# Supplementary Materials for "Monotonic Safety for Scalable Probabilistic Model Checking"

**ACM Reference Format:**

. 2021. Supplementary Materials for "Monotonic Safety for Scalable Probabilistic Model Checking". *ACM Trans. Embedd. Comput. Syst.* 0, 0, Article 0 ( 2021), 8 pages. https://doi.org/00.0000/0000000.0000000

## A APPENDIX

This document serves as the supplemental material to the EMSOFT 2021 submission titled: Monotonic Safety for Scalable Probabilistic Model Checking. It contains the proofs of propositions, theorems, and counterexamples from the main paper, as well as the additional details on the experimental setup.

### A.1 Proof of Conservatism for MoS Abstraction Techniques

Theorem A.1 (Conservative transition trimming). *Given a PA* $\mathsf{M}$ *with some non-determinism, safety property* $\psi$*, and an MoS assumption* $\succeq_{MoS}$ *such that* $\succeq_{MoS}$*,* $s_1 \succeq_{MoS} s_2 \implies Pr_{\mathsf{M}_1}(\psi) \geq Pr_{\mathsf{M}_2}(\psi)$*, for any model* $\mathsf{M}'$ *that is constructed by either destination-based transition trimming or source-based transition trimming in* $\mathsf{M}$*, it is true that* $Pr_{\mathsf{M}'}(\psi) \leq Pr_{\mathsf{M}}(\psi)$*.*

Proof. Let conservative model $\mathsf{M}_1$ be given. So $Pr_{\mathsf{M}_1}(\psi) \leq Pr_{\mathsf{M}}(\psi)$. In addition, we write that $\mathsf{M}_1 = (S, \bar{s}, \alpha_{\mathsf{M}_1}, \delta_{\mathsf{M}_1}, L)$ be given, where $S$ is the set of states, $\bar{s}$ is the initial state, $\alpha_{\mathsf{M}_1}$ is the alphabet, $\delta_{\mathsf{M}_1} \subseteq S \times (\alpha_{\mathsf{M}_1} \cup \{\tau\}) \times Dist(S)$ is a probabilistic transition relation and $L$ is a labelling function assigning atomic propositions from a set $AP$ to each state. Now we need to show that applying destination based transition trimming or source based transition trimming results in a new model $\mathsf{M}'$ which is conservative wrt $\mathsf{M}_1$.

**Proof that destination-based transition trimming is conservative**: We prove that destination based transition trimming of non-probabilistic transitions is conservative. The proof can easily be extended to probabilistic transitions by using the MoS partial order over states to establish a partial over of distributions of states.

To apply transition trimming to construct our abstraction $\mathsf{M}'$ of $\mathsf{M}_1$, we say that if state $s$ of model $\mathsf{M}_1$ has transitions to both states $s_1$ and $s_2$ with letters $\alpha_1 \in \alpha_{\mathsf{M}_1}$ and $\alpha_2 \in \alpha_{\mathsf{M}_1}$, respectively, and $s_1 \succeq_{MoS} s_2$, then we remove the transition from $s$ to $s_1$ from $\mathsf{M}_1$, which results in model $\mathsf{M}'$. Since $\mathsf{M}'$ is simply a copy of $\mathsf{M}_1$, but with a single transition removed, it follows that $\delta_{\mathsf{M}'} \subset \delta_{\mathsf{M}_1}$. We are assuming that the transitions in question map to single next states since we are only abstracting a non-probabilistic sub-component of $\mathsf{M}_1$.

We need to show that $Pr_{\mathsf{M}'}(\psi) \leq Pr_{\mathsf{M}_1}(\psi)$. First, note that:

$$Pr_{\mathsf{M}_1}(\psi) = inf_{\sigma \in Adv_{\mathsf{M}_1}} Pr^{\sigma}_{\mathsf{M}_1}(\psi) \text{ and } Pr_{\mathsf{M}'}(\psi) = inf_{\sigma \in Adv_{\mathsf{M}'}} Pr^{\sigma}_{\mathsf{M}'}(\psi)$$

Author's address:

1539-9087/2021/0-ART0 $15.00

https://doi.org/00.0000/0000000.0000000

So we just need to show that $Pr^{\sigma'}_{\mathsf{M}_1}(\psi) \geq Pr_{\mathsf{M}'}(\psi)\ \forall \sigma' \in Adv_{\mathsf{M}_1}/Adv_{\mathsf{M}'}$. Consider such a $\sigma'$. Due to the construction of $\mathsf{M}_1$ and $\mathsf{M}'$, the path which $\sigma'$ takes through $\mathsf{M}_1$ takes the transition from $s$ to $s_1$. But there is a corresponding $\sigma \in Adv_{\mathsf{M}'}$ which takes the same initial path, but then takes the transition from $s$ to $s_2$. Recall that our MoS assumption states that $Pr_{\mathsf{M}_1^{s_1}}(\psi) \geq Pr_{\mathsf{M}_1^{s_2}}(\psi)$, where $\mathsf{M}^s$ indicates the model where PA $\mathsf{M}$ has single initial state $s$. So that means $Pr^{\sigma'}_{M_1}(\psi) \geq Pr_{\mathsf{M}'}(\psi)$, completing the proof.

**Proof that source-based trimming is conservative**: Consider some abstract state $s \in S$, which corresponds to intervals of concrete states $\mathcal{I}_s = \{[a_{1sl}, a_{1su}], \ldots, [a_{nsl}, a_{nsu}]\}$. We need this correspondence of abstract states to concrete states to be preserved across abstractions because our MoS rule is defined over concrete system states. So without this correspondence souce based transition trimming is not possible. Now consider two concrete system states, $t_1$ and $t_2$, both contained in the abstract state $s$. Assume that $t_1 \geq_{MoS} t_2$. Now consider the transitions originating from abstract state $s$, specifically the transitions which are present from true states $t_1$ and $t_2$. To construct our abstraction $\mathsf{M}'$ of $\mathsf{M}$, we remove any transitions, both probabilistic and non-probabilistic, from state $s$ which correspond exclusively to true state $t_1$.

Let $P_{\mathsf{M}_1^{t_1}}(\psi)$ and $P_{\mathsf{M}_1^{t_2}}(\psi)$ denote the probability of model $\mathsf{M}_1$ satisfying property $G$ given that it's in state $t_1$ or $t_2$, respectively. By our MoS assumption, we have that $P_{\mathsf{M}_1^{t_1}}(\psi) \geq P_{\mathsf{M}_1^{t_2}}(\psi)$. In addition, our abstract state $s$ satisfies

$$P_{\mathsf{M}_1^{s}}(\psi) \leq P_{\mathsf{M}_1^{\bar{t}}}(\psi) \ \ \forall \bar{t} \in \mathcal{I}_s$$

By combining these two properties, it follows that the transitions present in abstract state $s$ due to true state $t_1$ do not affect the overall probability of safety in abstract state $s$, and thus do not affect the overall probability of safety model, $P_{\mathsf{M}_1}(\psi)$. So we can remove them from the model without increasing the probability of safety, giving us that $P_{\mathsf{M}_1}(\psi) \geq P_{\mathsf{M}'}(\psi)$.

□

## A.2 Counterexamples to Monotonic Safety in AEBS

In all counterexamples, we set $\tau = 1s$ and window size $N_F = 1$. We adopt the notation $Pr(\psi_{nocol} \mid d, s)$ as a shorthand for $Pr_{\mathsf{M}}(\psi_{nocol})$ where $\mathsf{M}$ has the initial state $(d, s)$.

Counterexample 1. *Assumption $\succeq_d$ does not hold in the AEBS with one BP and distance-dependent perception for shift $\Delta = (1m, 0)$ when $B = 10m/s^2, v_0 = 11m/s, d_0 = 13m$, and $Pr_{det}(d) = 1 - \lceil d \rceil/20$.*

Proof. We set $\bar{s}_b$ = (13, 11) as the initial state of $\mathsf{M}_b$, and $\bar{s}_a$ = (14, 11) as the initial state of $\mathsf{M}_a$. Then we calculate as follows.

