Reduce}(T_i, T_2, E), k)\) in \( \diamond S_1 \subseteq \diamond S_2 \). By induction, \( \llbracket T_i \rrbracket \subseteq \llbracket \text{Reduce}(T_i, T_2, E) \rrbracket \), hence \( \nu_i \in \llbracket \text{Reduce}(T_i, T_2, E) \rrbracket \), as required. If no matching \( k \) exists in \( S_2 \), then we have a triple \((k_i, T_i, !)\) in \( \diamond S_1 \setminus \diamond S_2 \), and \( \nu_i \in \llbracket T_i \rrbracket \) holds by hypothesis.

For the second property, every triple \((k_j, T_j, !)\) in
\[
(\diamond S_1 \subseteq \diamond S_2) \cup ?(\diamond S_1 \setminus \diamond S_2) \cup ?(\diamond S_2 \setminus \diamond S_1)
\]
comes from the \( \diamond S_1 \subseteq \diamond S_2 \) component and, by definition of \( q_1 \cdot q_2 \), it corresponds to a triple \((k_j, \ldots, !)\) in \( \diamond S_1 \), hence \( V \) contains a field with the key \( k_j \) by hypothesis.

We can now prove that the \( \text{Reduce}(T_1, T_2, E) \) operator enjoys the commutativity and associativity properties that enable an efficient distributed map-reduce implementation.

**Theorem 4 (Commutativity)**

1. Given two \( E \)-reduced types \( T_1, T_2 \), we have:
\[
\text{Reduce}(T_1, T_2, E) \equiv \text{Reduce}(T_2, T_1, E)
\]
2. Given two structural E-reduced types $S_1$ and $S_2$ we have:

$$E(S_1, S_2) \Rightarrow Fuse(S_1, S_2, E) = Fuse(S_2, S_1, E)$$

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 $M_1, M_2, M$, and for any sets $F_1, F_2, F$ of triples $(k, T_i, q_i)$, where each $T_i$ is an E-reduced type, the following equalities hold.

$$\begin{align*}
(M_1 \cup M_2) \bowtie_E M &= (M_1 \bowtie_E M) \cup (M_2 \bowtie_E M) \\
(F_1 \cup F_2) \bowtie_:: F &= (F_1 \bowtie_:: F) \cup (F_2 \bowtie_:: F) \\
M \bowtie_E (M_1 \cup M_2) &= (M \bowtie_E M_1) \cup (M \bowtie_E M_2) \\
F \bowtie_:: (F_1 \cup F_2) &= (F \bowtie_:: F_1) \cup (F \bowtie_:: F_2)
\end{align*}$$

Proof. By definition of $\bowtie_E$:

$$\begin{align*}
(M_1 \cup M_2) \bowtie_E M \\
= \{ \text{Fuse}(S, S', E) \mid S \in M_1 \cup M_2, \ S' \in M, E(S, S') \} \\
= \{ \text{Fuse}(S, S', E) \mid S \in M_1, \ S' \in M, E(S, S') \} \\
\cup \{ \text{Fuse}(S, S', E) \mid S \in M_2, \ S' \in M, E(S, S') \} \\
= (M_1 \bowtie_E M) \cup (M_2 \bowtie_E M)
\end{align*}$$

By definition of $\bowtie_::$:

$$\begin{align*}
(F_1 \cup F_2) \bowtie_:: F \\
= \{ (k, \text{Reduce}(T, T', E), q \cdot q') \\
\mid (k, T, q) \in (F_1 \cup F_2), (k, T', q') \in F \} \\
= \{ (k, \text{Reduce}(T, T', E), q \cdot q') \\
\mid (k, T, q) \in F_1, (k, T', q') \in F \} \\
\cup \{ (k, \text{Reduce}(T, T', E), q \cdot q') \\
\mid (k, T, q) \in F_2, (k, T', q') \in F \} \\
= (F_1 \bowtie_:: F) \cup (F_2 \bowtie_:: F)
\end{align*}$$

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 \( T_1, T_2 \) and \( T_3 \), we have
\[
\text{Reduce}(\text{Reduce}(T_1, T_2, E), T_3, E) = \text{Reduce}(T_1, \text{Reduce}(T_2, T_3, E), E)
\]

2. Given three \( E \)-reduced structural types \( S_1, S_2 \) and \( S_3 \) that are mutually \( E \)-equivalent, we have
\[
\text{Fuse}(\text{Fuse}(S_1, S_2, E), S_3, E) = \text{Fuse}(S_1, \text{Fuse}(S_2, S_3, E), E)
\]

\textbf{Proof.} We proof (1) and (2) by mutual induction.

