Online appendix for the paper

Enablers and Inhibitors in Causal Justifications of Logic Programs

published in Theory and Practice of Logic Programming

Pedro Cabalar and Jorge Fandinno
Department of Computer Science
University of Corunna, Spain
(*e-mail:* {cabalar, jorge.fandino}@udc.es)

submitted October 2015; revised December 2015; accepted January 2016

Appendix A. Auxiliary figures

Associativity	*Commutativity*	*Absorption*
$t + (u{+}w) = (t{+}u) + w$	$t + u = u + t$	$t = t + (t * u)$
$t * (u * w) = (t * u) * w$	$t * u = u * t$	$t = t * (t{+}u)$

Distributive	*Identity*	*Idempotency*	*Annihilator*
$t + (u * w) = (t{+}u) * (t{+}w)$	$t = t + 0$	$t = t + t$	$1 = 1 + t$
$t * (u{+}w) = (t * u) + (t * w)$	$t = t * 1$	$t = t * t$	$0 = 0 * t$

Fig. A 1. Sum and product satisfy the properties of a completely distributive lattice.

Appendix B. Proofs of Theorems and Implicit Results

In the following, by abuse of notation, for every function $f : \mathbf{V}_{Lb} \longrightarrow \mathbf{V}_{Lb}$, we will also denote by f a function over the set of interpretations such that $f(I)(A) = f(I(A))$ for every atom $A \in At$. We have organized the proofs into different subsections.

Appendix B.1. Proofs of Propositions 1 to 3

Proposition 1
Negation '$\sim$' is anti-monotonic. That is $t \leq u$ holds if and only if $\sim t \geq \sim u$ for any given two causal terms t and u. □

Proof. By definition $t \leq u$ iff $t * u = t$. Furthermore, by De Morgan laws, $\sim(t * u) = \sim t + \sim u$ and, thus, $\sim(t * u) = \sim t$ iff $\sim t + \sim u = \sim t$. Finally, just note that $\sim t + \sim u = \sim t$ iff $\sim t \geq \sim u$. Hence, $t \leq u$ holds iff $\sim t \geq \sim u$. □

Proposition 2
The map $t \mapsto {\sim}{\sim}t$ is a closure. That is, it is monotonic, idempotent and it holds that $t \leq {\sim}{\sim}t$ for any given causal term t. □

Proof. To show that $t \mapsto {\sim}{\sim}t$ is monotonic just note that $t \mapsto {\sim}t$ is antimonotonic (Proposition 1) and then $t \leq u$ iff ${\sim}t \geq {\sim}u$ iff ${\sim}{\sim}t \leq {\sim}{\sim}u$. Furthermore, ${\sim}{\sim}({\sim}{\sim}t) = {\sim}({\sim}{\sim}{\sim}t) = {\sim}{\sim}t$, that is, $t \mapsto {\sim}{\sim}t$ is idempotent. Finally, note that, by definition, $t \leq {\sim}{\sim}t$ iff $t * {\sim}{\sim}t = t$ and

$$\begin{aligned}
t * {\sim}{\sim}t &= t * {\sim}{\sim}t \ + \ 0 && \text{(identity)}\\
&= t * {\sim}{\sim}t \ + \ t * {\sim}t && \text{(pseudo-complement)}\\
&= (t * {\sim}{\sim}t + t) * (t * {\sim}{\sim}t + {\sim}t) && \text{(distributivity)}\\
&= (t + t) \ * \ ({\sim}{\sim}t + t) \ * \ (t + {\sim}t) \ * \ ({\sim}{\sim}t + {\sim}t) && \text{(distributivity)}\\
&= t \ * \ ({\sim}{\sim}t + t) \ * \ (t + {\sim}t) \ * \ ({\sim}{\sim}t + {\sim}t) && \text{(idempotency)}\\
&= t \ * \ (t + {\sim}t) \ * \ ({\sim}{\sim}t + t) \ * \ 1 && \text{(w. excluded middle)}\\
&= t \ * \ (t + {\sim}t) \ * \ ({\sim}{\sim}t + t) && \text{(identity)}\\
&= t \ * \ ({\sim}{\sim}t + t) && \text{(absorption)}\\
&= t && \text{(absorption)}
\end{aligned}$$

Hence, $t \mapsto {\sim}{\sim}t$ is a closure. □

Proposition 3
Given any term t, it can be rewritten as an equivalent term u in negation and disjuntive normal forms. □

Proof. This is a trivial proof by structural induction using the DeMorgan laws and negation of application axiom. Furthermore, using the axiom ${\sim}{\sim}{\sim}t = t$ no more than two nested negations are required. Furthermore, it is easy to see that by applying distributivity of "$\cdot$" and "$*$" over "$+$," every term can be equivalently represented as a term "$+$" is not in the scope of any other operation. Moreover, applying distributivity of "$\cdot$" over "$*$" every such term can be represented as one in every application subterm is elementary. □

Lemma B.1
Let t be a join irreducible causal value. Then, either $t * {\sim}{\sim}u = 0$ or $t * {\sim}{\sim}u$ is join irreducible for every causal value $u \in \mathbf{V}_{Lb}$. □

Proof. Suppose that $t * u$ is not join irreducible and let $W \subseteq \mathbf{V}_{Lb}$ a set of causal values such that $w \neq t * {\sim}{\sim}u$ for every $w \in W$ and $t * {\sim}{\sim}u = \sum_{w \in W} w$. Since $t * {\sim}{\sim}u = \sum_{w \in W} w$, it follows that $w \leq t * {\sim}{\sim}u$ for every $w \in W$ and, since $w \neq t * {\sim}{\sim}u$, it follows that $w < t * {\sim}{\sim}u$ for every $w \in W$. Furthermore, $t * {\sim}{\sim}u + t * {\sim}u = t * ({\sim}{\sim}u + {\sim}u) = t$.

Since t is join irreducible, it follows that either $t = t * {\sim}{\sim}u$ or $t = t * {\sim}u$. If $t = t * {\sim}u$, then $t * {\sim}{\sim}u = (t * {\sim}u) * {\sim}{\sim}u = 0$. Otherwise, $t = t * {\sim}{\sim}u$ and t is join irreducible by hypothesis. □

Lemma B.2
Let t be a term. Then $\lambda^p(\sim t) = \neg\lambda^p(t)$. □

Proof. We proceed by structural induction assuming that t is in negated normal form. In case that $t = a$ is elementary, it follows that $\lambda^p(\sim a) = \neg a = \neg\lambda^p(a)$. In case that $t = \sim a$ with a elementary, $\lambda^p(\sim t) = \lambda^p(\sim\sim a)$ and $\lambda^p(\sim\sim a) = a = \neg\neg a = \neg\lambda^p(\sim a) = \lambda^p(t)$. In case that $t = \sim\sim a$, with a elementary, $\lambda^p(\sim t) = \lambda^p(\sim\sim\sim a)$ and

$$\lambda^p(\sim\sim\sim a) = \lambda^p(\sim a) = \neg a = \neg\lambda^p(\sim\sim a) = \neg\lambda^p(t)$$

In case that $t = u + v$. Then

$$\lambda^p(\sim t) = \lambda^p(\sim u * \sim v) = \lambda^p(\sim u) \wedge \lambda^p(\sim v)$$

By induction hypothesis $\lambda^p(\sim u) = \neg\lambda^p(u)$ and $\lambda^p(\sim v) = \neg\lambda^p(v)$ and, therefore, it holds that $\lambda^p(\sim t) = \neg\lambda^p(u) \wedge \neg\lambda^p(v)$. Thus, $\neg\lambda^p(t) = \neg(\lambda^p(u) \vee \lambda^p(v)) = \neg\lambda^p(u) \wedge \neg\lambda^p(v) = \lambda^p(\sim t)$.

In case that $t = u \otimes v$ with $\otimes \in \{*, \cdot\}$. Then $\lambda^p(\sim t) = \lambda^p(\sim u + \sim v) = \lambda^p(\sim u) \vee \lambda^p(\sim v)$ and by induction hypothesis $\lambda^p(\sim u) = \neg\lambda^p(u)$ and $\lambda^p(\sim v) = \neg\lambda^p(v)$. Consequently it holds that $\lambda^p(\sim t) = \neg\lambda^p(t)$. □

Lemma B.3
Let t be a term and ϕ a provenance term. If $\phi \leq \lambda^p(t)$, then $\lambda^p(\sim t) \leq \neg\phi$ and if $\lambda^p(t) \leq \phi$, then $\neg\phi \leq \lambda^p(\sim t)$. □

Proof. If $\phi \leq \lambda^p(t)$, then $\phi = \lambda^p(t) * \phi$ and then $\neg\phi = \neg\lambda^p(t) + \neg\phi$ and, by Lemma B.2, it follows that $\neg\phi = \lambda^p(\sim t) + \neg\phi$. Hence $\lambda^p(\sim t) \leq \neg\phi$. Furthermore if $\lambda^p(t) \leq \phi$, then $\phi = \lambda^p(t) + \phi$ and then $\neg\phi = \neg\lambda^p(t) * \neg\phi$ and, by Lemma B.2, it follows that $\neg\phi = \lambda^p(\sim t) * \neg\phi$. Hence $\neg\phi \leq \lambda^p(\sim t)$. □

Appendix B.2. Proof of Theorem 1

The proof of Theorem 1 will relay on the definition of the following direct consequence operator

$$\tilde{T}_P(\tilde{I})(H) \stackrel{\text{def}}{=} \sum \{ \ (\tilde{I}(B_1) * \ldots * \tilde{I}(B_n)) \cdot r_i \ | \ (r_i : \ H \leftarrow B_1, \ldots, B_n) \in P \ \}$$

for any CG interpretation $\tilde{I}$ and atom $H \in At$. Note that the definition of this direct consequence operator $\tilde{T}_P$ is analogous to the T_P operator, but the domain and image of $\tilde{T}_P$ are the set of CG interpretations while the domain and image of T_P are the set of ECJ interpretations.