$$\begin{aligned}
&Pr_{\mathsf{M}_a}(\psi_{nocol}) = \\
&Pr_{det}(14) * Pr(\psi_{nocol}|3,1) + (1 - Pr_{det}(14)) * \underbrace{Pr(\psi_{nocol}|3,11)}_{=0} = \\
&Pr_{det}(14) * (Pr_{det}(3) * \underbrace{Pr(\psi_{nocol}|2,0)}_{=1} + (1 - Pr_{det}(3)) * \underbrace{Pr(\psi_{nocol}|2,1)}_{=Pr_{det}(2)}) = \\
&Pr_{det}(14) * (Pr_{det}(3) + (1 - Pr_{det}(3)) * Pr_{det}(2)) = \\
&Pr_{det}(14) * (Pr_{det}(3) + Pr_{det}(2) - Pr_{det}(3) * Pr_{det}(2)).
\end{aligned}$$

$$\begin{aligned}
&Pr_{\mathsf{M}_b}(\psi_{nocol}) = \\
&Pr_{det}(13) * Pr(safe|2,1) + (1 - Pr_{det}(13)) * \underbrace{Pr(safe|2,11)}_{=0} = \\
&Pr_{det}(13) * (Pr_{det}(2) * \underbrace{Pr(safe|1,0)}_{=1} + (1 - Pr_{det}(2)) * \underbrace{Pr(safe|1,1)}_{=0}) = \\
&Pr_{det}(13) * Pr_{det}(2).
\end{aligned}$$

If we set $Pr_{det}(d) = 1 - \lceil d \rceil/20$ (i.e., the detection probability grows linearly from 0% at 20 meters to 100% at 0 meters), then the above falsifies MoS:

$$Pr(\psi_{nocol}|13,11) = 0.315 > 0.2955 = Pr(\psi_{nocol}|14,11)$$

P.S. This inequality is reversed for less steep detection curves like $1 - \lceil d \rceil/40$.

□

The second counterexample shows that driving at a faster speed may be beneficial because it leads to closer distances, which result in higher detection chances.

Counterexample 2. *Assumption $\succeq_v$ does not hold in the AEBS with one BP and distance-dependent perception for shift $\Delta = (0, -1m/s)$ when $B = 10m/s^2, v_0 = 9m/s, d_0 = 20m$, and $Pr_{det}(d) = 1 - \lceil d \rceil/20$.*

Proof. We set $\bar{s}_b$ = (20, 9) as the initial state of $\mathsf{M}_b$, and $\bar{s}_a$ = (20, 8) as the initial state of $\mathsf{M}_a$. Then we calculate as follows.

$$\begin{aligned}
&Pr_{\mathsf{M}_a}(\psi_{nocol}) = \\
&Pr_{det}(20) * \underbrace{Pr(\psi_{nocol}|11,0)}_{=1} + (1 - Pr_{det}(20)) * Pr(\psi_{nocol}|11,9) = \\
&Pr_{det}(20) + (1 - Pr_{det}(20)) * \\
&\quad (Pr_{det}(11) * \underbrace{Pr(\psi_{nocol}|2,0)}_{=1} + (1 - Pr_{det}(11)) * \underbrace{Pr(\psi_{nocol}|2,9)}_{=0}) = \\
&Pr_{det}(20) + Pr_{det}(11) * (1 - Pr_{det}(20))
\end{aligned}$$

$$\begin{aligned}
&Pr_{\mathsf{M}_b}(\psi_{nocol}) = \\
&Pr_{det}(20) * \underbrace{Pr(\psi_{nocol}|12,0)}_{=1} + (1 - Pr_{det}(20)) * Pr(\psi_{nocol}|12,8) = \\
&Pr_{det}(20) + (1 - Pr_{det}(20)) * \\
&\quad (Pr_{det}(12) * \underbrace{Pr(\psi_{nocol}|4,0)}_{=1} + (1 - Pr_{det}(12)) * \underbrace{Pr(\psi_{nocol}|4,8)}_{=0}) = \\
&Pr_{det}(20) + Pr_{det}(12) * (1 - Pr_{det}(20))
\end{aligned}$$

For any detection curve where $Pr_{det}(11) > Pr_{det}(12)$ (e.g., $Pr_{det}(d) = 1 - d/20$), the following inequality holds, falsifying MoS:

$$Pr_{\mathsf{M}_b}(\psi_{nocol}) > Pr_{\mathsf{M}_a}(\psi_{nocol})$$

□

The third counterexample demonstrates that starting closer to the obstacle may be beneficial because it may lead to a stronger braking mode, which stops the car faster.