We first partition each of \( \circ T_1, \circ T_2 \) and \( \circ T_3 \) in four parts, that correspond to four possible combinations of \( \bigcap_E \) and \( \bigsetminus_E \), as follows.

\[
\begin{align*}
M_1^{23} & = \{ S_1 \in \circ T_1 \mid \exists S_2 \in \circ T_2, E(S_1, S_2), \\
& \quad \exists S_3 \in \circ T_3, E(S_1, S_3) \} \\
M_1^{22} & = \{ S_1 \in \circ T_1 \mid \exists S_2 \in \circ T_2, E(S_1, S_2), \\
& \quad \exists S_3 \in \circ T_3, E(S_1, S_3) \} \\
M_1^{21} & = \{ S_1 \in \circ T_1 \mid \exists S_2 \in \circ T_2, E(S_1, S_2), \\
& \quad \exists S_3 \in \circ T_3, E(S_1, S_3) \} \\
M_1^{12} & = \{ S_1 \in \circ T_1 \mid \exists S_2 \in \circ T_2, E(S_1, S_2), \\
& \quad \exists S_3 \in \circ T_3, E(S_1, S_3) \}
\end{align*}
\]

The partitions \( \{ M_1^{13}, M_2^{12}, M_1^{23}, M_2^{14} \} \) of \( \circ T_2 \) and \( \{ M_1^{32}, M_3^{12}, M_3^{24}, M_3^{14} \} \) of \( \circ T_3 \) are defined in the same way. Now we can decompose \( \circ \text{Reduce}(T_1, T_2, E) \) as follows. In all of our computations we will make use of distributivity of join over set union (Lemma 1.4).

\[
\circ \text{Reduce}(T_1, T_2, E) = ((M_1^{23} \cup M_1^{22}) \bowtie_E (M_2^{13} \cup M_2^{14})) \\
\quad \cup (M_1^{21} \cup M_2^{21}) \cup M_2^{13} \cup M_2^{14} \\
\quad = ((M_1^{23} \bowtie_E M_2^{13}) \cup (M_1^{22} \bowtie_E M_2^{14})) \\
\quad \cup (M_1^{21} \cup M_2^{21}) \cup M_2^{13} \cup M_2^{14}
\]

Now we compute \( \circ \text{Reduce}(\text{Reduce}(T_1, T_2, E), T_3, E) \). The first two lines join the components of \( \circ \text{Reduce}(T_1, T_2, E) \) that match some component of \( \circ T_3 \) with the corresponding component of \( \circ T_3 \), while the last line lists all the non-matching components of \( \circ \text{Reduce}(T_1, T_2, E) \) and \( \circ T_3 \).

\[
\circ \text{Reduce}(\text{Reduce}(T_1, T_2, E), T_3, E) = \\
\quad ((M_1^{23} \bowtie_E M_2^{13}) \bowtie_E M_3^{12}) \\
\quad \cup (M_1^{23} \bowtie_E M_3^{12}) \cup (M_2^{13} \bowtie_E M_3^{12}) \\
\quad \cup (M_1^{22} \bowtie_E M_1^{13}) \cup M_1^{22} \cup M_2^{12} \cup M_3^{22}
\]

8
By reordering the components, we have the following equation for $\circ Reduce(Reduce(T_1, T_2, E), T_3, E)$:

$$
\circ Reduce(Reduce(T_1, T_2, E), T_3, E) =
\left((M_1^{23} \bowtie_E M_2^{13}) \bowtie_E M_3^{12}\right)
\cup(M_1^{24} \bowtie_E M_2^{14})
\cup M_1^{12} \cup M_2^{14} \cup M_3^{12}
$$

The same computation for $\circ Reduce(T_1, Reduce(T_2, T_3, E), E)$ yields the same result with the only exception of the first term.

$$
\circ Reduce(T_1, Reduce(T_2, T_3, E), E) =
(M_1^{23} \bowtie_E (M_2^{13} \bowtie_E M_3^{12}))
\cup(M_1^{24} \bowtie_E M_2^{14})
\cup M_1^{12} \cup M_2^{14} \cup M_3^{12}
$$

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.

