

GADTs Meet Their Match:

Pattern-Matching Warnings That Account for GADTs, Guards, and Laziness

SEBASTIAN GRAF, Karlsruhe Institute of Technology, Germany

SIMON PEYTON JONES, Microsoft Research, UK

Authors' addresses: Sebastian Graf, Karlsruhe Institute of Technology, Karlsruhe, Germany, sebastian.graf@kit.edu; Simon Peyton Jones, Microsoft Research, Cambridge, UK, simonpj@microsoft.com.

Guard Syntax		
$K \in$	Con	$n \in \mathbb{N}$
$x, y, a, b \in$	Var	$\gamma \in \text{TyCt} ::= \tau_1 \sim \tau_2 \mid \dots$
$\tau, \sigma \in$	Type	$p \in \text{Pat} ::= _$
$e \in$	Expr	$::= x$
		$::= K \bar{\tau} \bar{\sigma} \bar{\gamma} \bar{e}$
		$::= \dots$
		$g \in \text{Grd} ::= \text{let } x : \tau = e$
		$::= K \bar{a} \bar{\gamma} \bar{y} : \bar{\tau} \leftarrow x$
		$::= !x$
Constraint Formula Syntax		
$\Gamma ::=$	$\emptyset \mid \Gamma, x : \tau \mid \Gamma, a$	Context
$\delta ::=$	$\checkmark \mid \times \mid K \bar{a} \bar{\gamma} \bar{y} : \bar{\tau} \leftarrow x \mid x \not\approx K \mid x \approx \perp \mid x \not\approx \perp \mid x \approx e$	Constraint Literals
$\Delta ::=$	$\delta \mid \Delta \wedge \Delta \mid \Delta \vee \Delta$	Formula
$\varphi ::=$	$\gamma \mid x \approx K \bar{a} \bar{\gamma} \mid x \not\approx K \mid x \approx \perp \mid x \not\approx \perp$	Simple constraints without scoping
$\Phi ::=$	$\emptyset \mid \Phi, \varphi$	Set of simple constraints
$\nabla ::=$	$\Gamma \triangleright \Phi$	Inert Set
Clause Tree Syntax		
$t_G, u_G \in$	Gdt	$::= \text{Rhs } n \mid t_G; u_G \mid \text{Guard } g \ t_G$
$t_A, u_A \in$	Ant	$::= \text{AccessibleRhs } n \mid \text{InaccessibleRhs } n \mid t_A; u_A \mid \text{MayDiverge } t_A$

Fig. 1. Syntax

1 END TO END EXAMPLE

We'll start from the following source Haskell program and see how each of the steps (translation to guard trees, checking guard trees and ultimately generating inhabitants of the occurring Δ s) work.

```
f :: Maybe Int -> Int
f Nothing = 0 -- RHS 1
f x | Just y <- x = y -- RHS 2
```

1.1 Translation to guard trees

The program (by a function we probably only give in the appendix?) corresponds to the following guard tree t_f :

```
Guard (!x) Guard (Nothing ← x) Rhs 1;
Guard (!x) Guard (Just y ← x) Rhs 2
```

Data constructor matches are strict, so we add a bang for each match.

1.2 Checking

1.2.1 Uncovered values. First compute the uncovered Δ s, after the first and the second clause respectively.

(1)

$$\begin{aligned} \Delta_1 &::= \mathcal{U}(\text{Guard } (!x) \text{ Guard } (\text{Nothing} \leftarrow x) \text{ Rhs } 1) \\ &= x \not\approx \perp \wedge (x \not\approx \text{Nothing} \vee \times) \end{aligned}$$

(2)

$$\Delta_2 ::= \mathcal{U}(t_f) = \Delta_1 \wedge x \not\approx \perp \wedge (x \not\approx \text{Just} \vee \times)$$