Counterexample 3. *Assumption $\succeq_d$ does not hold in the AEBS with multiple BPs and distance-independent perception for shift $\Delta = (2m, 0)$ when $B(d) = [10m/s^2 \; if \, d \leq 11m; 5m/s^2 \; otherwise]$, $v_0 = 9m/s, d_0 = 10m$, and $Pr_{det} = 0.5$.*

Proof. We set $\bar{s}_b$ = (10, 9) as the initial state of $\mathsf{M}_b$, and $\bar{s}_a$ = (12, 9) as the initial state of $\mathsf{M}_a$. Then we calculate as follows.

$$Pr_{\mathsf{M}_a}(\psi_{nocol}) = \\ Pr_{det} * \underbrace{Pr(\psi_{nocol}|3,4)}_{=0} + (1 - Pr_{det}) * \underbrace{Pr(\psi_{nocol}|3,8)}_{=0} = 0$$

$$Pr_{\mathsf{M}_b}(\psi_{nocol}) = \\ Pr_{det} * \underbrace{Pr(\psi_{nocol}|1,0)}_{=1} + (1 - Pr_{det}) * \underbrace{Pr(\psi_{nocol}|1,9)}_{=0} = Pr_{det}$$

Thus, for any $Pr_{det} > 0$, the above falsifies MoS:

$$Pr_{\mathsf{M}_b}(\psi_{nocol}) > Pr_{\mathsf{M}_a}(\psi_{nocol})$$

□

The fourth counterexample shows that going at a faster speed may be beneficial because it leads to closer distances, which result in stronger braking.

Counterexample 4. *Assumption $\succeq_v$ does not hold in the AEBS with multiple BPs and distance-independent perception for shift $\Delta = (0, -1m/s)$ when $B(d) = [10m/s^2 \; if \, d \leq 11m; 3m/s^2 \; otherwise]$, $v_0 = 9m/s, d_0 = 20m$, and $Pr_{det} = 0.5$.*

Proof. We set $\bar{s}_b$ = (20, 9) as the initial state of $\mathsf{M}_b$, and $\bar{s}_a$ = (20, 8) as the initial state of $\mathsf{M}_a$. Then we calculate as follows.

We consider the safety chance in two scenarios: $s_0 = 9m/s$ and $s_0 = 8m/s$.

$$Pr_{\mathsf{M}_a}(\psi_{nocol}) = \\ Pr_{det} * Pr(\psi_{nocol}|11,6) + (1 - Pr_{det}) * Pr(\psi_{nocol}|11,9) = \\ Pr_{det} * (Pr_{det} * \underbrace{Pr(\psi_{nocol}|5,0)}_{=1} + (1 - Pr_{det}) * \underbrace{Pr(\psi_{nocol}|5,6)}_{=0}) + \\ (1 - Pr_{det}) * (Pr_{det} * \underbrace{Pr(\psi_{nocol}|2,0)}_{=1} + (1 - Pr_{det}) * \underbrace{Pr(\psi_{nocol}|2,9)}_{=0} = \\ (Pr_{det})^2 + (1 - Pr_{det}) * Pr_{det} = Pr_{det}$$

$$
\begin{aligned}
&Pr_{\mathsf{M}_b}(\psi_{nocol}) = \\
&Pr_{det} * Pr(\psi_{nocol}|12, 5) + (1 - Pr_{det}) * Pr(\psi_{nocol}|12, 8) = \\
&Pr_{det} * (Pr_{det} * Pr(\psi_{nocol}|7, 2) + (1 - Pr_{det}) * \underbrace{Pr(\psi_{nocol}|7, 5)}_{=Pr_{det}})+ \\
&\quad (1 - Pr_{det}) * (Pr_{det} * \underbrace{Pr(\psi_{nocol}|4, 5)}_{=0} + (1 - Pr_{det}) * \underbrace{Pr(\psi_{nocol}|4, 8)}_{=0} = \cdots = \\
&Pr^2_{det}(1 + 2Pr_{det} - 3Pr^2_{det} + Pr^3_{det})
\end{aligned}
$$

It can be shown that for any $0 < Pr_{det} < 1$, $Pr_{det} > Pr^2_{det}(1 + 2Pr_{det} - 3Pr^2_{det} + Pr^3_{det})$. For example, for $Pr_{det} = 0.5$, $0.5 > 0.34375$. Thus, the above falsifies MoS:

$$Pr_{\mathsf{M}_b}(\psi_{nocol}) > Pr_{\mathsf{M}_a}(\psi_{nocol})$$

□

In summary, these counterexamples arise when due to shifting the model crosses a state boundary into a different braking mode or detection chance.