Theorem 11 (*Theorem 2 from Cabalar et al. 2014a*)
Let P be a (possibly infinite) positive logic program with n causal rules. Then, (*i*) $lfp(\tilde{T}_P)$ is the least model of P, and (*ii*) $lfp(\tilde{T}_P) = \tilde{T}_P{\uparrow}^{\omega}(\mathbf{0}) = \tilde{T}_P{\uparrow}^{n}(\mathbf{0})$. □

Proof of Theorem 1. Assume that every term occurring in P is NNF and let Q be the program obtained by renaming in P each occurrence of $\sim l$ as l' and each occurrence of $\sim\sim l$ as l'' with l' and l'' new symbols. Note that this renaming implies that $\sim l$ and $\sim\sim l$ are treated as completely independent symbols from l and, thus, all equalities among terms derived from program Q are also satisfied by P, although the converse does not hold. Note also that, since $\sim$ does not occur in Q, this is also a CG program. From Theorem 11, $\text{lfp}(\tilde{T}_Q) = \tilde{T}_Q\!\uparrow^{\omega}(\mathbf{0})$ is the least model of Q. By renaming back l' and l'' as $\sim l$ and $\sim\sim l$ in $\tilde{T}_Q\!\uparrow^{k}(\mathbf{0})$ we obtain $T_P\!\uparrow^{k}(\mathbf{0})$ for any k. Hence, $\text{lfp}(T_P) = T_P\!\uparrow^{\omega}(\mathbf{0})$ is the least model of P. Statement (ii) is proved in the same manner. □

Appendix B.3. Proof of Proposition 4

Lemma B.4
Let P_1 and P_2 be two programs and let U_1 and U_2 be two interpretations such that $P_1 \supseteq P_2$ and $U_1 \leq U_2$. Let also I_1 and I_2 be the least models of $P_1^{U_1}$ and $P_2^{U_2}$, respectively. Then $I_1 \geq I_2$. □

Proof. First, for any rule r_i and pair of interpretations J_1 and J_2 such that $J_1 \geq J_2$,

$$J_1(body^+(r_i^{U_1})) \;\geq\; J_2(body^+(r_i^{U_2}))$$

Furthermore, since $U_1 \leq U_2$, by Proposition 1, it follows

$$U_1(body^-(r_i^{U_1})) \;\geq\; U_2(body^-(r_i^{U_2}))$$

and, since by Definition 5 $J_j(body^-(r_i^{U_1})) \stackrel{\text{def}}{=} U_j(body^-(r_i^{U_1}))$, it follows that

$$J_1(body^-(r_i^{U_1})) \;\geq\; J_2(body^-(r_i^{U_2}))$$

Hence, we obtain that $J_1(body(r_i^{U_1})) \geq J_2(body(r_i^{U_2}))$.

Since $P_1 \supseteq P_2$, it follows that every rule $r_i \in P_2$ is in P_1 as well. Thus, $T_{P_1^{U_1}}(J_1)(H) \geq T_{P_2^{U_2}}(J_2)(H)$ for every atom H. Furthermore, since

$$T_{P_1^{U_1}}\!\uparrow^{0}(\mathbf{0})(H) \;=\; T_{P_2^{U_2}}\!\uparrow^{0}(\mathbf{0})(H) \;=\; 0$$

it follows $T_{P_1^{U_1}}\!\uparrow^{i}(\mathbf{0})(H) \;\geq\; T_{P_2^{U_2}}\!\uparrow^{i}(\mathbf{0})(H)$ for all $0 \leq i$. Finally,

$$T_{P_j^{U_j}}\!\uparrow^{\omega}(\mathbf{0})(H) \;\stackrel{\text{def}}{=}\; \sum_{i\leq\omega} T_{P_j^{U_j}}\!\uparrow^{i}(\mathbf{0})(H) \;=\; 0$$

and hence $T_{P_1^{U_1}}\!\uparrow^{\omega}(\mathbf{0})(H) \;\geq\; T_{P_2^{U_2}}\!\uparrow^{\omega}(\mathbf{0})(H)$. By Theorem 1, these are respectively the least models of $P_1^{U_1}$ and $P_2^{U_2}$. That is $I_1 \geq I_2$. □

Proposition 4
Γ_P operator is anti-monotonic and operator Γ_P^2 is monotonic. That is, $\Gamma_P(U_1) \geq \Gamma_P(U_2)$ and $\Gamma_P^2(U_1) \leq \Gamma_P^2(U_2)$ for any pair of interpretations U_1 and U_2 such that $U_1 \leq U_2$. □

Proof. Since $U_1 \leq U_2$, by Lemma B.4, it follows $I_1 \geq I_2$ with I_1 and I_2 being respectively the least models of P^{U_1} and P^{U_2}. Then, $\Gamma_P(U_1) = I_1$ and $\Gamma_P(U_2) = I_2$ and, thus, $\Gamma_P(U_1) \geq \Gamma_P(U_2)$. Since Γ_P is anti-monotonic it follows that Γ_P^2 is monotonic. □

Appendix B.4. Proof of Theorem 2

The proof of Theorem 2 will rely on the relation between ECJ justifications and non-hypothetical WnP justifications established by Theorem 9 and it can be found below the proof of that theorem in page 13.

Appendix B.5. Proof of Theorem 3

Definition 17
A term $t \in \mathbf{V}_{Lb}$ is *join irreducible* iff $t = \sum_{u \in U} u$ implies that $u = t$ for some $u \in U$ and it is *join prime* iff $t \leq \sum_{u \in U} u$ implies that $u \leq t$ for some $u \in U$. □

Proposition 5
The following results hold:

1. A term is join irreducible iff is join prime.
2. If Lb is finite, then every term t can be represented as a unique finite sum of pairwise incomparable join irreducible terms. □

Proof. The first result directly follows from Theorem 1 in (Balbes and Dwinger 1975, page 65). Furthermore, from Theorem 2 in (Balbes and Dwinger 1975, page 66), in every distributive lattice satisfying the descending chain condition, any element can be represented as a unique finite sum of pairwise incomparable join irreducible elements and it is clear that every finite lattice satisfies the descending chain condition. □

Lemma B.5
Let P be a positive program over a signature $\langle At, Lb \rangle$ where Lb is a finite set of labels and Q be the result of removing all rules labelled by some label $l \in Lb$. Let I and J be two interpretations such that J such that $\rho_{\sim l}(I) \geq J$. Then, $\rho_{\sim l}(\Gamma_P(I)) \leq \Gamma_Q(J)$. □

Proof. By definition $\Gamma_P(I)$ and $\Gamma_Q(J)$ are the least models of programs P^I and Q^J, respectively. Furthermore, from Theorem 1, the least model of any program P is the least fixpoint of the T_P operator, that is, $\Gamma_X(Y) = T_{X^Y} \uparrow^{\omega} (\mathbf{0})$ with $X \in \{P, Q\}$ and $X^Y \in \{P^I, P^J\}$. Then, the proof follows by induction assuming that $u \leq T_{Q^J} \uparrow^{\beta} (\mathbf{0})(H)$ implies $\rho_{\sim l}(u) \leq T_{Q^I} \uparrow^{\beta} (\mathbf{0})(H)$ for any join irreducible u, atom H and every ordinal $\beta < \alpha$.

Note that $T_{Q^J} \uparrow^{0} (\mathbf{0})(H) = 0 = \rho_{\sim l}(0) = T_{P^I} \uparrow^{0} (\mathbf{0})(A)$ for any atom H and, thus, the statement holds vacuous.

If α is a successor ordinal, since $u \leq T_{P^I} \uparrow^{\alpha} (\mathbf{0})(H)$, there is a rule in P of the form (4) such that

$$u \leq (u_{B_1} * \ldots * u_{B_m} * u_{C_1} * \ldots * u_{C_n}) \cdot r_i$$

where $u_{B_j} \leq T_{P^I} \uparrow^{\alpha-1} (\mathbf{0})(B_j)$ and $u_{C_j} \leq \sim I(C_j)$ for each positive literal B_j and each negative literal *not* C_j in the body of rule r_i. Then,

1. By induction hypothesis, it follows that $\rho_{\sim l}(u_{B_j}) \leq T_{Q^J} \uparrow^{\alpha-1} (\mathbf{0})(B_j)$, and
2. from $\rho_{\sim l}(I(H)) \geq J(H)$, it follows that $u_{C_j} \leq \sim I(C_j)$ implies $\rho_{\sim l}(u_{C_j}) \leq \sim J(C_j)$.

Furthermore, if $r_i \neq l$, then $r_i \in Q$ and, thus,

$$\rho_{\sim l}(u) \leq (\rho_{\sim l}(u_{B_1}) * \ldots * \rho_{\sim l}(u_{B_m}) * \rho_{\sim l}(u_{C_1}) * \ldots * \rho_{\sim l}(u_{C_n})) \cdot r_i \leq T_{Q^J} \uparrow^{\alpha} (\mathbf{0})(H)$$

If otherwise $r_i = l$, then $\rho_{\sim l}(u) = 0 \leq T_{Q^J} \uparrow^{\alpha} (\mathbf{0})(H)$.

In case that α is a limit ordinal, $u \leq T_{P^I} \uparrow^{\alpha} (\mathbf{0})$ iff $u \leq T_{P^I} \uparrow^{\beta} (\mathbf{0})$ for some $\beta < \alpha$ and any join irreducible u. Hence, by induction hypothesis, it follows that $\rho_{\sim I}(u) \leq T_{Q^J} \uparrow^{\beta} (\mathbf{0}) \leq T_{Q^J} \uparrow^{\alpha} (\mathbf{0})$ and, thus, $\rho_{\sim I}(T_{P^I} \uparrow^{\alpha} (\mathbf{0})) \leq T_{Q^J} \uparrow^{\alpha} (\mathbf{0})$. □

Proof of Theorem 3. In the sake of simplicity, we just write ρ instead of $\rho_{\sim r_i}$. Note that, by definition, for any atom H, it follows that $\mathbb{W}_X(H) = \mathbb{L}_X(H)$ with $X \in \{P, Q\}$. The proof follows by induction in the number of steps of the Γ^2 operator assuming as induction hypothesis that $\Gamma^2_Q \uparrow^{\beta} (\mathbf{0}) \leq \rho(\Gamma^2_P \uparrow^{\beta} (\mathbf{0}))$ for every $\beta < \alpha$. Note that $\Gamma^2_Q \uparrow^{0} (\mathbf{0})(H) = 0 \leq \rho(\Gamma^2_P \uparrow^{0} (\mathbf{0}))(H)$ and, thus, the statement trivially holds for $\alpha = 0$.

