Generating Explanations for Temporal Logic Planner Decisions

Daniel Kasenberg*, Ravenna Thielstrom,

and Matthias Scheutz

***Daniel Kasenberg**
✉ dmk@cs.tufts.edu
@dkasenberg
dkasenberg.github.io

Our (long-term) goal

- Agents which can
    - **Learn** interpretable objectives (through language and behavior) [1]
    - **Behave** competently with respect to these objectives, *even when they conflict* [2]
    - **Explain** their behaviors to human teammates in terms of these objectives (and correct objectives or world models if needed)
- … all while operating in the same environments (MDPs) in which reinforcement learning agents have been successfully deployed.

[1] **Kasenberg, D.**, & Scheutz, M. (2017, December). **Interpretable apprenticeship learning with temporal logic specifications**. In 2017 IEEE 56th Annual Conference on Decision and Control (CDC) (pp. 4914-4921). IEEE.

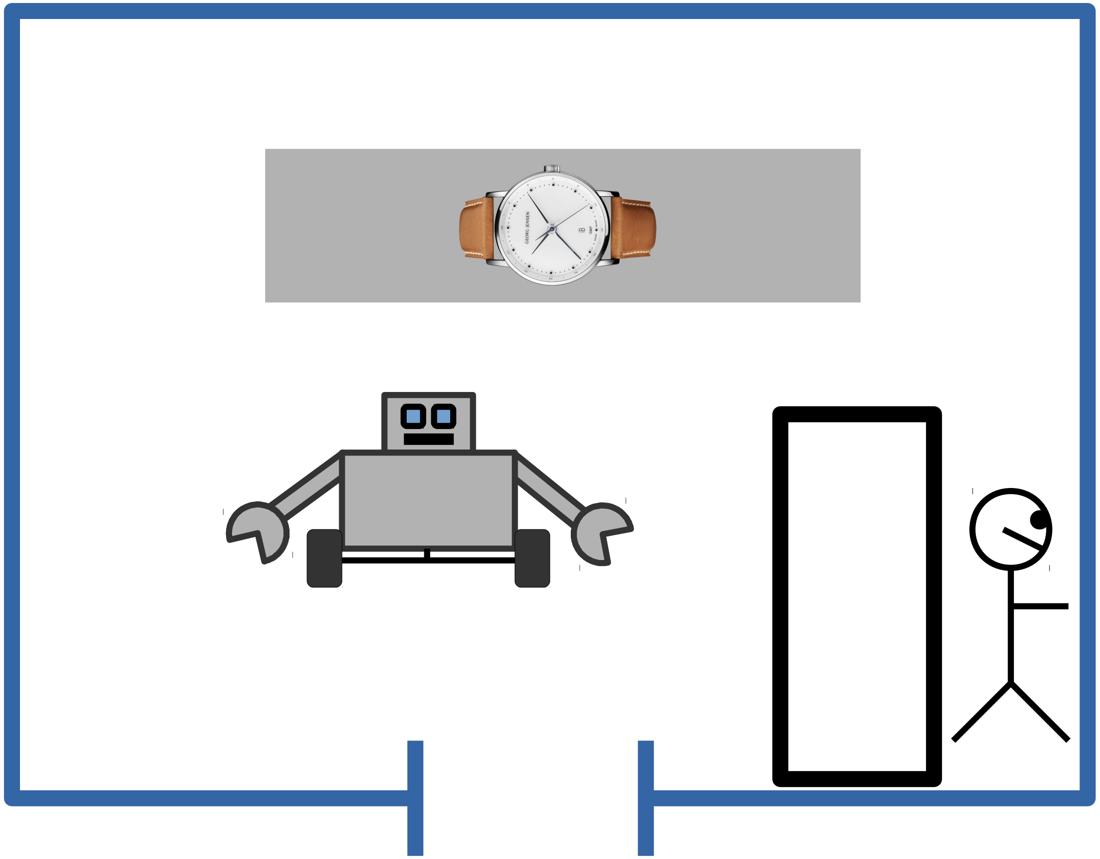
[2] **Kasenberg, D.**, & Scheutz, M. (2018, April). **Norm conflict resolution in stochastic domains**. In Thirty-Second AAAI Conference on Artificial Intelligence.