### A.3 Proposition Proofs for Monotonic Safety in the Simple Model

Proposition A.2. *For AEBS with distance-independent perception and shifts* $(\Delta_d, 0)$ *and* $(0, \Delta_v)$, S1 = 0.

Proof. S1 is a probability difference between two sets of paths: $\dot{\Pi}^{ss}_a \mathbin{\cap\mkern-11mu\circ} \dot{\Pi}^{ss}_b$ and $\dot{\Pi}^{ss}_b \mathbin{\cap\mkern-11mu\circ} \dot{\Pi}^{ss}_a$. Based on the definition of $\mathbin{\cap\mkern-11mu\circ}$, it is easy to see that the sets $\dot{\Pi}^{ss}_a \mathbin{\cap\mkern-11mu\circ} \dot{\Pi}^{ss}_b$ and $\dot{\Pi}^{ss}_b \mathbin{\cap\mkern-11mu\circ} \dot{\Pi}^{ss}_a$ are isomorfic: for every $\pi_a$ in the former, there exists a unique $\pi_b$ in the latter such that $o(\pi_a) = o(\pi_b)$; the inverse is also true.

Paths $\pi_a$ and $\pi_b$ have the same number of steps and the same low- and high-level detection outcomes, so these paths differ only in distances and speeds. The probability of a path is invariant of speed and, based on this proposition's premise, invariant of the distance. Therefore, $Pr(\pi_a) = Pr(\pi_b)$.

Summing up that equation over all pairs of traces in the two sets, it follows that $Pr(\dot{\Pi}^{ss}_a \mathbin{\cap\mkern-11mu\circ} \dot{\Pi}^{ss}_b) = Pr(\dot{\Pi}^{ss}_b \mathbin{\cap\mkern-11mu\circ} \dot{\Pi}^{ss}_a)$. Therefore, S1 = 0.

□

Proposition A.3. *For AEBS with one BP and shift* $(\Delta_d, 0)$, S2 = 0.

Proof. S2 is determined by the probabilities of two sets: $\dot{\Pi}^{ss}_a \mathbin{\setminus\mkern-4mu\circ}\, \dot{\Pi}^{ss}_b$ and $\dot{\Pi}^{ss}_b \mathbin{\setminus\mkern-4mu\circ}\, \dot{\Pi}^{ss}_a$. Assume for contradiction that they are not empty.

Consider some path $\pi_a$ from the former set. By the definition of $\dot{\Pi}^{ss}_a$, $\pi_a$ achieves a stop in $\mathsf{M}_a$ and some shorter or equally long related path $\pi_b$ achieves a stop in $\mathsf{M}_a$. By the definition of $\mathbin{\setminus\mkern-4mu\circ}$, $\pi_b$ has to be strictly shorter than $\pi_a$ — otherwise $\pi_a$ would not be in the former set.

So we arrive at two paths $\pi_a$ and $\pi_b$: they have the same detections up to $l(\pi_b)$ and both lead to a stop. Since there is only one braking power $B$ and the same initial speed $v_0$, the braking commands are the same in both models, and in both models the car stops after $\lceil v_0/(B \cdot \tau) \rceil$ braking commands. Therefore, in both paths $v = 0$ after $l(\pi_b)$ steps.

Path $\pi_a$ continued to execute after reaching a terminal state with $v = 0$, which contradicts our model. Therefore, our assumption was wrong, and $\dot{\Pi}^{ss}_a \mathbin{\setminus\mkern-4mu\circ}\, \dot{\Pi}^{ss}_b = \emptyset$. Similar reasoning leads to $\dot{\Pi}^{ss}_b \mathbin{\setminus\mkern-4mu\circ}\, \dot{\Pi}^{ss}_a = \emptyset$. We conclude S2 = 0. □

Proposition A.4. *For AEBS with one BP and shifts* $(\Delta_d, 0)$ *and* $(0, \Delta_v)$, S3 ≥ 0.