In case that α is a successor ordinal, by induction hypothesis, it follows that

$$\Gamma^2_Q \uparrow^{\alpha-1} (\mathbf{0}) \;\leq\; \rho(\Gamma^2_P \uparrow^{\alpha-1} (\mathbf{0}))$$

and, from Lemma B.5, it follows that

$$\begin{array}{lcl} \Gamma_Q(\Gamma^2_Q \uparrow^{\alpha-1} (\mathbf{0})) & \geq & \rho(\Gamma_P(\Gamma^2_P \uparrow^{\alpha-1} (\mathbf{0}))) \\ \Gamma^2_Q(\Gamma^2_Q \uparrow^{\alpha-1} (\mathbf{0})(H))) & \leq & \rho(\Gamma^2_P(\Gamma^2_P \uparrow^{\alpha-1} (\mathbf{0}))) \end{array}$$

That is, $\Gamma^2_Q \uparrow^{\alpha} (\mathbf{0}) \;\leq\; \rho(\Gamma^2_P \uparrow^{\alpha} (\mathbf{0})$.

Finally, in case that α is a limit ordinal, every join irreducible u satisfies $u \leq \Gamma^2_Q \uparrow^{\alpha} (\mathbf{0}) = \sum_{\beta<\alpha} \Gamma^2_Q \uparrow^{\beta} (\mathbf{0})$ iff $u \leq \Gamma^2_Q \uparrow^{\beta} (\mathbf{0})$ for some $\beta < \alpha$ and, thus, by induction hypothesis $\rho(u) \leq \Gamma^2_P \uparrow^{\beta} (\mathbf{0}) \leq \Gamma^2_P \uparrow^{\alpha} (\mathbf{0})$. Consequently, $\Gamma^2_Q \uparrow^{\infty} (\mathbf{0}) \leq \rho(\Gamma^2_P \uparrow^{\infty} (\mathbf{0})$ and $\mathbb{W}_Q(A) \leq \rho(\mathbb{W}_P(A)$ for any atom A. □

Appendix B.6. Proof of Theorem 5

By $\tilde{\Gamma}_P(\tilde{I})$ we denote the least model of a program $P^{\tilde{I}}$. Note that the relation between $\tilde{\Gamma}_P$ and Γ_P is similar to the relation between $\tilde{T}_P$ and T_P: the $\tilde{\Gamma}_P$ operator is a function in the set of CG interpretations while Γ_P is a function in the set of ECJ interpretations. Note also that the evaluation of negated literals with respect to CG and ECJ interpretations and, thus, the reducts $P^{\tilde{I}}$ and P^I may be different even if $\tilde{I}(A) = I(A)$ for every atom A.

Lemma B.6
Let P be a labelled logic program, $\tilde{I}$ and J be respectively an CG and a ECJ interpretation such that $\tilde{I} \geq \lambda^c(J)$. Then $\tilde{\Gamma}_P(\tilde{I}) \leq \lambda^c(\Gamma_P(J))$. □

Proof. By definition $\tilde{\Gamma}_P(\tilde{I})$ and $\Gamma_P(J)$ are respectively the least model of the programs $P^{\tilde{I}}$ and P^J. Furthermore, from Theorem 1 the least model of any program P is the least fixpoint of the T_P operator, that is, $\tilde{\Gamma}_P(\tilde{I}) = \tilde{T}_{P^{\tilde{I}}} \uparrow^{\omega} (\mathbf{0})$ and $\Gamma_P(J) = T_{P^J} \uparrow^{\omega} (\mathbf{0})$. In case that $\alpha = 0$, it follows that $\tilde{T}_{P^{\tilde{I}}} \uparrow^{0} (\mathbf{0})(H) \;=\; 0 \;\leq\; \lambda^c(T_{P^J} \uparrow^{0} (\mathbf{0}))(H)$ for every atom H. We assume as induction hypothesis that $\tilde{T}_{P^{\tilde{I}}} \uparrow^{\beta} (\mathbf{0}) \leq \lambda^c(T_{P^J} \uparrow^{\beta} (\mathbf{0}))$ for all $\beta < \alpha$.

In case that α is a successor ordinal, $E \leq \tilde{T}_{P^{\tilde{I}}} \uparrow^{\alpha} (\mathbf{0})(H) = \tilde{T}_{P^{\tilde{I}}}(\tilde{T}_{P^{\tilde{I}}} \uparrow^{\alpha-1} (\mathbf{0}))(H)$ if and only if there is a rule R^I in $P^{\tilde{I}}$ of the form

$$r_i : H \leftarrow B_1, \ldots, B_m,$$

which is the reduct of a rule R of the form (4) in P and that satisfies $E \leq (E_{B_1} * \ldots * E_{B_m}) \cdot r_i$ with each $E_{B_j} \leq \tilde{T}_{P^{\tilde{I}}} \uparrow^{\alpha-1} (\mathbf{0})(B_j)$ and $\tilde{I}(C_j) = 0$ for all B_j and C_j in $body(R)$. Hence there is a rule in P^J of the form

$$r_i : H \leftarrow B_1, \ldots, B_m,\ J(not\, C_1), \ldots,\ J(not\, C_n)$$

and, by induction hypothesis, $E_{B_j} \leq \lambda^c\big(T_{P^J} \uparrow^{\alpha-1} (\mathbf{0})(B_j)\big)$ for all B_j. Furthermore, by definition

$$\big(T_{P^J} \uparrow^{\alpha-1} (\mathbf{0})(B_1) * \ldots * T_{P^J} \uparrow^{\alpha-1} (\mathbf{0})(B_m) * J(not\, C_1) * \ldots * J(not\, C_m)\big) \cdot r_i \;\; \leq \;\; T_{P^J} \uparrow^{\alpha} (\mathbf{0})(H)$$

From the fact that $\tilde{I}(C_j) = 0$ and the lemma's hypothesis $\tilde{I} \geq \lambda^c(J)$, it follows that $0 \geq \lambda^c(J(C_j))$ and, thus, $1 \leq \lambda^c(\sim J(C_j)) = \lambda^c(J(not\, C_j))$. Hence,

$$\begin{aligned}
&\lambda^c\big((T_{P^J} \uparrow^{\alpha-1} (\mathbf{0})(B_1) * \ldots * T_{P^J} \uparrow^{\alpha-1} (\mathbf{0})(B_m) * J(not\, C_1) * \ldots * J(not\, C_m)) \cdot r_i\big) \;\; = \\
&\quad = \;\; \lambda^c\big((T_{P^J} \uparrow^{\alpha-1} (\mathbf{0})(B_1) * \ldots * T_{P^J} \uparrow^{\alpha-1} (\mathbf{0})(B_m)\big) * \lambda^c\big(J(not\, C_1)\big) * \ldots * \lambda^c\big(J(not\, C_m)\big)\big) \cdot r_i \\
&\quad = \;\; \lambda^c\big((T_{P^J} \uparrow^{\alpha-1} (\mathbf{0})(B_1) * \ldots * T_{P^J} \uparrow^{\alpha-1} (\mathbf{0})(B_m)\big) * 1 * \ldots * 1\big) \cdot r_i \\
&\quad = \;\; \lambda^c\big((T_{P^J} \uparrow^{\alpha-1} (\mathbf{0})(B_1) * \ldots * T_{P^J} \uparrow^{\alpha-1} (\mathbf{0})(B_m)\big)\big) \cdot r_i
\end{aligned}$$

and, thus,

$$\lambda^c\big((T_{P^J} \uparrow^{\alpha-1} (\mathbf{0})(B_1) * \ldots * T_{P^J} \uparrow^{\alpha-1} (\mathbf{0})(B_m)\big)\big) \cdot r_i \;\; \leq \;\; \lambda^c\big(T_{P^J} \uparrow^{\alpha} (\mathbf{0})(H)\big)$$

Since $E_{B_j} \leq \lambda^c\big(T_{P^J} \uparrow^{\alpha-1} (\mathbf{0})(B_j)\big)$ for all B_j, it follows that

$$E \;\; \leq \;\; (E_{B_1} * \ldots * E_{B_m}) \cdot r_i \;\; \leq \;\; \lambda^c(T_{P^J} \uparrow^{\alpha} (\mathbf{0}))(H)$$

Finally, in case that α is a limit ordinal, it follows from Theorem 1 that $\alpha = \omega$. Furthermore, since $\tilde{I}$ is a CG interpretation, it follows that $P^{\tilde{I}}$ is a CG program and, thus, $E \leq T_{P^{\tilde{I}}} \uparrow^{\omega} (\mathbf{0})$ iff $E \leq T_{P^{\tilde{I}}} \uparrow^{n} (\mathbf{0})$ for some $n < \omega$ (see Cabalar et al. 2014a). Hence, by induction hypothesis, it follows that $E \leq T_{P^J} \uparrow^{n} (\mathbf{0}) \leq T_{P^J} \uparrow^{\omega} (\mathbf{0})$. □

Lemma B.7
Let P be a labelled logic program over a signature $\langle At, Lb \rangle$ where Lb is a finite set of labels, $\tilde{I}$ and J respectively be a CG and a ECJ interpretation such that $\tilde{I} \leq \lambda^c(J)$. Then $\tilde{\Gamma}_P(\tilde{I}) \geq \lambda^c(\Gamma_P(J))$. □

Proof. Since Lb is finite, it follows that $\mathbf{V}_{Lb}$ is also finite. Furthermore, since $\mathbf{V}_{Lb}$ is a finite distributive lattice, every element $t \in \mathbf{V}_{Lb}$ can be represented as a unique sum of join irreducible elements (Proposition 5).

Assume as induction hypothesis that $u \leq T_{P^J} \uparrow^{\beta} (\mathbf{0})(H)$ implies $\lambda^c(u) \leq \tilde{T}_{P^{\tilde{I}}} \uparrow^{\beta} (\mathbf{0})(H)$ for every join irreducible u, atom $H \in At$ and ordinal $\beta < \alpha$.