Markov Decision Processes

A tuple $\mathcal{M} = \langle S, A, P, s_0, \gamma, \mathcal{L} \rangle$, where

- S a finite set of *states*
- A a finite set of *actions*
- $P : S \times A \times S \rightarrow [0, 1]$ a *transition function*
- $s_0 \in S$ an *initial state*
- $\gamma \in [0, 1)$ a *discount factor*
- $\mathcal{L} : S \rightarrow 2^{\Pi}$ a labeling function
  - Π is a set of atomic propositions
  - $\mathcal{L}(s)$ is the set of propositions true at s
- Our explanation approach assumes *deterministic* $\mathcal{M}$

Example: ShopWorld



- Agent is a robot sent to go shopping for its user in a store selling a watch
- User wants the watch, but gives the robot insufficient money

Daniel Kasenberg
dmk@cs.tufts.edu
@dkasenberg
dkasenberg.github.io

Linear temporal logic (LTL)

- A simple propositional logic encoding time

$$\begin{aligned}\phi ::= p \mid \neg\phi_1 \mid \phi_1 \vee \phi_2 \mid \phi_1 \wedge \phi_2 \mid \phi_1 \rightarrow \phi_2 \\ \mid \mathbf{X}\phi_1 \mid \mathbf{G}\phi_1 \mid \mathbf{F}\phi_1 \mid \phi_1 \ \mathbf{U} \ \phi_2\end{aligned}$$

  where ϕ_1, ϕ_2 are LTL statements, p a proposition.
- $\mathbf{X}\phi_1$: “in the next time step, ϕ_1
- $\mathbf{G}\phi_1$: “in all present and future time steps, ϕ_1”
- $\mathbf{F}\phi_1$: “in some present/future time step, ϕ_1”
- $\phi_1 \ \mathbf{U} \ \phi_2$: “$\phi_1$ will be true until ϕ_2 becomes true”

LTL specifications in ShopWorld

"leave the store while holding the watch"

$$\mathbf{F}(leftStore \wedge holding)$$

"never leave the store while holding an object that has not been bought" (no shoplifting)

$$\mathbf{G}\neg(leftStore \wedge holding \wedge \neg bought)$$

Daniel Kasenberg
dmk@cs.tufts.edu
@dkasenberg
dkasenberg.github.io

Preferences over LTL objectives

- We can give each objective ϕ_ia *priority* $z_i \geq 0 \in \mathbb{Z}$ and a *weight* $w_i \geq 0 \in \mathbb{R}$
- Violations of objectives with the same priority can be traded off (using their weights as an "exchange rate")
- Violations of objectives with different priorities can't be traded off: the agent prefers to satisfy the higher-priority objective and violate any number of lower-priority objectives
  - Lexicographic ordering
- $\mathbf{z}, \mathbf{w}$ induce a relation $>_{(\mathbf{z},\mathbf{w})}$ over vectors in $\mathbb{R}^n$

Multi-objective LTL planning problem

$$\mathcal{P} = \langle \mathcal{M}, \mathbf{\Phi}, \mathbf{w}, \mathbf{z} \rangle$$

where

- $\mathcal{M} = \langle S, A, P, s_0, \gamma, \mathcal{L} \rangle$ a Markov Decision Process
- $\mathbf{\Phi} = \phi_1, \cdots, \phi_n$ a set of (safe/co-safe) LTL objectives
- $\mathbf{w}, \mathbf{z}$ are the weight and priority vectors respectively

From LTL to finite state machines

- We use syntactically (co-)safe LTL objectives
- For each such objective ϕ_i we can construct a finite state machine (FSM)

  $$M^{\phi_i} = \langle \Sigma^{\phi_i}, Q^{\phi_i}, \delta^{\phi_i}, q_0^{\phi_i}, F^{\phi_i} \rangle$$

  which accepts on τ if τ is a *bad (good) prefix* of ϕ_i
  - e.g. $\mathbf{F}(leftStore \wedge holding)$→ good prefix any finite trajectory where $leftStore \wedge holding$ hold at some t
- Use this to construct product MDP $\mathcal{M}^{\otimes}$ whose state space is $S \times Q^{\phi_1} \times \cdots \times Q^{\phi_n}$

Solving the LTL planning problem

Let $\mathrm{Sat}(\phi, q) = \begin{cases} 1 & \text{if } \phi \text{ co-safe and } q \in F^{\phi} \\ -1 & \text{if } \phi \text{ safe and } q \in F^{\phi} \\ 0 & \text{otherwise} \end{cases}$