Proof. To show that S3 $\geq 0$, it is sufficient to prove that $\dot{\Pi}_b^{su} = \emptyset$. Assume for contradiction that $\dot{\Pi}_b^{su} \neq \emptyset$ and consider some safe path $\pi_b \in \dot{\Pi}_b^{su}$. This path has a corresponding unsafe path $\pi_a$ that shares the outcome sequence $o(\pi_b)$, which, due to having just one BP, uniquely determines some sequence of braking actions.

For shift $(\Delta_d, 0)$, $\pi_a$ and $\pi_b$ have the same initial speed, which that sequence of braking actions reduces to 0 because $\pi_b$ is safe. Therefore, $\pi_a$ is safe too, contradicting our assumption $\pi_b \in \dot{\Pi}_b^{su}$.

For shift $(0, \Delta_v)$, $\pi_a$ has a smaller initial speed than $\pi_b$, and thus should be stopped in no more braking actions than $\pi_b$. So similarly to the above, $\pi_a$ is safe, contradicting $\pi_b \in \dot{\Pi}_b^{su}$.

Thus for both shifts, we conclude that $\dot{\Pi}_b^{su} = \emptyset$, and so S3 $\geq 0$. □

Proposition A.5. *For AEBS with one BP, distance-independent perception, and shift* $(0, \Delta_v)$*,* S2 + S3 $\geq 0$.

Proof. To show S2 + S3 $\geq 0$, it is sufficient to prove that $Pr(\dot{\Pi}_a^{us}) \geq Pr(\dot{\Pi}_b^{ss} \mathbin{\setminus\!\circ} \dot{\Pi}_a^{ss})$ because under our premises $\dot{\Pi}_b^{su} = \emptyset$ as shown in Proposition A.4.

Based on the definitions of these subsets, paths in $\dot{\Pi}_a^{us}$ and $\dot{\Pi}_a^{ss}$ have $\lceil (v_0 - \Delta_v)/(B \cdot \tau) \rceil$ braking actions, and paths in $\dot{\Pi}_b^{ss}$ have $\lceil v_0/(B \cdot \tau) \rceil$ braking actions.

First, consider the simple case when $\lceil (v_0 - \Delta_v)/(B \cdot \tau) \rceil = \lceil v_0/(B \cdot \tau) \rceil$. Then the outcome trace function $o$ establishes an isomorphism between $\dot{\Pi}_b^{ss}$ and $\dot{\Pi}_a^{ss}$ because each safe sequence of braking actions corresponds to a related pair of unique paths, one in each set. Then $\dot{\Pi}_b^{ss} \mathbin{\setminus\!\circ} \dot{\Pi}_a^{ss} = \emptyset$ and hence S2 + S3 $\geq 0$.

Second, consider the more complex case when $\lceil (v_0 - \Delta_v)/(B \cdot \tau) \rceil < \lceil v_0/(B \cdot \tau) \rceil$. To start, we show that every outcome trace in $o(\dot{\Pi}_b^{ss} \mathbin{\setminus\!\circ} \dot{\Pi}_a^{ss})$ has a strict prefix in $o(\dot{\Pi}_a^{us})$. This is true because the elements of both sets rely, by definition, on the existence of a safe path $\pi_a \in \dot{\Pi}_a$ such that its braking actions do not stop the vehicle in $\mathsf{M}_b$. Thus, the safe extensions of the braking actions of $\pi_a$ belong to $\dot{\Pi}_b^{ss} \mathbin{\setminus\!\circ} \dot{\Pi}_a^{ss}$ and unsafe extensions belong to $o(\dot{\Pi}_a^{us})$. Therefore, $o(\pi_a)$ is a prefix of its safe extensions in $\dot{\Pi}_b^{ss} \mathbin{\setminus\!\circ} \dot{\Pi}_a^{ss}$. Notice that the safe extensions need to have at least one more detection, otherwise they cannot safely stop the vehicle.

