

Automating Public Announcement Logic and the Wise Men Puzzle in Isabelle/HOL

Christoph Benz Müller and Sebastian Reiche

October 13, 2025

Abstract

We present a shallow embedding of public announcement logic (PAL) with relativized general knowledge in HOL. We then use PAL to obtain an elegant encoding of the wise men puzzle, which we solve automatically using sledgehammer.

Contents

1 Public Announcement Logic (PAL) in HOL	1
2 Automating the Wise Men Puzzle	3

1 Public Announcement Logic (PAL) in HOL

An earlier encoding and automation of the wise men puzzle, utilizing a shallow embedding of higher-order (multi-)modal logic in HOL, has been presented in [1, 2]. However, this work did not convincingly address the interaction dynamics between the involved agents. Here we therefore extend and adapt the universal (meta-)logical reasoning approach of [1] for public announcement logic (PAL) and we demonstrate how it can be utilized to achieve a convincing encoding and automation of the wise men puzzle in HOL, so that also the interaction dynamics as given in the scenario is adequately addressed. For further background information on the work presented here we refer to [3, 4].

theory *PAL* **imports** *Main* **begin**
nitpick-params[*user-axioms, expect=genuine*]

Type *i* is associated with possible worlds

typedecl *i*
type-synonym $\sigma = i \Rightarrow \text{bool}$
type-synonym $\tau = \sigma \Rightarrow i \Rightarrow \text{bool}$
type-synonym $\alpha = i \Rightarrow i \Rightarrow \text{bool}$

type-synonym $\rho = \alpha \Rightarrow \text{bool}$

Some useful relations (for constraining accessibility relations)

definition *reflexive*:: $\alpha \Rightarrow \text{bool}$

where *reflexive* $R \equiv \forall x. R\ x\ x$

definition *symmetric*:: $\alpha \Rightarrow \text{bool}$

where *symmetric* $R \equiv \forall x\ y. R\ x\ y \longrightarrow R\ y\ x$

definition *transitive*:: $\alpha \Rightarrow \text{bool}$

where *transitive* $R \equiv \forall x\ y\ z. R\ x\ y \wedge R\ y\ z \longrightarrow R\ x\ z$

definition *euclidean*:: $\alpha \Rightarrow \text{bool}$

where *euclidean* $R \equiv \forall x\ y\ z. R\ x\ y \wedge R\ x\ z \longrightarrow R\ y\ z$

definition *intersection-rel*:: $\alpha \Rightarrow \alpha \Rightarrow \alpha$

where *intersection-rel* $R\ Q \equiv \lambda u\ v. R\ u\ v \wedge Q\ u\ v$

definition *union-rel*:: $\alpha \Rightarrow \alpha \Rightarrow \alpha$

where *union-rel* $R\ Q \equiv \lambda u\ v. R\ u\ v \vee Q\ u\ v$

definition *sub-rel*:: $\alpha \Rightarrow \alpha \Rightarrow \text{bool}$

where *sub-rel* $R\ Q \equiv \forall u\ v. R\ u\ v \longrightarrow Q\ u\ v$

definition *inverse-rel*:: $\alpha \Rightarrow \alpha$

where *inverse-rel* $R \equiv \lambda u\ v. R\ v\ u$

definition *big-union-rel*:: $\rho \Rightarrow \alpha$

where *big-union-rel* $X \equiv \lambda u\ v. \exists R. (X\ R) \wedge (R\ u\ v)$

definition *big-intersection-rel*:: $\rho \Rightarrow \alpha$

where *big-intersection-rel* $X \equiv \lambda u\ v. \forall R. (X\ R) \longrightarrow (R\ u\ v)$

In HOL the transitive closure of a relation can be defined in a single line.

definition *tc*:: $\alpha \Rightarrow \alpha$

where *tc* $R \equiv \lambda x\ y. \forall Q. \text{transitive}\ Q \longrightarrow (\text{sub-rel}\ R\ Q \longrightarrow Q\ x\ y)$