Checking Guard Trees

$$\boxed{\mathcal{U}(t_G) = \Delta}$$

$$\begin{aligned} \mathcal{U}(\text{Rhs } n) &= \times \\ \mathcal{U}(t; u) &= \mathcal{U}(t) \wedge \mathcal{U}(u) \\ \mathcal{U}(\text{Guard } (!x) \ t) &= (x \neq \perp) \wedge \mathcal{U}(t) \\ \mathcal{U}(\text{Guard } (\text{let } x = e) \ t) &= (x \approx e) \wedge \mathcal{U}(t) \\ \mathcal{U}(\text{Guard } (K \ \bar{a} \ \bar{y} \ \bar{y} : \bar{\tau} \leftarrow x) \ t) &= (x \neq K) \vee ((K \ \bar{a} \ \bar{y} \ \bar{y} : \bar{\tau} \leftarrow x) \wedge \mathcal{U}(t)) \end{aligned}$$

$$\boxed{\mathcal{A}_\Gamma(\Delta, t_G) = t_A}$$

$$\begin{aligned} \mathcal{A}_\Gamma(\Delta, \text{Rhs } n) &= \begin{cases} \text{InaccessibleRhs } n, & \mathcal{G}(\Gamma, \Delta) = \emptyset \\ \text{AccessibleRhs } n, & \text{otherwise} \end{cases} \\ \mathcal{A}_\Gamma(\Delta, (t; u)) &= \mathcal{A}_\Gamma(\Delta, t); \mathcal{A}_\Gamma(\Delta \wedge \mathcal{U}(t), u) \\ \mathcal{A}_\Gamma(\Delta, \text{Guard } (!x) \ t) &= \begin{cases} \mathcal{A}_\Gamma(\Delta \wedge (x \neq \perp), t), & \mathcal{G}(\Gamma, \Delta \wedge (x \approx \perp)) = \emptyset \\ \text{MayDiverge } \mathcal{A}_\Gamma(\Delta \wedge (x \neq \perp), t) & \text{otherwise} \end{cases} \\ \mathcal{A}_\Gamma(\Delta, \text{Guard } (\text{let } x = e) \ t) &= \mathcal{A}_\Gamma(\Delta \wedge (x \approx e), t) \\ \mathcal{A}_\Gamma(\Delta, \text{Guard } (K \ \bar{a} \ \bar{y} \ \bar{y} : \bar{\tau} \leftarrow x) \ t) &= \mathcal{A}_\Gamma(\Delta \wedge (K \ \bar{a} \ \bar{y} \ \bar{y} : \bar{\tau} \leftarrow x), t) \end{aligned}$$

Putting it all together

- (0) Input: Context with match vars Γ and desugared Gdt t
 - (1) Report n pattern vectors of $\mathcal{G}(\Gamma, \mathcal{U}(t))$ as uncovered
 - (2) Report the collected redundant and not-redundant-but-inaccessible clauses in $\mathcal{A}_\Gamma(\checkmark, t)$
- (TODO: Write a function that collects the RHSs).

Fig. 2. Pattern-match checking

The right operands of \vee are vacuous, but the purely syntactical transformation doesn't see that.

We can see that Δ_2 is in fact uninhabited, because the three constraints $x \neq \perp$, $x \neq \text{Nothing}$ and $x \neq \text{Just}$ cover all possible data constructors of the Maybe data type. And indeed $\mathcal{G}(x : \text{Maybe Int}, \Delta_2) = \emptyset$, as we'll see later.

1.2.2 Redundancy. In order to compute the annotated clause tree $\mathcal{A}_\Gamma(\checkmark, t_f)$, we need to perform the following four inhabitation checks, one for each bang (for knowing whether we need to wrap a MayDiverge and one for each RHS (where we have to decide for InaccessibleRhs or AccessibleRhs):

- (1) The first divergence check: $\Delta_3 := \checkmark \wedge x \approx \perp$
- (2) Upon reaching the first RHS: $\Delta_4 := \checkmark \wedge x \neq \perp \wedge \text{Nothing} \leftarrow x$
- (3) The second divergence check: $\Delta_5 := \checkmark \wedge \Delta_1 \wedge x \approx \perp$
- (4) Upon reaching the second RHS: $\Delta_6 := \checkmark \wedge \Delta_1 \wedge x \neq \perp \wedge \text{Just } y \leftarrow x$

Except for Δ_5 , these are all inhabited, i.e. $\mathcal{G}(x : \text{Maybe Int}, \Delta_i) \neq \emptyset$ (as we'll see in the next section).