In case that α is a successor ordinal. For any join irreducible justification $u \leq T_{P^J} \uparrow^{\alpha} (\mathbf{0})(H)$ there is a rule R^J in P^J of the form (6) and there are join irreducible terms $u_{B_j} \leq T_{P^J} \uparrow^{\alpha-1} (\mathbf{0})(B_j)$ and $u_{C_j} \leq \sim J(C_j)$ for all B_j and C_j such that

$$u \;\; \leq \;\; (u_{B_1} * \ldots * u_{B_m} * u_{C_1} * \ldots * u_{C_n}) \cdot r_i$$

If u_{C_j} contains an oddly negated label for some C_j, then $\lambda^c(u_{C_j}) = 0$ and it consequently follows that $\lambda^c(u) = 0 \; \leq \; \tilde{T}_{P^{\tilde{I}}} \uparrow^{\alpha} (\mathbf{0})(H)$. Thus, we assume that u_{C_j} only contains evenly negated labels for any C_j. Note that, since $u_{C_j} \leq \sim J(C_j)$, then u_{C_j} cannot contain any non-negated label, that is, all occurrences of labels in u_{C_j} are strictly evenly negated and, thus, every term $u'_{C_j} \leq J(C_j)$ must contain some oddly negated label. Hence, $\tilde{I}(C_j) \leq \lambda^c(J(C_j)) = 0$ for any C_j and there is a rule $R^{\tilde{I}}$ in $Q^{\tilde{I}}$ of the form

$$r_i : \ H \leftarrow B_1, \ldots, B_m$$

By induction hypothesis, $u_{B_j} \leq T_{P^J} \uparrow^{\alpha-1} (\mathbf{0})(B_j)$ implies $\lambda^c(u_{B_j}) \leq \tilde{T}_{P^{\tilde{I}}} \uparrow^{\alpha-1} (\mathbf{0})(B_j)$ and, consequently, $\lambda^c(u) \leq \tilde{T}_{P^{\tilde{I}}} \uparrow^{\alpha} (\mathbf{0})(H)$.

Since $T_{P^J} \uparrow^{\alpha} (\mathbf{0})(H) = \sum_{u \in U_H} u$ where every $u \in U_H$ is join irreducible and every $u \in U_H$ satisfies $u \leq T_{P^J} \uparrow^{\alpha} (\mathbf{0})(H)$, it follows that $\lambda^c(u) \leq \tilde{T}_{P^{\tilde{I}}} \uparrow^{\alpha} (\mathbf{0})(H)$ and, thus, $\sum_{u \in U_H} \lambda^c(u) \leq \tilde{T}_{P^{\tilde{I}}} \uparrow^{\alpha} (\mathbf{0})(H)$. Note that, by definition, $\lambda^c(\sum_{u \in U_H} u) = \sum_{u \in U_H} \lambda^c(u)$ and, thus,

$$\lambda^c(T_{P^J} \uparrow^{\alpha} (\mathbf{0})(H)) \ = \ \lambda^c(\sum_{u \in U_H} u) \ \leq \ \tilde{T}_{P^{\tilde{I}}} \uparrow^{\alpha} (\mathbf{0})(H)$$

In case that α is a limit ordinal, it follows $u \leq T_{P^J} \uparrow^{\alpha} (\mathbf{0})(H)$ iff $u \leq T_{P^J} \uparrow^{\beta} (\mathbf{0})(H)$ for some $\beta < \omega$ and, by induction hypothesis, it follows that $\lambda^c(u) \leq \tilde{T}_{P^{\tilde{I}}} \uparrow^{\beta} (\mathbf{0})(H) \leq \tilde{T}_{P^{\tilde{I}}} \uparrow^{\alpha} (\mathbf{0})(H)$ and, thus, $\tilde{T}_{P^{\tilde{I}}} \uparrow^{\alpha} (\mathbf{0}) \geq \lambda^c(T_{P^J} \uparrow^{\alpha} (\mathbf{0}))$.

Finally, by definition $\tilde{\Gamma}_P(\tilde{I})$ and $\Gamma_P(J)$ are respectively the least models of $P^{\tilde{I}}$ and P^J and, from Theorem 11, these are precisely $\tilde{T}_{P^{\tilde{I}}} \uparrow^{\omega} (\mathbf{0})$ and $T_{P^J} \uparrow^{\omega} (\mathbf{0})$. Hence, $\tilde{T}_{P^{\tilde{I}}} \uparrow^{\omega} (\mathbf{0}) \geq \lambda^c(T_{P^J} \uparrow^{\omega} (\mathbf{0}))$ implies $\tilde{\Gamma}_P(\tilde{I}) \geq \lambda^c(\Gamma_P(J))$. □

Proposition 6
Given a program P over a signature $\langle At, Lb \rangle$ where Lb is a finite set of labels, any ECJ interpretation I satisfies $\tilde{\Gamma}_P(\lambda^c(I)) = \lambda^c(\Gamma_P(I)))$. □

Proof of Proposition 6. Let $\tilde{I}$ be a CG interpretation such that $I(H) = \tilde{I}(H)$ for every atom H. Then, it follows that $\tilde{I} = \lambda^c(I)$. Hence, from Lemmas B.6 and B.7, it respectively follows that $\tilde{\Gamma}_P(\tilde{I}) \leq \lambda^c(\Gamma_P(I))$ and $\tilde{\Gamma}_P(\tilde{I}) \geq \lambda^c(\Gamma_P(I))$. Then, $\tilde{\Gamma}_P(\tilde{I}) = \tilde{\Gamma}_P(\lambda^c(I)) = \lambda^c(\Gamma_P(I))$. □

Proof of Theorem 5. According to (Cabalar et al. 2014a), a CG interpretation $\tilde{I}$ is a CG stable model of P iff $\tilde{I}$ is the least model of the program $P^{\tilde{I}}$. Then, the CG stable models are just the fixpoints of the $\tilde{\Gamma}_P$ operator.

Let $\tilde{I}$ be a CG stable model according to (Cabalar et al. 2014a), let I be a ECJ interpretation such that $I(H) = \tilde{I}(H)$ for every atom $H \in At$ and let $J \stackrel{\text{def}}{=} \Gamma^2_P \uparrow^{\infty} (I)$ be the least fixpoint of Γ^2_P iterating from I. Since $I(H) = \tilde{I}(H)$ for every atom $H \in At$, it follows that $\tilde{I} = \lambda^c(I)$ and, by definition of CG stable model, it follows that $\tilde{I} = \tilde{\Gamma}_P(\tilde{I})$. Thus, from Proposition 6, it follows that $\tilde{I} = \lambda^c(\Gamma_P(I))$. Applying $\tilde{\Gamma}_P$ to both sides of this equality, we obtain that $\tilde{\Gamma}_P(\tilde{I}) = \tilde{\Gamma}_P(\lambda^c(\Gamma_P(I)))$. From Proposition 6 again, it follows that $\tilde{\Gamma}_P(\lambda^c(\Gamma_P(I))) = \lambda^c(\Gamma_P(\Gamma_P(I))) = \lambda^c(\Gamma^2_P(I))$ and, thus, $\tilde{\Gamma}_P(\tilde{I}) = \lambda^c(\Gamma^2_P(I))$. Furthermore, since $\tilde{I} = \tilde{\Gamma}_P(\tilde{I})$, it follows that $\tilde{I} = \lambda^c(\Gamma^2_P(I))$. Inductively applying this argument, it follows that $\tilde{I} = \lambda^c(\Gamma^2_P \uparrow^{\alpha} (I))$ for any successor ordinal α. Moreover, for a limit ordinal α,

$$\lambda^c\big(\Gamma^2_P \uparrow^{\alpha} (I)\big) \ = \ \lambda^c\Big(\sum_{\beta < \alpha} \Gamma^2_P \uparrow^{\beta} (I)\Big) \ = \ \sum_{\beta < \alpha} \lambda^c\big(\Gamma^2_P \uparrow^{\beta} (I)\big) \ = \ \tilde{I}$$

Then, since we have defined $J = \Gamma^2_P \uparrow^{\infty} (I)$, it follows that $\tilde{I} = \lambda^c(J) = \lambda^c(I)$ and, since we also have that $\tilde{I} = \lambda^c(\Gamma_P(I))$, we obtain that $\lambda^c(I) = \lambda^c(\Gamma_P(I))$.

The other way around. Let I be a fixpoint of Γ_P^2 such that $\lambda^c(I) = \lambda^c(\Gamma_P(I))$ and let $\tilde{I} \stackrel{\text{def}}{=} \lambda^c(I)$. In the same way as above, it follows that $\tilde{\Gamma}_P(\tilde{I}) = \lambda^c(\Gamma_P(I)) = \lambda^c(I) = \tilde{I}$. That is, $\tilde{\Gamma}_P(\tilde{I}) = \tilde{I}$ and so that $\tilde{I}$ is a causal stable model of P according to (Cabalar et al. 2014a). □

Appendix B.7. Proof of Theorem 6

Proof of Theorem 6 . Let $\tilde{I}$ be a causal stable model of P and I be the correspondent fixpoint of Γ_P^2 with $\tilde{I} = \lambda^c(I)$. Since E is a enabled justification of A, i.e. $E \leq \mathbb{W}_P(A)$, then $E \leq \mathbb{L}_P(A)$ with $\mathbb{L}_P$ the least fixpoint of Γ_P^2. Since, I is a fixpoint of Γ_P^2, if follows that $E \leq \mathbb{L}_P(A) \leq I(A)$ and, thus, $\lambda^c(E) \leq \lambda^c(I(A)) = \tilde{I}(A)$. Then $G \stackrel{\text{def}}{=} graph(\lambda^c(E))$ is, by definition, a causal explanation of the atom A.

Appendix B.8. Proof of Theorem 7

The proof of Theorem 7 will need the following definition.