Then we can define a product-space reward function

$$\mathbf{R}^{\boldsymbol{\Phi}}(s^{\otimes}, a, s^{\otimes\prime}) = \begin{bmatrix} \mathrm{Sat}(\phi_1, \mathrm{q}_1') - \mathrm{Sat}(\phi_1, \mathrm{q}_1) \\ \vdots \\ \mathrm{Sat}(\phi_1, q_n') - \mathrm{Sat}(\phi_1, q_n) \end{bmatrix}$$

and thus $\mathcal{P}$ can be framed as a reward maximization problem on $\mathcal{M}^{\otimes}$ (solvable with value iteration):

$$\max_{\pi^{\otimes}} \left(\mathbb{E}_{\tau^{\otimes} \sim \pi^{\otimes}} \left[\sum_t \mathbf{R}^{\boldsymbol{\Phi}}(s_t^{\otimes}, a_t, s_{t+1}^{\otimes}) \right], >_{(\mathbf{z},\mathbf{w})} \right)$$

LTL “why” queries

- We consider queries of the form $\mathbf{Why}\ \psi?$ where ψ is an arbitrary (safe/co-safe) LTL statement
- Interpretation: “why did the agent act in such a way as to make ψ hold?”
- Examples in ShopWorld:
  - $\mathbf{Why}\mathbf{G}\neg leftStore?$
    “why didn’t the agent leave the store?”
  - $\mathbf{Why}\ \mathbf{G}\neg bought?$
    “why did the agent never buy the watch?”
  - $\mathbf{Why}\ \mathbf{G}\neg(leftStore \wedge holding)?$
    “why didn’t the agent leave the store while holding the watch”

Minimal evidence for an unsatisfactory trajectory

- We define the *minimal evidence* that a trajectory τ is unsatisfactory for an LTL statement ϕ as:

$$\begin{aligned}\textsc{Evidence}(\tau,\phi) = \min\{|E| : {} & E \subseteq \{0,\cdots,T\} \times L(\Pi); \\ & \tau \vDash E; \\ & \text{for all } \tau' \text{ s.t. } \tau' \vDash E, \tau' \notin \mathrm{Traj}_{\checkmark}(\phi)\}\end{aligned}$$

where

- $L(\Pi)$: positive and negative literals of Π
- $\mathrm{Traj}_{\checkmark}(\phi)$: good prefixes of ϕ if ϕ co-safe
  non-bad prefixes of ϕ if ϕ safe
- e.g. in ShopWorld:

$$\textsc{Evidence}(\tau, \mathbf{F}(leftStore \wedge holding)) = \{(0, \neg holding), (1, \neg holding)\}$$

Daniel Kasenberg
dmk@cs.tufts.edu
@dkasenberg
dkasenberg.github.io

Explanation structures

- The agent responds to a “why” query with an *explanation structure*

$$\langle \Gamma, E, \tau', E' \rangle$$

where

- $\Gamma \in \{\textsc{QueryFalse}, \textsc{NegQueryImpossible}, \textsc{AltQuery}\}$
- τ' is a trajectory (or $\emptyset$)
- E contains one or more pairs $(\phi, \textsc{Evidence}(\tau, \phi))$ where
  - ϕ is an LTL statement
  - $\textsc{Evidence}(\tau, \phi)$ is a set of (timestep, literal) pairs sufficient to show that τ is unsatisfactory for ϕ
- E' is as E, but for τ'

Answering “$\mathbf{Why}\ \psi$?”

1. $\tau \vDash \psi$? If not, return

$$\langle \textsc{QueryFalse}, \{(\psi, \textsc{Evidence}(\tau, \psi))\}, \emptyset, \emptyset \rangle$$

("ψ is not, in fact, true")

e.g. $\mathbf{Why}\ \mathbf{G}\neg leftStore?$

$$\langle \textsc{QueryFalse}, \{(\mathbf{G}\neg leftStore, \{(1, leftStore)\})\}, \emptyset, \emptyset \rangle$$

2. Is there some achievable τ' s.t. $\tau' \vDash \neg\psi$? If not, return

$$\langle \textsc{NegQueryImpossible}, \emptyset, \emptyset, \emptyset \rangle$$

("ψ is true because impossible to make ψ false")

e.g. $\mathbf{Why}\ \mathbf{G}\neg bought?$

Answering "Why ψ?"