Logical connectives for PAL

abbreviation *patom*:: $\sigma \Rightarrow \tau$ ($\langle^A \neg \rangle$ [79]80)

where $^A p \equiv \lambda W\ w. W\ w \wedge p\ w$

abbreviation *ptop*:: τ ($\langle^T \rangle$)

where $\top \equiv \lambda W\ w. \text{True}$

abbreviation *pneg*:: $\tau \Rightarrow \tau$ ($\langle^{\neg} \neg \rangle$ [52]53)

where $\neg \varphi \equiv \lambda W\ w. \neg(\varphi\ W\ w)$

abbreviation *pand*:: $\tau \Rightarrow \tau \Rightarrow \tau$ (**infixr** $\langle^{\wedge} \rangle$ 51)

where $\varphi \wedge \psi \equiv \lambda W\ w. (\varphi\ W\ w) \wedge (\psi\ W\ w)$

abbreviation *por*:: $\tau \Rightarrow \tau \Rightarrow \tau$ (**infixr** $\langle^{\vee} \rangle$ 50)

where $\varphi \vee \psi \equiv \lambda W\ w. (\varphi\ W\ w) \vee (\psi\ W\ w)$

abbreviation *pimp*:: $\tau \Rightarrow \tau \Rightarrow \tau$ (**infixr** $\langle^{\rightarrow} \rangle$ 49)

where $\varphi \rightarrow \psi \equiv \lambda W\ w. (\varphi\ W\ w) \longrightarrow (\psi\ W\ w)$

abbreviation *pequ*:: $\tau \Rightarrow \tau \Rightarrow \tau$ (**infixr** $\langle^{\leftrightarrow} \rangle$ 48)

where $\varphi \leftrightarrow \psi \equiv \lambda W\ w. (\varphi\ W\ w) \longleftrightarrow (\psi\ W\ w)$

abbreviation *pknow*:: $\alpha \Rightarrow \tau \Rightarrow \tau$ ($\langle^{\mathbf{K}} \neg \rangle$)

where $\mathbf{K}\ r \equiv \lambda W\ w. \forall v. (W\ v \wedge r\ w\ v) \longrightarrow (\varphi\ W\ v)$

abbreviation *ppal*:: $\tau \Rightarrow \tau \Rightarrow \tau$ ($\langle^{\mathbf{!}} \neg \rangle$)

where $\mathbf{!}\varphi \equiv \lambda W\ w. (\varphi\ W\ w) \longrightarrow (\psi\ (\lambda z. W\ z \wedge \varphi\ W\ z)\ w)$

Global validity of PAL formulas

abbreviation *pvalid*:: $\tau \Rightarrow \text{bool}$ ($\langle^{\mathbf{!}} \neg \rangle$ [7]8)

where $[\varphi] \equiv \forall W. \forall w. W w \longrightarrow \varphi W w$

Introducing agent knowledge (K), mutual knowledge (E), distributed knowledge (D) and common knowledge (C).

abbreviation $EVR::\varrho \Rightarrow \alpha$

where $EVR G \equiv \text{big-union-rel } G$

abbreviation $DIS::\varrho \Rightarrow \alpha$

where $DIS G \equiv \text{big-intersection-rel } G$

abbreviation $agtknows::\alpha \Rightarrow \tau \Rightarrow \tau \ (\langle \mathbf{K}_- \rangle)$

where $\mathbf{K}_r \varphi \equiv \mathbf{K} r \varphi$

abbreviation $evrknows::\varrho \Rightarrow \tau \Rightarrow \tau \ (\langle \mathbf{E}_- \rangle)$

where $\mathbf{E}_G \varphi \equiv \mathbf{K} (EVR G) \varphi$

abbreviation $disknows :: \varrho \Rightarrow \tau \Rightarrow \tau \ (\langle \mathbf{D}_- \rangle)$

where $\mathbf{D}_G \varphi \equiv \mathbf{K} (DIS G) \varphi$

abbreviation $prck::\varrho \Rightarrow \tau \Rightarrow \tau \Rightarrow \tau \ (\langle \mathbf{C}_- \rangle)$

where $\mathbf{C}_G(\varphi|\psi) \equiv \lambda W w. \forall v. (tc (\text{intersection-rel } (EVR G) (\lambda u v. W v \wedge \varphi W v)) w v) \longrightarrow (\psi W v)$

abbreviation $pcmn::\varrho \Rightarrow \tau \Rightarrow \tau \ (\langle \mathbf{C}_- \rangle)$

where $\mathbf{C}_G \varphi \equiv \mathbf{C}_G(\top|\varphi)$

Postulating S5 principles for the agent's accessibility relations.

abbreviation $S5Agent::\alpha \Rightarrow \text{bool}$

where $S5Agent i \equiv \text{reflexive } i \wedge \text{transitive } i \wedge \text{euclidean } i$

abbreviation $S5Agents::\varrho \Rightarrow \text{bool}$

where $S5Agents A \equiv \forall i. (A i \longrightarrow S5Agent i)$

Introducing "Defs" as the set of the above definitions; useful for convenient unfolding.

named-theorems $Defs$

declare $\text{reflexive-def}[Defs] \text{symmetric-def}[Defs] \text{transitive-def}[Defs]$

$\text{euclidean-def}[Defs] \text{intersection-rel-def}[Defs] \text{union-rel-def}[Defs]$