Definition 18
Given a program P, a *WnP interpretation* is a mapping $\mathfrak{I} : At \longrightarrow \mathbf{B}_{Lb}$ assigning a Boolean formula to each atom. The evaluation of a negated literal $not A$ with respect to a WnP interpretation is given by $\mathfrak{I}(not A) = \neg\mathfrak{I}(A)$. An interpretation $\mathfrak{I}$ is a WnP model of rule like (4) iff

$$\mathfrak{I}(B_1) * \ldots * \mathfrak{I}(B_m) * \mathfrak{I}(not\, C_1) * \ldots * \mathfrak{I}(not\, C_n) * r_i \ \leq\ \mathfrak{I}(H)$$

The operator $\mathfrak{G}_P(\mathfrak{I})$ maps a WnP interpretation $\mathfrak{I}$ to the least model of the program $P^{\mathfrak{I}}$. □

Note that the only differences in the model evaluation between ECJ and WnP comes from the valuation of negative literals and the use of '$*$' instead of '$\cdot$' for keeping track of rule application. Besides, we will also use the following facts whose proof is addressed in an appendix.

Definition 19
Given a positive program P, we define a direct consequence operator $\mathfrak{T}_P$ such that

$$\mathfrak{T}_P(\mathfrak{I})(H) \ \stackrel{\text{def}}{=}\ \sum \left\{ \mathfrak{I}(B_1) * \ldots * \mathfrak{I}(B_n) * r_i \ \mid\ (r_i :\ H \leftarrow B_1, \ldots, B_n) \in P \right\}$$

for any WnP interpretation $\mathfrak{I}$ and atom $H \in At$. □

Definition 20 (*From Damásio et al. 2013*)
Given a program P, its why-not program is given by $\mathscr{P} \stackrel{\text{def}}{=} P \cup P'$ here P' contains a labelled fact of the form

$$\neg not(A) : A$$

for each atom $A \in At$ not occurring in P as a fact. The why-not provenance information under the well-founded semantics is defined as follows: $Why_{\mathscr{P}}(H) = [\mathfrak{T}_{\mathscr{P}}(H)]$; $Why_{\mathscr{P}}(H) = [\neg\mathfrak{TU}_{\mathscr{P}}(H)]$; and $Why_{\mathscr{P}}(undef\, A) = [\neg\mathfrak{T}_{\mathscr{P}}(H) \wedge \mathfrak{TU}_Q(H)]$ where $\mathfrak{T}_{\mathscr{P}}$ and $\mathfrak{TU}_{\mathscr{P}} = \mathfrak{G}_{\mathscr{P}}(\mathfrak{T}_{\mathscr{P}})$ be the least and greates fixpoints of $\mathfrak{G}_{\mathscr{P}}^2$, respectively. □

Lemma B.8
Let P be a labelled logic program over a signature $\langle At, Lb\rangle$ where Lb is a finite set of labels and let I and $\mathfrak{I}$ be respectively a ECJ and a WnP interpretation such that $\lambda^p(I) \geq \mathfrak{I}$. Then, $\lambda^p(\Gamma_{\mathfrak{P}}(I)) \leq \mathfrak{G}_{\mathscr{P}}(\mathfrak{I})$.

Proof. By definition $\Gamma_{\mathfrak{P}}(I)$ and $\mathfrak{G}_{\mathscr{P}}(\mathfrak{I})$ are the least model of the programs $\mathfrak{P}^I$ and $\mathscr{P}^{\mathfrak{I}}$, respectively. Furthermore, the least model of programs $\mathfrak{P}^I$ and $\mathscr{P}^{\mathfrak{I}}$ are the least fixpoint of the $T_{\mathfrak{P}^I}$ and $\mathfrak{T}_{\mathscr{P}^{\mathfrak{I}}}$ operators, that is, $\Gamma_{\mathfrak{P}}(I) = T_{\mathfrak{P}^I}\uparrow^{\omega}(\mathbf{0})$ and $\mathfrak{G}_{\mathscr{P}}(J) = \mathfrak{T}_{\mathscr{P}^{\mathfrak{I}}}\uparrow^{\omega}(\bot)$.

In case that $\alpha = 0$, it follows that $\lambda^p(T_{\mathfrak{P}^I}\uparrow^{0}(\mathbf{0})(H)) = \mathfrak{T}_{\mathscr{P}^{\mathfrak{I}}}\uparrow^{0}(\bot)(H) = 0$ for every atom H. We assume as induction hypothesis that $\lambda^p(T_{\mathfrak{P}^I}\uparrow^{\beta}(\mathbf{0})) \leq \mathfrak{T}_{\mathscr{P}^{\mathfrak{I}}}\uparrow^{\beta}(\bot)$ for all $\beta < \alpha$.

In case that α is a successor ordinal. Assume that $u \leq T_{\mathfrak{P}^I}\uparrow^{\alpha-1}(\mathbf{0})(H)$ for some join irreducible u and atom H. Then there is a rule $r_i \in P$ of the form (4) and

$$u \;\leq\; (u_{B_1} * \ldots * u_{B_1} * u_{C_1} * \ldots * u_{C_1}) \cdot r_i$$

where $u_{B_j} \leq T_{\mathfrak{P}^I}\uparrow^{\alpha-1}(\mathbf{0})(B_j)$ and $u_{C_j} \leq {\sim}I(C_j)$. Hence, by induction hypothesis, it follows that $\lambda^p(u_{B_j}) \leq \mathfrak{T}_{\mathscr{P}^{\mathfrak{I}}}\uparrow^{\alpha-1}(\bot)(B_j)$ and, since $u_{C_j} \leq {\sim}I(C_j)$, it also follows that $\lambda^p(u_{C_j}) \leq \neg\mathfrak{I}(C_j)$ for all C_j. Consequently, we have that $\lambda^p(u) \leq \mathfrak{T}_{\mathscr{P}^{\mathfrak{I}}}\uparrow^{\alpha}(\bot)(H)$.

In case that α is a limit ordinal, $u \leq T_{\mathfrak{P}^I}\uparrow^{\alpha}(\mathbf{0})$ iff $u \leq T_{\mathfrak{P}^I}\uparrow^{\beta}(\mathbf{0})$ for some $\beta < \alpha$ and all join irreducible u. Hence, by induction hypothesis, it follows that $\lambda^p(u) \leq T_{\mathscr{P}^{\mathfrak{I}}}\uparrow^{\beta}(\mathbf{0}) \leq T_{\mathscr{P}^{\mathfrak{I}}}\uparrow^{\alpha}(\mathbf{0})$ and, thus, $\lambda^p(T_{\mathfrak{P}^I}\uparrow^{\alpha}(\mathbf{0})) \leq \mathfrak{T}_{\mathscr{P}^{\mathfrak{I}}}\uparrow^{\alpha}(\bot)$. □

Lemma B.9
Let P be a labelled logic program over a signature $\langle At, Lb\rangle$ where Lb is a finite set of labels and let I and $\mathfrak{I}$ be respectively a ECJ and a WnP interpretation such that $\lambda^p(I) \leq \mathfrak{I}$. Therefore, $\lambda^p(\Gamma_{\mathfrak{P}}(I)) \geq \mathfrak{G}_{\mathscr{P}}(\mathfrak{I})$. □

Proof. The proof is similar to the proof of Lemma B.8 and we just show the case in which α is a successor ordinal.

Assume that $u \leq \mathfrak{T}_{\mathscr{P}^{\mathfrak{I}}}\uparrow^{\alpha}(\bot)(H)$ for some join irreducible u and atom H. Hence, there is some rule $r_i \in P$ of the form (4) and

$$u \;\leq\; u_{B_1} * \ldots * u_{B_m} * u_{C_1} * \ldots * u_{C_n} * r_i$$

where $u_{B_j} \leq \mathfrak{T}_{\mathscr{P}^{\mathfrak{I}}}\uparrow^{\alpha-1}(\bot)(B_j)$ for each B_j and $u_{C_j} \leq \neg\mathfrak{I}(C_j)$ for each C_j. By induction hypothesis, $u_{B_j} \leq \lambda^p(T_{\mathfrak{P}^I}\uparrow^{\alpha-1}(\mathbf{0}))(B_j)$ for all B_j. Furthermore, since $\lambda^p(I) \leq \mathfrak{I}$ it follows, from Lemma B.3, that $\lambda^p({\sim}I) \geq \neg\mathfrak{I}$ and, since $u_{C_j} \leq \neg\mathfrak{I}(C_j)$, it also follows that $u_{C_j} \leq \lambda^p({\sim}I(C_j))$. Hence,

$$\lambda(u) \;\leq\; (\lambda^p(u_{B_1}) * \ldots * \lambda^p(u_{B_1}) * \lambda^p(u_{C_1}) * \ldots * \lambda^p(u_{C_1})) * r_i \;\;\leq\;\; \lambda^p(T_{\mathfrak{P}^I}\uparrow^{\alpha}(\mathbf{0})(H))$$

Thus, $\mathfrak{T}_{\mathscr{P}^{\mathfrak{I}}}\uparrow^{\alpha}(\bot)(B_j) \leq \lambda^p(T_{\mathfrak{P}^I}\uparrow^{\alpha}(\mathbf{0})(B_j))$. □

Note that the image of λ^p is a boolean algebra and the set of causal values corresponding to negated terms $\{\, {\sim}t \mid t \in \mathbf{V}_{Lb} \,\}$ are also a boolean algebra. Consequently, we define a function $\lambda^q(t) = {\sim}{\sim}t$ which is analogous to λ^p but whose image is in $\mathbf{V}_{Lb}$.

Lemma B.10
Let P be a labelled logic program and let I be an ECJ interpretation. Then, $\Gamma_{\mathfrak{P}}(I) = \Gamma_{\mathfrak{P}}(\lambda^q(I))$ and $\lambda^p(t) = \lambda^p(\lambda^q(t))$. □

Proof. For $\Gamma_{\mathfrak{P}}(I) = \Gamma_{\mathfrak{P}}(\lambda^q(I))$. Since $\lambda^q(t) = {\sim}{\sim}t$ and ${\sim}{\sim}{\sim}t = {\sim}t$, it follows that $\lambda^q({\sim}I) = {\sim}{\sim}{\sim}I = {\sim}I$ and, thus, $\mathfrak{P}^I = \mathfrak{P}^{\lambda^q(I)}$. Since by definition $\Gamma_{\mathfrak{P}}(I)$ and $\Gamma_{\mathfrak{P}}(\lambda^q(I))$ are respectively the least models of programs $\mathfrak{P}^I$ and $\mathfrak{P}^{\lambda^q(I)}$ it is clear that $\Gamma_{\mathfrak{P}}(I) = \Gamma_{\mathfrak{P}}(\lambda^q(I))$.