Thus, we will get the following annotated tree:

MayDiverge AccessibleRhs 1; AccessibleRhs 2

1.3 Generating inhabitants

The last section left open how $\mathcal{G}(\cdot)$ works, which was used to establish or refute vacuity of a Δ .

Generate inhabitants of Δ

$$\boxed{\mathcal{G}(\Gamma, \Delta) = \mathcal{P}(\bar{p})}$$

$$\mathcal{G}(\Gamma, \Delta) = \bigcup \{ \mathcal{E}(\nabla, \text{dom}(\Gamma)) \mid \nabla \in C(\Gamma \triangleright \emptyset, \Delta) \}$$

Construct inhabited ∇ s from Δ

$$\boxed{C(\nabla, \Delta) = \mathcal{P}(\nabla)}$$

$$\begin{aligned} C(\nabla, \delta) &= \begin{cases} \{\nabla'\} & \text{where } \nabla' = \nabla \oplus \delta \\ \emptyset & \text{otherwise} \end{cases} \\ C(\nabla, \Delta_1 \wedge \Delta_2) &= \bigcup \{ C(\nabla', \Delta_2) \mid \nabla' \in C(\nabla, \Delta_1) \} \\ C(\nabla, \Delta_1 \vee \Delta_2) &= C(\nabla, \Delta_1) \cup C(\nabla, \Delta_2) \end{aligned}$$

Convert Δ to ∇ (without checking consistency)

$$\boxed{\mathcal{T}(\Gamma, \Delta) = \mathcal{P}(\bar{p})}$$

$$\mathcal{G}(\Gamma, \Delta) = \bigcup \{ \mathcal{E}(\nabla, \text{dom}(\Gamma)) \mid \nabla \in C(\Gamma \triangleright \emptyset, \Delta) \}$$

Expand variables to Pat with ∇

$$\boxed{\mathcal{E}(\nabla, \bar{x}) = \mathcal{P}(\bar{p})}$$

$$\begin{aligned} \mathcal{E}(\nabla, \epsilon) &= \{ \epsilon \} \\ \mathcal{E}(\Gamma \triangleright \Phi, x_1 \dots x_n) &= \begin{cases} \{ (K \ q_1 \dots q_m) p_2 \dots p_n \mid (q_1 \dots q_m \ p_2 \dots p_n) \in \mathcal{E}(\Gamma \triangleright \Phi, y_1 \dots y_m \ x_2 \dots x_n) \} & \text{if } \Phi(x) \approx K \ \bar{a} \ \bar{y} \in \\ \{ _ \ p_2 \dots p_n \mid (p_2 \dots p_n) \in \mathcal{E}(\Gamma \triangleright \Phi, x_2 \dots x_n) \} & \text{otherwise} \end{cases} \end{aligned}$$

Finding the representative of a variable in Φ

$$\boxed{\Phi(x) = y}$$

$$\Phi(x) = \begin{cases} \Phi(y) & x \approx y \in \Phi \\ x & \text{otherwise} \end{cases}$$

Fig. 3. Bridging between the facade Δ and ∇

$\mathcal{G}(\cdot)$ proceeds in two steps: First it constructs zero, one or many *inert sets* ∇ with $C(\cdot, \cdot)$ (each of them representing a set of mutually compatible constraints) and then expands each of the returned inert sets into one or more pattern vectors \bar{p} with $\mathcal{E}(\cdot, \cdot)$, which is the preferred representation to show to the user.