$$
\begin{align*}
((M_1^{23} \bowtie_E M_2^{13}) &\bowtie_E M_3^{12}) \\
= \| \text{Fuse}(S_1, S_2, E) \\
&| S_1 \in M_1^{23}, S_2 \in M_2^{13}, E(S_1, S_2) \| \bowtie_E M_3^{12} \\
= \| \text{Fuse}(\text{Fuse}(S_1, S_2, E), S_3, E) \\
&| S_1 \in M_1^{23}, S_2 \in M_2^{13}, S_3 \in M_3^{12}, \\
&\quad E(S_1, S_2), E(\text{Fuse}(S_1, S_2, E), S_3) \| \\
(M_1^{23} \bowtie_E (M_2^{13} \bowtie_E M_3^{12})) \\
= \| \text{Fuse}(S_1, \text{Fuse}(S_2, S_3, E), E) \\
&| S_1 \in M_1^{23}, S_2 \in M_2^{13}, S_3 \in M_3^{12}, \\
&\quad E(S_2, S_3), E(S_1, \text{Fuse}(S_2, S_3, E)) \|
\end{align*}
$$

By stability, both

$$
E(S_1, S_2) \land E(\text{Fuse}(S_1, S_2, E), S_3)
$$

and

$$
E(S_2, S_3) \land E(S_1, \text{Fuse}(S_2, S_3, E))
$$

can be rewritten as

$$
E(S_1, S_2) \land E(S_2, S_3),
$$
while \( Fuse(Fuse(S_1, S_2, E), S_3, E) \) is equivalent to
\[
Fuse(S_1, Fuse(S_2, S_3, E), E)
\]
by induction, hence we conclude.

(2) Observe that \( S_1, S_2, \) and \( S_3 \) 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 \( S_1 = [T_1], S_2 = [T_2], \) and \( S_3 = [T_3], \)
for some \( T_1, T_2, \) and \( T_3, \) and we have:
\[
\begin{align*}
Fuse(Fuse([T_1], [T_2], E), [T_3], E) &=Fuse([Reduce(T_1, T_2, E)], [T_3], E) \\
&= [Reduce(Reduce(T_1, T_2, E), T_3, E)] \\
Fuse([T_1], Fuse([T_2], [T_3], E), E) &= Fuse([T_1], [Reduce(T_2, T_3, E)], E) \\
&= [Reduce(T_1, Reduce(T_2, T_3, E), E)]
\end{align*}
\]
The thesis follows by case (1) and mutual induction.

The last case is that of record types, that is, \( S_1 = \{ \omega S_1 \}, S_2 = \{ \omega S_2 \}, \)
and \( S_3 = \{ \omega S_3 \} \).

We will follow the same structure as in the proof of the first case, that of \( Reduce(Reduce(T_1, T_2, E), T_3, E). \)

As in the first case, we partition \( \omega S_1 \) in four parts \( F_1^{23}, F_1^{24}, F_1^{23}, F_1^{24}, \)
according to the existence of a matching field in \( \omega S_2 \) and of a matching field in \( \omega S_3 \).
\[
\begin{align*}
F_1^{23} &= (\omega S_1 \cap_\omega \omega S_2) \cap_\omega \omega S_3 \\
F_1^{24} &= (\omega S_1 \cap_\omega \omega S_2) \setminus_\omega \omega S_3 \\
F_1^{23} &= (\omega S_1 \setminus_\omega \omega S_2) \cap_\omega \omega S_3 \\
F_1^{24} &= (\omega S_1 \setminus_\omega \omega S_2) \setminus_\omega \omega S_3
\end{align*}
\]

Now we can decompose \( \omega Fuse(S_1, S_2, E) \) as follows.
\[
\begin{align*}
\omega Fuse(S_1, S_2, E) &= (\omega (M_1^{23} \cup M_1^{24}) \bowtie_E (M_1^{23} \cup M_1^{24})) \\
&\cup (M_1^{23} \setminus_\omega (M_1^{23} \cup M_1^{24}) \cup M_1^{24}) \\
&\cup (M_1^{23} \setminus_\omega (M_1^{23} \cup M_1^{24}) \cup M_1^{24}) \\
&\cup (M_1^{23} \setminus_\omega (M_1^{23} \cup M_1^{24}) \cup M_1^{24})
\end{align*}
\]
Now we compute \( \omega Fuse(Fuse(S_1, S_2, E), S_3, E) \). The first two lines join the components of \( \omega Fuse(S_1, S_2, E) \) that match some component of \( \omega S_3 \) with
the corresponding component of $\diamondsuit S_3$, while the last line lists all the non-
matching components of $\diamondsuit Fuse(S_1, S_2, E)$ and $\diamondsuit S_3$.

$$\diamondsuit Fuse(Fuse(S_1, S_2, E), S_3, E) =$$

$$= ((F_{1}^{23} \bowtie_{\cdot} F_{2}^{13}) \bowtie_{\cdot} F_{3}^{12})$$

$$\cup (F_{1}^{23} \bowtie_{\cdot} F_{2}^{13}) \cup (F_{2}^{12} \bowtie_{\cdot} F_{3}^{12})$$

$$\cup (F_{1}^{23} \bowtie_{\cdot} F_{2}^{13}) \cup F_{2}^{12} \cup F_{2}^{12} \cup F_{3}^{12}$$