For $\lambda^p(t) = \lambda^p(\lambda^q(t))$, just note $\lambda^p(\lambda^q(t)) = \lambda^p({\sim}{\sim}t) = \neg\neg\lambda^p(t) = \lambda^p(t)$. □

Proposition 7
Let P be a program over a signature $\langle At, Lb\rangle$ where Lb is a finite set of labels. Then, any causal interpretation I satisfies:

(i). $\mathfrak{G}_{\mathscr{P}}(\lambda^p(I)) \ = \ \lambda^p(\Gamma_{\mathfrak{P}}(I))$,
(ii). $\Gamma_{\mathfrak{P}}(\lambda^q(I)) \ = \ \Gamma_{\mathfrak{P}}(I)$ and
(iii). $\lambda^p(t) = \lambda^p(\lambda^q(t))$. □

Proof. (i) From Lemmas B.8 and B.9, it respectively follows that $\lambda^p(\Gamma_{\mathfrak{P}}(I)) \leq \mathfrak{G}_{\mathscr{P}}(\lambda^p(I))$ and that $\lambda^p(\Gamma_{\mathfrak{P}}(I)) \geq \mathfrak{G}_{\mathscr{P}}(\lambda^p(I))$. Then, $\mathfrak{G}_P(\lambda^p(I)) = \lambda^p(\Gamma_{\mathscr{P}}(I))$. (ii) and (iii) follow from Lemma B.10. □

Proof of Theorem 7. Note that $Why_{\mathscr{P}}(A) = \mathfrak{T}_{\mathscr{P}}(A)$ and that, by λ^p definition, it follows that $\lambda^p(\mathbf{0}) = \mathbf{0}$ and thus, from Proposition 7 (i), it follows that $\mathfrak{G}_{\mathscr{P}}(\bot) = \mathfrak{G}_{\mathscr{P}}(\lambda^p(\mathbf{0})) = \lambda^p(\Gamma_{\mathfrak{P}}(\mathbf{0}))$ and

$$\mathfrak{G}_{\mathscr{P}}(\bot) \ = \ \mathfrak{G}_{\mathscr{P}}(\lambda^p(\mathbf{0})) \ = \ \lambda^p(\Gamma_{\mathscr{P}}(\mathbf{0})) \ = \ \lambda^p(\lambda^q(\Gamma_{\mathscr{P}}(\mathbf{0})))$$

Hence, from Proposition 7, it follows that

$$\begin{aligned} \mathfrak{G}^2_{\mathscr{P}}(\bot) \ &= \ \mathfrak{G}_{\mathscr{P}}(\mathfrak{G}_{\mathscr{P}}(\bot)) \ = \ \mathfrak{G}_{\mathscr{P}}(\lambda^p(\lambda^q(\Gamma_{\mathscr{P}}(\mathbf{0})))) \\ &= \ \lambda^p(\Gamma_{\mathfrak{P}}(\lambda^q(\Gamma_{\mathfrak{P}}(\mathbf{0})))) \ = \ \lambda^p(\Gamma_{\mathfrak{P}}(\Gamma_{\mathfrak{P}}(\mathbf{0}))) \ = \ \lambda^p(\Gamma^2_{\mathfrak{P}}(\mathbf{0})) \end{aligned}$$

Inductively applying this reasoning it follows that $\mathfrak{G}^2_{\mathscr{P}}\!\uparrow^\infty(\mathbf{0}) = \lambda^p(\Gamma^2_{\mathfrak{P}}\!\uparrow^\infty(\mathbf{0}))$ which, by Knaster-Tarski theorem are the least fixpoints of the operators, that is, $\mathfrak{T}_{\mathscr{P}} = \lambda^p(\mathbb{L}_{\mathfrak{P}})$ and, consequently, $Why_{\mathscr{P}}(A) = \mathfrak{T}_{\mathscr{P}}(A) = \lambda^p(\mathbb{L}_{\mathfrak{P}}(A)) = \lambda^p(\mathbb{W}_{\mathfrak{P}}(A)) = Why_P(A)$. Similarly, by definition, it follows that $Why_{\mathscr{P}}(not A) = \neg\mathfrak{TU}_{\mathscr{P}}(A)$ where $\mathfrak{TU}_{\mathscr{P}}$ is the greatest fixpoint of the operator $\mathfrak{G}^2_{\mathscr{P}}$. Thus,

$$Why_{\mathscr{P}}(not A) \ = \ \neg\mathfrak{G}_{\mathscr{P}}(\mathfrak{T}_{\mathscr{P}}) \ = \ \lambda^p({\sim}\Gamma_{\mathfrak{P}}(\mathbb{L}_{\mathfrak{P}})) \ = \ \lambda^p({\sim}\mathbb{U}_{\mathfrak{P}}(A)) \ = \ \lambda^p(\mathbb{W}_{\mathfrak{P}}(not A))$$

Finally, $Why_{\mathscr{P}}(undef A) \ = \ \neg\mathfrak{T}_{\mathscr{P}}(A) * \mathfrak{TU}_{\mathscr{P}}(A)$ and, thus

$$\begin{aligned} Why_{\mathscr{P}}(undef A) \ &= \ \lambda^p({\sim}\mathbb{L}_{\mathfrak{P}}(A)) * \lambda^p({\sim}{\sim}\mathbb{U}_{\mathfrak{P}}(A)) \\ &= \ \lambda^p({\sim}\mathbb{L}_{\mathfrak{P}}(A) * {\sim}{\sim}\mathbb{U}_{\mathfrak{P}}(A)) \\ &= \ \lambda^p({\sim}\mathbb{W}_{\mathfrak{P}}(A) * {\sim}\mathbb{W}_{\mathfrak{P}}(not A)) \ = \ \lambda^p(\mathbb{W}_{\mathfrak{P}}(undef A)) \end{aligned}$$

and, thus, $Why_{\mathscr{P}}(undef A) = \lambda^p(\mathbb{W}_{\mathfrak{P}}(undef A)) = Why_P(not A)$. □

Appendix B.9. Proof of Theorem 8

Lemma B.11
Let P be a labelled logic program over a signature $\langle At, Lb\rangle$ where Lb is a finite set of labels and no rule is a labelled by $not(A)$ nor $\sim\sim not(A)$. Let Q be the result of removing all rules labelled by $\sim not(A)$ for some atom A. Let I and J be two interpretations such that $J = \rho_{not(A)}(I)$. Then, $\Gamma_Q(J) = \rho_{not(A)}(\Gamma_P(I))$. □

Proof. In the sake of simplicity, we just write ρ instead of $\rho_{not(A)}$. By definition $\Gamma_P(I)$ and $\Gamma_Q(J)$ are respectively the least model of P^I and Q^J. The proof follows then by induction on the steps of the T_P operator assuming that $\rho(T_{P^I}\uparrow^{\beta}(\mathbf{0})) = T_{Q^J}\uparrow^{\beta}(\mathbf{0})$ for all $\beta < \alpha$.

Note that, $T_X\uparrow^0(\mathbf{0})(H) = 0$ for any program X and atom H and, thus, the statement trivially holds.

In case that α is a successor ordinal. Let $u \in \mathbf{V}_{Lb}$ be a join irreducible causal value such that $u \leq T_{P^I}\uparrow^{\alpha}(\mathbf{0})(H)$. Then, there is a rule in P of the form (4) such that

$$u \;\; \leq \;\; (u_{B_1} * \ldots * u_{B_m} * u_{C_1} * \ldots * u_{C_n}) \cdot r_i$$

where $u_{B_j} \leq T_{P^I}\uparrow^{\alpha-1}(\mathbf{0})(B_j)$ and $u_{C_j} \leq \sim I(C_j)$ for each positive literal B_j and each negative literal $not\, C_j$ in the body of rule r_i.

If $r_i = \sim not(A)$, then $\rho(u) = 0 \leq T_Q\uparrow^{\alpha-1}(\mathbf{0})(H)$. Otherwise,

1. By induction hypothesis, it follows that $\rho(u_{B_j}) \leq T_Q\uparrow^{\alpha-1}(\mathbf{0})(B_j)$, and
2. from $J(H) = \rho(I(H))$ and $u_{C_j} \leq \sim I(C_j)$, it follows that $\rho(u_{C_j}) \leq \sim J(C_j)$.

Furthermore, no rule in the program P is labelled with $not(A)$ nor $\sim\sim not(A)$ and, thus, $r_i \neq not(A)$ and $r_i \neq \sim\sim not(A)$. Hence, $\rho(u) \leq T_Q\uparrow^{\alpha-1}(\mathbf{0})(H)$.

The other way around is similar. Since $u \leq T_{Q^J}\uparrow^{\alpha}(\mathbf{0})(H)$ there is a rule in Q of the form (4) such that

$$u \;\; \leq \;\; (u_{B_1} * \ldots * u_{B_m} * u_{C_1} * \ldots * u_{C_n}) \cdot r_i$$

and $u_{B_j} \leq T_{Q^J}\uparrow^{\alpha-1}(\mathbf{0})(B_j)$ and $u_{C_j} \leq \sim J(C_j)$ for each positive literal B_j and each negative literal $not\, C_j$ in the body of rule r_i. By induction hypothesis, $u_{B_j} \leq \rho(T_{P^I}\uparrow^{\alpha-1}(\mathbf{0})(B_j))$ for each B_j with $1 \leq j \leq m$ and, since $J(H) = \rho(I(H))$ and $u_{C_j} \leq \sim J(C_j)$, it follows that $u_{C_j} \leq \rho(\sim I(C_j))$. Then, $u \leq \rho(T_{P^I}\uparrow^{\alpha}(\mathbf{0})(H))$.