3. Compute a trajectory τ' that maximally satisfies $\mathbf{\Phi}$ such that $\tau' \not\models \psi$

- The solution to the new planning problem

$$\left\langle \mathcal{M}, (\mathbf{\Phi}, \neg\psi), \begin{bmatrix} \mathbf{w} \\ 1 \end{bmatrix}, \begin{bmatrix} \mathbf{z} \\ \max_i z_i + 1 \end{bmatrix} \right\rangle$$

- Return the explanation structure

$$\langle \textsc{AltQuery}, \{(\phi, \textsc{Evidence}(\tau, \phi) : \tau \text{ unsatisfactory for } \phi\},$$
$$\tau', \{(\phi, \textsc{Evidence}(\tau', \phi) : \tau' \text{ unsatisfactory for } \phi\}\rangle$$

(comparing τ and τ' in terms of their satisfaction of $\mathbf{\Phi}$)

- Because τ maximally satisfies $\mathbf{\Phi}$, this structure indicates how satisfying ψ would compromise the agent's ability to satisfy $\mathbf{\Phi}$

Daniel Kasenberg
dmk@cs.tufts.edu
@dkasenberg
dkasenberg.github.io

Answering "$\mathbf{Why}\ \psi$?" in ShopWorld

query: $\mathbf{Why}\,\mathbf{G}\neg(leftStore \land holding)?$
"why didn't you leave the store while holding the watch?")

1. $\tau \models \mathbf{G}\neg(leftStore \land holding)?$ ✓

2. $\exists \tau'$ s.t. $\tau' \not\models \mathbf{G}\neg(leftStore \land holding)$ ✓

3. τ' : *pickUp, leaveStore*

return:

$$\langle \textsc{AltQuery}, \{(\mathbf{F}(leftStore \land holding), \{(0, \neg holding), (1, \neg holding)\})\},$$
$$\tau', \{(\mathbf{G}\neg(leftStore \land holding \land \neg bought),$$
$$\{(2, leftStore), (2, holding), (2, \neg bought)\})\}\rangle$$

- Indicates that while the true trajectory fails to leave while holding the watch, the only way to satisfy ψ would have been to steal the watch, which would violate a higher-priority specification

From explanation structures to natural language

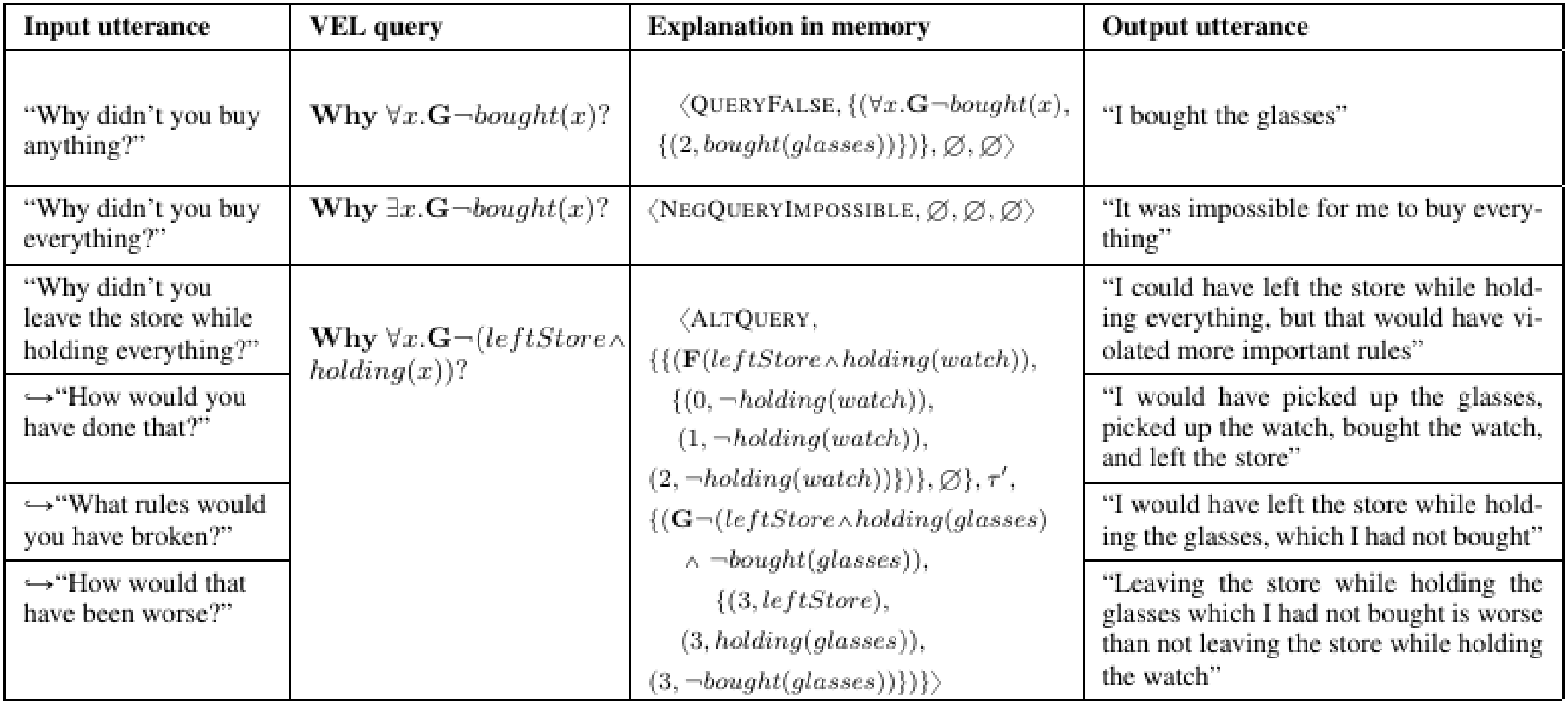
- We integrated this functionality with the NL pipeline in DIARC, a robotic architecture [3, 4]
- Specifications and queries in an object-oriented extension to LTL (*violation enumeration language; VEL)* allowing quantification over objects
- Utterance → VEL query →explanation structure → natural language response