The interesting bit happens in $C(\cdot, \cdot)$, where a Δ is basically transformed into disjunctive normal form, represented by a set of independently inhabited ∇ . This happens by gradually adding individual constraints to the incoming inert set with \oplus , which starts out empty in $\mathcal{G}(\cdot, \cdot)$. Conjunction is handled by performing the equivalent of a `concatMap`, whereas disjunction simply translates to set union.

Let's see how that works for Δ_3 above. Recall that $\Gamma = x : \text{Maybe Int}$ and $\Delta_3 = \checkmark \wedge x \approx \perp$:

Add a constraint to the inert set	
$\nabla \oplus \delta = \nabla$	
$\nabla \oplus \times$	$= \perp$
$\nabla \oplus \checkmark$	$= \nabla$
$\Gamma \triangleright \Phi \oplus \gamma$	$= \begin{cases} \Gamma \triangleright (\Phi, \gamma) & \text{if type checker deems } \gamma \text{ compatible with } \Phi \\ & \text{and } \forall x \in \text{dom}(\Gamma) : \Gamma \triangleright (\Phi, \gamma) \vdash \Phi(x) \\ \perp & \text{otherwise} \end{cases}$
$\Gamma \triangleright \Phi \oplus K \bar{a} \bar{\gamma} \bar{y} : \bar{\tau} \leftarrow x$	$= \begin{cases} \Gamma, \bar{a}, \bar{y} : \bar{\tau} \triangleright \Phi \oplus \bar{a} \sim \bar{b} \oplus \bar{\gamma} \oplus \bar{y} \approx \bar{z} & \text{if } \Phi(x) \approx K \bar{b} \bar{z} \in \Phi \\ \Gamma' \triangleright (\Phi', \Phi(x) \approx K \bar{a} \bar{y}) & \text{where } \Gamma' \triangleright \Phi' = \Gamma, \bar{a}, \bar{y} : \bar{\tau} \triangleright \Phi \oplus \bar{\gamma} \\ & \text{and } \Phi(x) \approx K \notin \Phi \\ & \text{and } \Gamma' \triangleright \Phi' \vdash \bar{y} \\ \perp & \text{otherwise} \end{cases}$
$\Gamma \triangleright \Phi \oplus x \approx K$	$= \begin{cases} \perp & \text{if } \Phi(x) \approx K \bar{a} \bar{y} \in \Phi \\ \perp & \text{if } \Phi(x) : \tau \in \Gamma \\ & \text{and } \forall K' \in \text{Cons}(\Gamma \triangleright \Phi, \tau) : \Phi(x) \approx K' \in (\Phi, \Phi(x) \approx K) \\ \perp & \text{if not } \Gamma \triangleright (\Phi, \Phi(x) \approx K) \vdash \Phi(x) \\ \Gamma \triangleright (\Phi, \Phi(x) \approx K) & \text{otherwise} \end{cases}$
$\Gamma \triangleright \Phi \oplus x \approx \perp$	$= \begin{cases} \perp & \text{if } \Phi(x) \approx \perp \in \Phi \\ \Gamma \triangleright (\Phi, \Phi(x) \approx \perp) & \text{otherwise} \end{cases}$
$\Gamma \triangleright \Phi \oplus x \approx \perp$	$= \begin{cases} \perp & \text{if } \Phi(x) \approx \perp \in \Phi \\ \perp & \text{if not } \Gamma \triangleright (\Phi, \Phi(x) \approx \perp) \vdash \Phi(x) \\ \Gamma \triangleright (\Phi, \Phi(x) \approx \perp) & \text{otherwise} \end{cases}$
$\Gamma \triangleright \Phi \oplus x \approx y$	$= \begin{cases} \Gamma \triangleright \Phi & \text{if } \Phi(x) = \Phi(y) \\ \Gamma \triangleright (\Phi, \Phi(x) \approx \Phi(y)) \oplus ((\Phi \cap \Phi(x))[\Phi(y)/\Phi(x)]) & \text{otherwise} \end{cases}$
$\Gamma \triangleright \Phi \oplus x \approx K \bar{\tau}' \bar{\tau} \bar{\gamma} \bar{e}$	$= \Gamma, \bar{a}, \bar{y} : \bar{\sigma} \triangleright \Phi \oplus K \bar{a} \bar{\gamma} \bar{y} \leftarrow x \oplus \bar{a} \sim \bar{\tau} \oplus \bar{y} \approx \bar{e} \text{ where } \bar{a} \# \Gamma, \bar{y} \# \Gamma, \bar{e} : \bar{\sigma}$
$\nabla \oplus x \approx e$	$= \nabla$
$\Phi \cap x = \Phi$	
$\emptyset \cap x$	$= \emptyset$
$(\Phi, x \approx K \bar{a} \bar{y}) \cap x$	$= (\Phi \cap x), x \approx K \bar{a} \bar{y}$
$(\Phi, x \approx K) \cap x$	$= (\Phi \cap x), x \approx K$
$(\Phi, x \approx \perp) \cap x$	$= (\Phi \cap x), x \approx \perp$
$(\Phi, x \approx \perp) \cap x$	$= (\Phi \cap x), x \approx \perp$
$(\Phi, \varphi) \cap x$	$= \Phi \cap x$