In case that α is a limit ordinal $T_X\uparrow^{\alpha}(\mathbf{0}) = \sum_{\beta<\alpha} T_X\uparrow^{\beta}(\mathbf{0})(H)$ and, thus, $u \leq T_X\uparrow^{\alpha}(\mathbf{0})$ if and only if $u \leq T_X\uparrow^{\beta}(\mathbf{0})(H)$ with $\beta < \alpha$. By induction hypothesis, $\rho(T_{P^I}\uparrow^{\beta}(\mathbf{0})(H)) = T_{Q^J}\uparrow^{\beta}(\mathbf{0})(H)$ and, thus, $u \leq \rho(T_{P^I}\uparrow^{\alpha}(\mathbf{0}))$ if and only if $u \leq T_{Q^J}\uparrow^{\alpha}(\mathbf{0})$. Hence, $\rho(T_{P^I}\uparrow^{\alpha}(\mathbf{0})) = T_{Q^J}\uparrow^{\alpha}(\mathbf{0})$ and, consequently, $\Gamma_Q(J) = \rho(\Gamma_P(I))$. □

Proposition 8
Let P be a labelled logic program over a signature $\langle At, Lb\rangle$ where Lb is a finite set of labels where no rule is a labelled by $not(A)$ nor $\sim\sim not(A)$. Let Q be the result of removing all rules labelled by $\sim not(A)$ for some atom A. Then, $\mathbb{L}_Q = \rho_{not(A)}(\mathbb{L}_P)$ and $\mathbb{U}_Q = \rho_{not(A)}(\mathbb{U}_P)$. □

Proof. Note that $\mathbb{L}_X = \Gamma_X^2 \uparrow^{\infty} (\mathbf{0})$ with $X \in \{P, Q\}$. Furthermore, by definition, it follows that $\Gamma_P^2 \uparrow^0 (\mathbf{0}) = \Gamma_Q^2 \uparrow^0 (\mathbf{0}) = 0$. Then, assume as induction hypothesis that $\Gamma_Q^2 \uparrow^\beta (\mathbf{0}) = \rho(\Gamma_P^2 \uparrow^\beta (\mathbf{0}))$ for all $\beta < \alpha$. When α is a successor ordinal, by definition $\Gamma_X^2 \uparrow^\alpha (\mathbf{0}) = \Gamma_X^2(\Gamma_X^2 \uparrow^{\alpha-1} (\mathbf{0})) = \Gamma_X(\Gamma_X(\Gamma_X^2 \uparrow^{\alpha-1} (\mathbf{0})))$ with $X \in \{P, Q\}$ and, thus, the statement follows from Lemma B.11.

In case that α is a limit ordinal $\Gamma_X^2 \uparrow^\alpha (\mathbf{0}) = \sum_{\beta<\alpha} \Gamma_X^2 \uparrow^\beta (\mathbf{0})$. Then, for every join irreducible u it follows that $u \leq \Gamma_P^2 \uparrow^\alpha (\mathbf{0})$ if and only if $u \leq \Gamma_P^2 \uparrow^\beta (\mathbf{0})$ for some $\beta < \alpha$ (by induction hypothesis) iff $\rho(u) \leq \Gamma_P^2 \uparrow^\beta (\mathbf{0})$ iff $\rho(u) \leq \Gamma_P^2 \uparrow^\alpha (\mathbf{0})$. Hence, $\Gamma_Q^2 \uparrow^\alpha (\mathbf{0}) = \rho(\Gamma_P^2 \uparrow^\alpha (\mathbf{0}))$ and, conseuqntly, $\mathbb{L}_Q = \rho(\mathbb{L}_P)$

Finally, note that $\mathbb{U}_X = \Gamma_X(\mathbb{L}_X)$ with $X \in \{P, Q\}$ and, thus, the statement follows directly from Lemma B.11. □

Proof of Theorem 8. By definition, program P is the result of removing all rules labelled with $\sim not(A)$ in $\mathfrak{P}$. In case that L is some atom H, by definition, it follows that $\mathbb{W}_P(H) = \mathbb{L}_P(H)$ and $\mathbb{W}_{\mathfrak{P}}(H) = \mathbb{L}_{\mathfrak{P}}(H)$ and, from Proposition 8, it follows that $\mathbb{L}_P = \rho(\mathbb{L}_{\mathfrak{P}})$ and, thus $\mathbb{W}_P = \rho(\mathbb{W}_{\mathfrak{P}})$.

Similarly, in case that L is a negative literal ($L = not\,H$), then $\mathbb{W}_P(H) = \sim\mathbb{U}_P(H)$ and $\mathbb{W}_{\mathfrak{P}}(H) = \sim\mathbb{U}_{\mathfrak{P}}(H)$ and, from Proposition 8, it follows that $\mathbb{U}_P = \rho(\mathbb{U}_{\mathfrak{P}})$. Just note tha $\rho_x(\sim u) = \sim\rho_x(u)$ for any elementary term x and any value u. Hence, $\mathbb{U}_P = \rho(\mathbb{U}_{\mathfrak{P}})$ implies that $\sim\mathbb{U}_P = \rho(\sim\mathbb{U}_{\mathfrak{P}})$ and, consequently, $\mathbb{W}_P = \rho(\mathbb{W}_{\mathfrak{P}})$.

In case that L is an undefined literal ($L = undef\,H$), by definition, it follows that $\mathbb{W}_P(H) = \sim\mathbb{W}_P(H) * \sim\mathbb{W}_P(not\,H) = \sim\mathbb{L}_P(H) * \sim\sim\mathbb{U}_P(H)$ and $\mathbb{W}_{\mathfrak{P}}(H) = \sim\mathbb{L}_{\mathfrak{P}}(H) * \sim\sim\mathbb{U}_{\mathfrak{P}}(H)$ and the result follows as before from Proposition 8. □

Appendix B.10. Proof of Theorem 9

Proof of Theorem 9. Note that $\rho(\lambda^p(u)) = \lambda^p(\rho(u))$ for any causal value $u \in \mathbf{V}_{Lb}$. By definition $Why_P(L) = \lambda^p(\mathbb{W}_{\mathfrak{P}})(L)$ and, thus

$$\rho(Why_P(L)) \;=\; \rho(\lambda^p(\mathbb{W}_{\mathfrak{P}})(L)) \;=\; \lambda^p(\rho(\mathbb{W}_{\mathfrak{P}}))(L)$$

From Theorem 8, it follows that $\mathbb{W}_P = \rho(\mathbb{W}_{\mathfrak{P}})$ and, thus, $\rho(Why_P(L)) = \lambda^p(\mathbb{W}_P)(L)$. □

Appendix B.11. Proof of Theorem 2

The proof of Theorem 2 will rely on the relation between ECJ justifications and non-hypothetical WnP justifications established by Theorem 9 plus the following result from (Damásio et al. 2013). First, we need some notation. Given a conjuntion of labels D, by $Remove(D)$ we denote the set of negated labels in D, by $Keep(D)$ the set of positive labels, by $AddFacts(D)$ the set of facts A such that $\neg not(A)$ occurs in D and by $NoFacts(D)$ the set of facts A such that $not(A)$ occurs in D.

Theorem 12 (*Theorem 3 from Damásio et al. 2013*)
Given a labelled logic program P, let N be a set of facts not in program P and R be a subset of rules of P. A literal L belongs to the WFM of $(P \backslash R) \cup N$ iff there is a conjunction of literals $D \models Why_P(L)$, such that $Remove(D) \subseteq R$, $Keep(D) \cap R = \emptyset$, $AddFacts(D) \subseteq N$, and $NoFacts(D) \cap N = \emptyset$. □

Definition 21
Given a positive program P, we define a direct consequence operator $\hat{T}_P$ such that

$$\hat{T}_P(\hat{I})(H) \stackrel{\text{def}}{=} \sum \left\{ \hat{I}(B_1) * \ldots * \hat{I}(B_n) \mid (r_i : H \leftarrow B_1, \ldots, B_n) \in P \right\}$$

for any standard interpretation interpretation $\hat{I}$ and atom $H \in At$. □

Lemma B.12
Let P be a labelled logic program over a signature $\langle At, Lb \rangle$ where Lb is a finite set of labels and let I and $\hat{I}$ be respectively a ECJ and a standard interpretation satisfying that there is some enable justification $E \leq {\sim}I(H)$ for every atom H such that $\hat{I}(H) = 0$. Then, every atom H satisfies $\hat{\Gamma}_P(\hat{I})(H) = 1$ iff there is some enabled justification $E \leq \Gamma_P(I)(H)$. □

Proof. By definition $\Gamma_P(I)$ and $\hat{\Gamma}_P(\hat{I})$ are the least model of the programs P^I and $P^{\hat{I}}$, respectively. Furthermore, the least model of programs P^I and $P^{\hat{I}}$ are the least fixpoint of the T_P and $\hat{T}_P$ operators, that is, $\Gamma_P(I) = T_{P^I}\!\uparrow^{\omega}(\mathbf{0})$ and $\hat{\Gamma}_P(J) = \hat{T}_{P^{\hat{I}}}\!\uparrow^{\omega}(\mathbf{0})$. In case that $\alpha = 0$, it follows that $\hat{T}_{P^{\hat{I}}}\!\uparrow^{0}(\mathbf{0})(H)$ for every atom H and, thus, the statement holds vacuous. We assume as induction hypothesis that for every atom H and ordinal $\beta < \alpha$ such that $\hat{T}_{P^{\hat{I}}}\!\uparrow^{\beta}(\mathbf{0})(H) = 1$, there is some enabled justification $E \leq T_{P^I}\!\uparrow^{\beta}(\mathbf{0})(H)$.

In case that α is a successor ordinal. If $\hat{T}_{P^I}\!\uparrow^{\alpha-1}(\mathbf{0})(H) = 1$, then there is a rule $r_i \in P$ of the form (4) such that $\hat{T}_{P^I}\!\uparrow^{\alpha-1}(\mathbf{0})(B_j) = 1$ and $I(C_j) = 0$. On the one hand, by induction hypothesis, it follows that there is some enabled justification $E_{B_j} \leq T_{P^I}\!\uparrow^{\alpha-1}(\mathbf{0})(B_j)$ and, by hypothesis, there is some enabled justification $E_{C_j} \leq {\sim}I(C_j)$. Hence,

$$E \stackrel{\text{def}}{=} (E_{B_1} * \ldots E_{B_m} * E_{C_1} * \ldots * E_{C_n}) \cdot r_i$$

is an enabled justification $E \leq T_{P^I}\!\uparrow^{\alpha}(\mathbf{0})(H)$.