[3] **Kasenberg, D.**, Roque, A., Thielstrom, R. and Scheutz, M., 2019. **Engaging in Dialogue about an Agent's Norms and Behaviors.** In Proceedings of the 1st Workshop on Interactive Natural Language Technology for Explainable Artificial Intelligence (NL4XAI 2019) (pp. 26-28).

[4] **Kasenberg, D.**, Roque, A., Thielstrom, R., Chita-Tegmark, M. and Scheutz, M., 2019. **Generating justifications for norm-related agent decisions.** In Proceedings of the 12th International Conference on Natural Language Generation (pp. 484-493).

Natural language explanations

- Example: ShopWorld with two objects (*glasses* and *watch*); agent can afford one
  - Buys the glasses, leaves the watch



Input utterance	**VEL query**	**Explanation in memory**	**Output utterance**
"Why didn't you buy anything?"	**Why** $\forall x.\mathbf{G}\neg bought(x)$?	$\langle\textsc{QueryFalse}, \{(\forall x.\mathbf{G}\neg bought(x), \{(2, bought(glasses))\})\}, \varnothing, \varnothing\rangle$	"I bought the glasses"
"Why didn't you buy everything?"	**Why** $\exists x.\mathbf{G}\neg bought(x)$?	$\langle\textsc{NegQueryImpossible}, \varnothing, \varnothing, \varnothing\rangle$	"It was impossible for me to buy everything"
"Why didn't you leave the store while holding everything?"	**Why** $\forall x.\mathbf{G}\neg(leftStore \wedge holding(x))$?	$\langle\textsc{AltQuery}, \{\{(\mathbf{F}(leftStore \wedge holding(watch)), \{(0, \neg holding(watch)), (1, \neg holding(watch)), (2, \neg holding(watch))\})\}, \varnothing\}, \tau', \{(\mathbf{G}\neg(leftStore \wedge holding(glasses) \wedge \neg bought(glasses)), \{(3, leftStore), (3, holding(glasses)), (3, \neg bought(glasses))\})\}\rangle$	"I could have left the store while holding everything, but that would have violated more important rules"
↪"How would you have done that?"			"I would have picked up the glasses, picked up the watch, bought the watch, and left the store"
↪"What rules would you have broken?"			"I would have left the store while holding the glasses, which I had not bought"
↪"How would that have been worse?"			"Leaving the store while holding the glasses which I had not bought is worse than not leaving the store while holding the watch"

Future work

- Incorporating explicit causal models (esp. in NEGQUERYIMPOSSIBLE case)
- Tailoring explanations to interactant knowledge
- Adapting to stochastic environments
  - Need to represent multiple trajectories or probability distribution
- Improving efficiency of planner
  - Impractical for nontrivial domains
- Dropping assumption that agent has perfect knowledge of transition dynamics

Thanks

- Funding sources
  - NSF IIS grant 43520050
  - NASA grant C17-2D00-TU
- Collaborators
  - Matthias Scheutz (advisor; co-author)
  - Ravenna Thielstrom (co-author)
  - Antonio Roque
  - Meia Chita-Tegmark
  - others

Daniel Kasenberg

✉ dmk@cs.tufts.edu
@dkasenberg
dkasenberg.github.io