Fig. 4. Adding a constraint to the inert set ∇

$$\begin{aligned}
& C(\Gamma, \checkmark \wedge x \approx \perp) \\
&= \{ \text{Conjunction} \} \\
& \cup \{ C(\Gamma' \triangleright \nabla', x \approx \perp) \mid \Gamma' \triangleright \nabla' \in C(\Gamma \triangleright \emptyset, \checkmark) \} \\
&= \{ \text{Single constraint} \} \\
& \begin{cases} C(\Gamma' \triangleright \nabla', x \approx \perp) & \text{where } \Gamma' \triangleright \nabla' = \Gamma \triangleright \emptyset \oplus \checkmark \\ \emptyset & \text{otherwise} \end{cases} \\
&= \{ \checkmark \text{ case of } \oplus \} \\
& C(\Gamma \triangleright \emptyset, x \approx \perp) \\
&= \{ \text{Single constraint} \} \\
& \begin{cases} \{ \Gamma' \triangleright \nabla' \} & \text{where } \Gamma' \triangleright \nabla' = \Gamma \triangleright \emptyset \oplus x \approx \perp \end{cases}
\end{aligned}$$

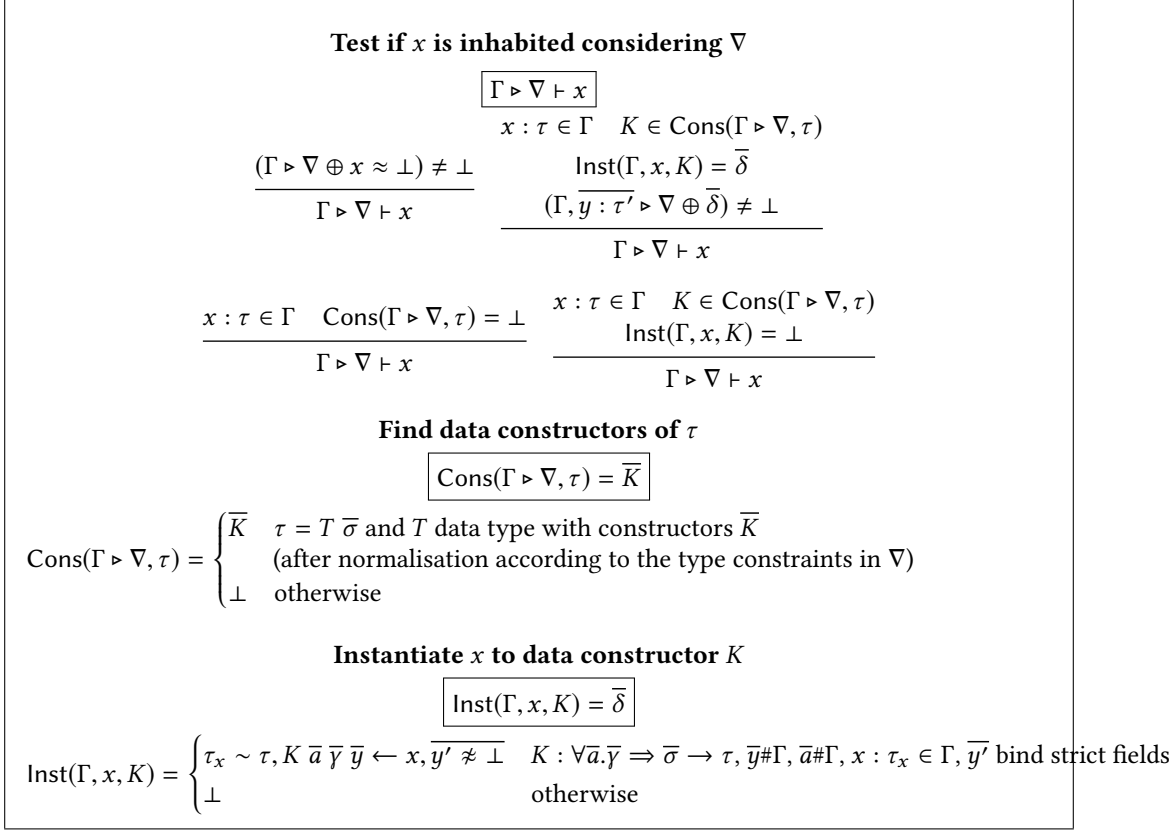


Fig. 5. Inhabitation test

Let's start with $\mathcal{G}(\Gamma, \Delta_3)$, where $\Gamma = x : \text{Maybe Int}$ and recall that $\Delta_3 = \checkmark \wedge x \approx \perp$. The first constraint \checkmark is added very easily to the initial nabla by discarding it, the second one ($x \approx \perp$) is not conflicting with any $x \not\approx \perp$ constraint in the incoming, still empty (\emptyset) nabla, so we end up with $\Gamma \triangleright x \approx \perp$ as proof that Δ_3 is in fact inhabited. Indeed, $\mathcal{E}(\Gamma \triangleright x \approx \perp, x)$ generate $_$ as the inhabitant (which is rather unhelpful, but correct).