The other way around, let E be some join irreducible justification. If $E \leq T_{P^I}\!\uparrow^{\alpha}(\mathbf{0})(H)$, then there is a rule $r_i \in P$ of the form (4) such that

$$E \leq (E_{B_1} * \ldots E_{B_m} * E_{C_1} * \ldots * E_{C_n}) \cdot r_i$$

where $E_{B_j} \leq T_{P^I}\!\uparrow^{\alpha}(\mathbf{0})(B_j)$ and $E_{C_j} \leq {\sim}I(C_j)$ are enabled justifications. Hence, it follows that $\hat{T}_{P^{\hat{I}}}\!\uparrow^{\alpha}(\mathbf{0})(B_j) = 1$ and $\hat{I}(C_j) = 0$.

In case that α is a limit ordinal, $\hat{T}_{P^{\tilde{I}}}\!\uparrow^{\alpha}(\mathbf{0}) = 1$ iff $\hat{T}_{P^{\tilde{I}}}\!\uparrow^{\beta}(\mathbf{0}) = 1$ for some $\beta < \alpha$ iff there is a join irreducible enabled justification $E \leq T_{P^I}\!\uparrow^{\beta}(\mathbf{0})) \leq \lambda^p(T_{P^I}\!\uparrow^{\alpha}(\mathbf{0})$. □

Proof of Theorem 2. Let $E \leq \mathbb{W}_P(L)$ be an enabled justification of $L \in \{A, \mathit{not}\,A, \mathit{undef}\,A\}$. From Theorem 9, it follows that $\lambda^p(E) \leq \lambda^p(\mathbb{W}_P(L)) = \rho(Why_P(L))$, that is, $\lambda^p(E) \leq \rho(Why_P(L))$. Note that the minimum causal value t such that $\rho(t) = \rho(Why_P(L))$ is $Why_P(L) \wedge \bigwedge_{A \in At} \mathit{not}(A)$ and, thus, $D \leq Why_P(L)$ where D is defined by $D = \lambda^p(E) \wedge \bigwedge_{A \in At} \mathit{not}(A)$. Furthermore, since E is an enabled justification, $\lambda^p(E)$ is a positive conjunction and, thus, so it is D. Hence, there is a positive conjunction D such that $D \leq Why_P(L)$ and, from Theorem 12, it follows that L holds with respect to the standard WFM of P.

The other way around. If $L = A$ is an atom, then L holds with respect to the standard WFM iff $\text{lfp}(\hat{\Gamma}_P^2)(L) = 1$. Furthermore, $\hat{\Gamma}_P^2\!\uparrow^{0}(\mathbf{0})(H) = \Gamma_P^2\!\uparrow^{0}(\mathbf{0}) = 0$ for any atom H and, thus, there is an enabled justification $E \leq \sim\Gamma_P^2\uparrow^0(\mathbf{0}) = \sim 0 = 1$ for any atom H. Then, from Lemma B.12, for any atom H , there is an enabled justification $E \leq \Gamma_P(\Gamma_P^2\uparrow^0(\mathbf{0}))(H)$ iff $\hat{\Gamma}_P(\hat{\Gamma}_P^2\uparrow^0(\mathbf{0}))(H) = 1$. Applying this result again, it follows that $E \leq \Gamma_P^2\uparrow^1(\mathbf{0})(H) = \Gamma_P^2(\Gamma_P^2\uparrow^0(\mathbf{0}))(H)$ if and only if $\hat{\Gamma}_P^2\uparrow^1(\mathbf{0}))(H) = \hat{\Gamma}_P^2(\hat{\Gamma}_P^2\uparrow^0(\mathbf{0}))(H) = 1$. Inductively applying this reasoning it follows that $\hat{\Gamma}_P^2\uparrow^\infty(\mathbf{0})(H) = 1$ iff there is an enabled justification $E \leq \Gamma_P^2\uparrow^\infty(\mathbf{0})(H)$ which, by Knaster-Tarski theorem are the least fixpoints respectively of the $\hat{\Gamma}_P$ and Γ_P operators.

Similarly, if $L = not\,A$, then L holds with respect to the standard WFM if and only if $\mathrm{gfp}(\hat{\Gamma}_P^2)(L) = \hat{\Gamma}_P(\mathrm{lfp}(\hat{\Gamma}_P^2))(L) = 0$ iff there is not any an enabled justification $E \leq \Gamma_P(\mathrm{lfp}(\Gamma_P^2))(L) = \mathrm{gfp}(\Gamma_P^2)(L)$ iff there is an enabled justification $E \leq \mathbb{W}_P(L) = \sim\mathrm{gfp}(\Gamma_P^2)(L)$.

Finally, if $L = undef\,A$, then L holds with respect to the standard WFM iff $\mathrm{lfp}(\hat{\Gamma}_P^2)(L) = 0$ and $\mathrm{gfp}(\hat{\Gamma}_P^2)(L) = 1$ if and only if there is not any enabled justification $E \leq \mathbb{W}_P(L)$ and there is not any enabled justification $E \leq \mathbb{W}_P(not\,L)$ iff there is some enabled justification $E \leq \sim\mathbb{W}_P(L)$ and there is some enabled justification $E \leq \sim\mathbb{W}_P(not\,L)$ iff there is some enabled justification $\mathbb{W}_P(undef\,A) = \sim\mathbb{W}_P(A) * \sim\mathbb{W}_P(not\,A)$. □

Appendix B.12. Proof of Theorem 10

Lemma B.13

Let t and u be two causal terms such that no-sums occur in t ant $t \leq u$. Then, $\rho_x(t) \leq \rho_x(u)$. □

Proof. By definition $t \leq u$ if and only if $t = t * u$. Then, $\rho_x(t) = \rho_x(t * u) = \rho_x(t) * \rho_x(u)$ and, thus if follows that $\rho_x(t) \leq \rho_x(u)$. □

Lemma B.14

Let t be a causal term. Then, $\lambda^c(\lambda^p(t)) \leq \lambda^p(\lambda^c(t))$. □

Proof. If $t \in Lb$ is a label, then $\lambda^c(t) = t$ and $\lambda^p(t) = t$ and, thus, $\lambda^c(\lambda^p(t)) = t \leq t = \lambda^p(\lambda^c(t))$. If $t = \sim l$ with $l \in Lb$ a label, then $\lambda^c(t) = 0$ and $\lambda^p(t) = \neg l$ and, thus, $\lambda^c(\lambda^p(t)) = 0 \leq 0 = \lambda^p(\lambda^c(t))$. If $t = \sim\sim l$ with $l \in Lb$ a label, then $\lambda^c(t) = 1$ and $\lambda^p(t) = l$ and, thus, $\lambda^c(\lambda^p(t)) = l \leq 1 = \lambda^p(\lambda^c(t))$.

Assume as induction hypothesis that $\lambda^c(\lambda^p(u)) \leq \lambda^p(\lambda^c(u))$ for every subterm u of t. If $t = u_1 \cdot u_2$, then

$$\lambda^c(\lambda^p(u_1 \cdot u_2)) \;=\; \lambda^c(\lambda^p(u_1) * \lambda^c(\lambda^p(u_2) \;\leq\; \lambda^p(\lambda^c(u_1) * \lambda^p(\lambda^c(u_2) \;=\; \lambda^p(\lambda^c(u_1 \cdot u_2))$$

Similarly, if $t = \sum_{u \in U} u$, then

$$\lambda^c(\lambda^p(\sum_{u \in U} u) \;=\; \sum_{u \in U} \lambda^c(\lambda^p(u) \;\leq\; \sum_{u \in U} \lambda^p(\lambda^c((u)) \;=\; \lambda^p(\lambda^c(\sum_{u \in U} u))$$

and if $t = \prod_{u \in U} u$, then

$$\lambda^c(\lambda^p(\prod_{u \in U} u) \;=\; \prod_{u \in U} \lambda^c(\lambda^p(u) \;\leq\; \prod_{u \in U} \lambda^p(\lambda^c((u)) \;=\; \lambda^p(\lambda^c(\prod_{u \in U} u))$$

□

Proof of Theorem 10. From Theorem 9, it follows that $\rho(Why_P(A)) = \lambda^p(\mathbb{W}_P)(A)$. Furthermore, since $D \leq Why_P(A)$, from Lemma B.13, it follows that

$$\rho(D) \;\leq\; \rho(Why_P(A)) \;=\; \lambda^p(\mathbb{W}_P)(A) \;=\; \lambda^p(\mathbb{L}_P)(A)$$

and, thus, $\lambda^c(\rho(D)) \leq \lambda^c(\lambda^p(\mathbb{L}_P))(A)$. Let $\tilde{I}$ be any CG stable model. Then, since $\tilde{I} = \lambda^c(I)$ for some fixpoint I of Γ^2_P, it follows that $\lambda^c(\mathbb{L}_P) \leq \tilde{I}$ and, thus, $\lambda^p(\lambda^c(\mathbb{L}_P)) \leq \lambda^p(\tilde{I})$. Furthermore, from Lemma B.14, it follows that $\lambda^c(\lambda^p(\mathbb{L}_P)) \leq \lambda^p(\lambda^c(\mathbb{L}_P))$ and, thus

$$\lambda^c(\rho(D)) \;\leq\; \lambda^c(\lambda^p(\mathbb{L}_P))(A) \;\leq\; \lambda^p(\lambda^c(\mathbb{L}_P))(A) \;\leq\; \lambda^p(\tilde{I})(A)$$

Note that, since D is non-hypothetical and enabled, it does not contain negated labels and, thus, $\lambda^c(\rho(D)) = \rho(D)$. Consequently, $\rho(D) \leq \lambda^p(\tilde{I})(A)$. □