By reordering the components, we have the following equation for $\diamondsuit Fuse(Fuse(S_1, S_2, E), S_3, E)$.

$$\diamondsuit Fuse(Fuse(S_1, S_2, E), S_3, E) =$$

$$= ((F_{1}^{23} \bowtie_{\cdot} F_{2}^{13}) \bowtie_{\cdot} F_{3}^{12})$$

$$\cup (F_{1}^{23} \bowtie_{\cdot} F_{2}^{13}) \cup (F_{1}^{23} \bowtie_{\cdot} F_{3}^{12}) \cup (F_{1}^{23} \bowtie_{\cdot} F_{3}^{12})$$

$$\cup F_{2}^{12} \cup F_{2}^{12} \cup F_{3}^{12}$$

The same computation for $\diamondsuit Fuse(S_1, Fuse(S_2, S_3, E), E)$ yields the same re-

$$\diamondsuit Fuse(S_1, Fuse(S_2, S_3, E), E) =$$

$$= (F_{1}^{23} \bowtie_{\cdot} (F_{2}^{13} \bowtie_{\cdot} F_{3}^{12}))$$

$$\cup (F_{1}^{23} \bowtie_{\cdot} (F_{1}^{23} \bowtie_{\cdot} F_{3}^{12}) \cup (F_{1}^{23} \bowtie_{\cdot} F_{3}^{12})$$

$$\cup F_{1}^{23} \cup F_{2}^{12} \cup F_{3}^{12}$$

Hence, we only have to prove that

$$((F_{1}^{23} \bowtie_{\cdot} F_{2}^{13}) \bowtie_{\cdot} F_{3}^{12}) = (F_{1}^{23} \bowtie_{\cdot} (F_{2}^{13} \bowtie_{\cdot} F_{3}^{12}))$$

By definition, we have the following equalities.

$$((F_{1}^{23} \bowtie_{\cdot} F_{2}^{13}) \bowtie_{\cdot} F_{3}^{12})$$

$$= \{ (k, Reduce(T_1, T_2, E), q_1 \cdot q_2)$$

$$| (k, T_1, q_1) \in F_{2}^{13}, (k, T_2, q_2) \in F_{2}^{13} \} \bowtie_{\cdot} F_{3}^{12}$$

$$= \{ (k, Reduce(\text{Reduce}(T_1, T_2, E), T_3, E), (q_1 \cdot q_2) \cdot q_3)$$

$$| (k, T_1, q_1) \in F_{2}^{13}, (k, T_2, q_2) \in F_{2}^{13},$$

$$\quad (k, T_3, q_3) \in F_{3}^{12} \}$$

$$(F_{1}^{23} \bowtie_{\cdot} (F_{2}^{13} \bowtie_{\cdot} F_{3}^{12}))$$

$$= \{ (k, Reduce(T_1, \text{Reduce}(T_2, T_3, E), E), q_1 \cdot (q_2 \cdot q_3))$$

$$| (k, T_1, q_1) \in F_{2}^{13}, (k, T_2, q_2) \in F_{2}^{13},$$

$$\quad (k, T_3, q_3) \in F_{3}^{12} \}$$

By induction $\text{Reduce}(\text{Reduce}(T_1, T_2, E), T_3, E)$ is equivalent to $\text{Reduce}(T_1, \text{Reduce}(T_2, T_3, E), E)$,

associativity of $q' \cdot q''$ is immediate, hence we conclude.
Theorem 5

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

$$\vdash^E J : S \Rightarrow [J] \in [S]$$
$$\vdash^E J_1, \ldots, J_n : ^c T \Rightarrow \{[J_1], \ldots, [J_n]\} \subseteq [T]$$

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

$$\{[J_1], \ldots, [J_i]\} \subseteq [T_1]$$
$$\{[J_{i+1}], \ldots, [J_n]\} \subseteq [T_2]$$

By Theorem 2,

$$T_1 \leq Reduce(T_1, T_2, E) \text{ and } T_2 \leq Reduce(T_1, T_2, E)$$

Hence, by transitivity, we have that

$$\{[J_1], \ldots, [J_i]\} \subseteq [Reduce(T_1, T_2, E)]$$
$$\{[J_{i+1}], \ldots, [J_n]\} \subseteq [Reduce(T_1, T_2, E)]$$

hence

$$\{[J_1], \ldots, [J_n]\} \subseteq [Reduce(T_1, T_2, E)].$$

$\blacksquare$