The result of $\mathcal{G}(\Gamma, \Delta_3)$ is thus $\{_ \}$, which is not empty. Thus, $\mathcal{A}_\Gamma(\Delta, t)$ will wrap a `MayDiverge` around the first RHS.

Similarly, $\mathcal{G}(\Gamma, \Delta_4)$ needs $C(\Gamma \triangleright \emptyset, \Delta_4)$, which in turn will add $x \not\approx \perp$ to an initially empty ∇ . That entails an inhabitation check to see if x might take on any values besides \perp .

This is one possible derivation of the $\Gamma \triangleright x \not\approx \perp \vdash x$ predicate:

$$\frac{\begin{array}{c} x : \text{Maybe Int} \in \Gamma \quad \text{Nothing} \in \text{Cons}(\Gamma \triangleright x \not\approx \perp, \text{Maybe Int}) \\ \text{Inst}(\Gamma, x, \text{Nothing}) = \text{Nothing} \leftarrow x \\ (\Gamma \triangleright x \not\approx \perp \oplus \text{Nothing} \leftarrow x) \neq \perp \end{array}}{\Gamma \triangleright x \not\approx \perp \vdash x}$$

The subgoal $\Gamma \triangleright x \not\approx \perp \oplus \text{Nothing} \leftarrow x$ is handled by the second case of the match on constructor pattern constraints, because there are no other constructor pattern constraints yet in the incoming ∇ . Since there are no type constraints carried by `Nothing`, no fields and no constraints of the form

$x \neq K$ in ∇ , we end up with $\Gamma \triangleright x \neq \perp, \text{Nothing} \leftarrow x$. Which is not \perp , thus we conclude our proof of $\Gamma \triangleright x \neq \perp \vdash x$.