$\text{sub-rel-def}[Defs] \text{inverse-rel-def}[Defs] \text{big-union-rel-def}[Defs]$

$\text{big-intersection-rel-def}[Defs] \text{tc-def}[Defs]$

Consistency: nitpick reports a model.

lemma $\text{True nitpick [satisfy] } \langle \text{proof} \rangle$

2 Automating the Wise Men Puzzle

Agents are modeled as accessibility relations.

consts $a::\alpha \ b::\alpha \ c::\alpha$

abbreviation $Agent::\alpha \Rightarrow \text{bool} \ (\langle \mathcal{A} \rangle)$ **where** $\mathcal{A} x \equiv x = a \vee x = b \vee x = c$

axiomatization where $\text{group-S5: } S5Agents \ \mathcal{A}$

Common knowledge: At least one of a, b and c has a white spot.

consts $ws::\alpha \Rightarrow \sigma$

axiomatization where $WM1: [\mathbf{C}_{\mathcal{A}} (^A ws a \vee ^A ws b \vee ^A ws c)]$

Common knowledge: If x does not have a white spot then y knows this.

axiomatization where

$WM2ab: [C_A (\neg(A_{ws} a) \rightarrow (K_b (\neg(A_{ws} a))))]$ **and**
 $WM2ac: [C_A (\neg(A_{ws} a) \rightarrow (K_c (\neg(A_{ws} a))))]$ **and**
 $WM2ba: [C_A (\neg(A_{ws} b) \rightarrow (K_a (\neg(A_{ws} b))))]$ **and**
 $WM2bc: [C_A (\neg(A_{ws} b) \rightarrow (K_c (\neg(A_{ws} b))))]$ **and**
 $WM2ca: [C_A (\neg(A_{ws} c) \rightarrow (K_a (\neg(A_{ws} c))))]$ **and**
 $WM2cb: [C_A (\neg(A_{ws} c) \rightarrow (K_b (\neg(A_{ws} c))))]$

Positive introspection principles are implied.

lemma $WM2ab'$: $[C_A ((A_{ws} a) \rightarrow K_b (A_{ws} a))]$
 $\langle proof \rangle$

lemma $WM2ac'$: $[C_A ((A_{ws} a) \rightarrow K_c (A_{ws} a))]$
 $\langle proof \rangle$

lemma $WM2ba'$: $[C_A ((A_{ws} b) \rightarrow K_a (A_{ws} b))]$
 $\langle proof \rangle$

lemma $WM2bc'$: $[C_A ((A_{ws} b) \rightarrow K_c (A_{ws} b))]$
 $\langle proof \rangle$

lemma $WM2ca'$: $[C_A ((A_{ws} c) \rightarrow K_a (A_{ws} c))]$
 $\langle proof \rangle$

lemma $WM2cb'$: $[C_A ((A_{ws} c) \rightarrow K_b (A_{ws} c))]$
 $\langle proof \rangle$

Automated solutions of the Wise Men Puzzle.

theorem $whitespot-c$: $[(!\neg K_a(A_{ws} a))(!\neg K_b(A_{ws} b))(K_c(A_{ws} c))]$
 $\langle proof \rangle$

For the following, alternative formulation a proof is found by sledgehammer, while the reconstruction of this proof using trusted methods (often) fails; this hints at further opportunities to improve the reasoning tools in Isabelle/HOL.

theorem $whitespot-c'$:
 $[(!\neg((K_a(A_{ws} a)) \vee (K_a(\neg A_{ws} a)))(!\neg((K_b(A_{ws} b)) \vee (K_b(\neg A_{ws} b)))(K_c(A_{ws} c)))]$
 $\langle proof \rangle$

Consistency: nitpick reports a model.

lemma *True nitpick* $[satisfy]$ $\langle proof \rangle$
end

References

- [1] C. Benzmüller. Universal (meta-)logical reasoning: Recent successes. *Science of Computer Programming*, 172:48–62, 2019.
- [2] C. Benzmüller. Universal (meta-)logical reasoning: The wise men puzzle (Isabelle/HOL Dataset). *Data in Brief*, 24(103823):1–5, 2019.

- [3] C. Benzmüller and S. Reiche. Modeling and automating public announcement logic with relativized common knowledge as a fragment of HOL in LogiKEy. Technical Report arXiv:2111.01654, CoRR, 2021.
- [4] S. Reiche and C. Benzmüller. Public announcement logic in HOL. In M. A. Martins and S. Igor, editors, *Dynamic Logic. New Trends and Applications. DaLi 2020*, volume 12569 of *Lecture Notes in Computer Science*. Springer, Cham, 2020.