Next, we have to add $\text{Nothing} \leftarrow x$ to our $\nabla = x \neq \perp$, which amounts to computing $\Gamma \triangleright x \neq \perp \oplus \text{Nothing} \leftarrow x$. Conveniently, we just did that! So the result of $C(\Gamma \triangleright \emptyset, \Delta_4)$ is $\Gamma \triangleright x \neq \perp, \text{Nothing} \leftarrow x$.

Now, we see that $\mathcal{E}(\Gamma \triangleright (x \neq \perp, \text{Nothing} \leftarrow x), x) = \{\text{Nothing}\}$, which is also the result of $\mathcal{G}(\Gamma, \Delta_4)$.

The checks for Δ_5 and Δ_6 are quite similar, only that we start from $C(\Gamma \triangleright \emptyset, \Delta_1)$ (which occur syntactically in Δ_5 and Δ_6) as the initial ∇ . So, we first compute that.

Fast forward to computing $\Gamma \triangleright x \neq \perp \oplus x \neq \text{Nothing}$. Ultimately, this entails a proof of $\Gamma \triangleright x \neq \perp, x \neq \text{Nothing} \vdash x$, for which we need to instantiate the `Just` constructor:

$$\frac{\begin{array}{l} x : \text{Maybe Int} \in \Gamma \quad \text{Just} \in \text{Cons}(\Gamma \triangleright (x \neq \perp, x \neq \text{Nothing}), \text{Maybe Int}) \\ \text{Inst}(\Gamma, x, \text{Just}) = \text{Just } y \leftarrow x \\ (\Gamma, y : \text{Int} \triangleright (x \neq \perp, x \neq \text{Nothing}) \oplus \text{Just } y \leftarrow x) \neq \perp \end{array}}{\Gamma \triangleright x \neq \perp, x \neq \text{Nothing} \vdash x}$$

$\Gamma, y : \text{Int} \triangleright (x \neq \perp, x \neq \text{Nothing}) \oplus \text{Just } y \leftarrow x$ is in fact not \perp , which is enough to conclude $\Gamma \triangleright x \neq \perp, x \neq \text{Nothing} \vdash x$.

The second operand of \vee in Δ_1 is similar, but ultimately ends in \times , so will never produce a ∇ , so $C(\Gamma \triangleright \emptyset, \Delta_1) = \Gamma \triangleright x \neq \perp, x \neq \text{Nothing}$.

$C(\Gamma \triangleright \emptyset, \Delta_5)$ will then just add $x \approx \perp$ to that ∇ , which immediately refutes with $x \neq \perp$. So no `MayDiverge` around the second RHS.

$C(\Gamma \triangleright \emptyset, \Delta_6)$ is very similar to the situation with Δ_4 , just with more (non-conflicting) constraints in the incoming ∇ and with $\text{Just } y \leftarrow x$ instead of $\text{Nothing} \leftarrow x$. Thus, $\mathcal{G}(\Gamma, \Delta_6) = \{\text{Just } _ \}$.

The last bit concerns $\mathcal{G}(\Gamma, \Delta_2)$, which is empty because we ultimately would add $x \neq \text{Just}$ to the inert set $x \neq \perp, x \neq \text{Nothing}$, which refutes by the second case of $_ \oplus _$. (The \vee operand with \times in it is empty, as usual).

So we have $\mathcal{G}(\Gamma, \Delta_2) = \emptyset$ and the pattern-match is exhaustive.

The result of $\mathcal{A}_\Gamma(\Gamma, t)$ is thus `MayDiverge AccessibleRhs 1; AccessibleRhs 2`.