Since we assumed that detections are distance-independent, comparing the original set probabilities is equivalent to comparing the probabilities of the detection outcome sets: $Pr(o(\dot{\Pi}_a^{us}))$ and $Pr(o(\dot{\Pi}_b^{ss} \mathbin{\setminus\!\circ} \dot{\Pi}_a^{ss}))$. The second set contains strict extensions of the first set with $n$ more detections, where $n = \lceil v_0/(B \cdot \tau) \rceil - \lceil (v_0 - \Delta_v)/(B \cdot \tau) \rceil > 0$.

Let us take an arbitrary prefix $x \in o(\dot{\Pi}_a^{us})$ and consider the worst case: $o(\dot{\Pi}_b^{ss} \mathbin{\setminus\!\circ} \dot{\Pi}_a^{ss})$ contains all possible extensions of $x$ with a least one detection and an arbitrary number of misdetections. The probability of all such extensions is

$$Pr(x) \cdot (Pr_{det})^n + Pr(x) \cdot (Pr_{det})^n \cdot (1 - Pr_{det}) + Pr(x) \cdot (Pr_{det})^n \cdot (1 - Pr_{det})^2 + \ldots$$

This expression is a sum of a geometric series equal to $Pr(x) \cdot (Pr_{det})^n/(1 - (1 - Pr_{det})) = Pr(x) \cdot (Pr_{det})^{n-1} \leq Pr(x)$. Thus, summing over all such prefixes $x$, we have proven the following:

$$Pr(o(\dot{\Pi}_a^{us})) \geq Pr(o(\dot{\Pi}_b^{ss} \mathbin{\setminus\!\circ} \dot{\Pi}_a^{ss}))$$
$$Pr(\dot{\Pi}_a^{us}) \geq Pr(\dot{\Pi}_b^{ss} \mathbin{\setminus\!\circ} \dot{\Pi}_a^{ss})$$

As a result, S2 + S3 $\geq 0$. □

These propositions are composed to prove monotonic safety in models with one BP and distance-independent perception.

Corollary A.6. *For AEBS with one BP, distance-independent perception, and shifts* $(\Delta_d, 0)$ *and* $(0, \Delta_v)$*, collision safety is monotonic.*

Proof. For shift $(\Delta_d, 0)$, we apply Propositions A.2 to A.4 to get that $S1 + S2 + S3 \geq 0$. For shift $(0, \Delta_v)$, we apply Propositions A.2, A.4 and A.5 to get that $S1 + S2 + S3 \geq 0$. □

## A.4 Dimension Removal Conservatism Proof

First, we introduce the following lemma:

Lemma A.7. *For the AEBS system in a constant braking region, applying a control sequence of (brake, no brake) results in a greater distance to the obstacle than a control sequence of (no brake, brake).*

Proof. Let the car be $d$ meters away from the obstacle and have speed $v$m/s. Assume the controller runs every $\delta_t$ seconds. Now assume the car brakes for one action, then doesn't brake for the next. After these two actions the car will be $d - \delta_t * v - \delta_t * (v - \delta_t * b)$ meters away and have speed $v - \delta_t * b$.

However, if the car doesn't brake, then brakes, the car will end up with distance $d - 2 * \delta_t * v$ and speed $v - \delta_t * b$. So braking first puts the car at a farther distance from the obstacle with an identical speed. □

Theorem A.8. *Given the MoS assumptions $\succeq_d$ and $\succeq_v$, the $\mathsf{M}_{dr}$ abstraction is conservative for the AEBS system: $Pr_{\mathsf{M}_{dr}}(\psi_{nocol}) \leq Pr_{\mathsf{M}_{cpl}}(\psi_{nocol})$, where $\mathsf{M}_{cpl}$ is the actual system.*

Proof. At a high level, the purpose of this abstraction is to first divide a known safe trace of the AEBS into a sequence of distance and velocity values: $(d_1, v_1), \ldots, (d_n, v_n)$. Now consider pair $(d_i, v_i)$. The algorithm then determines some number of braking actions $K$ and total control iterations $N$, such that when any possible ordering of the $K$ braking actions out of $N$ control iterations are applied to the car starting in state $(d_i, v_i)$ will bring the car to some new state $(d'_i, v'_i)$ such that $d'_i \geq d_{i+1}$ and $v'_i \leq v_{i+1}$. If the perception returns at least $K_i$ detections out of $N_i$ readings, then the model moves on to the next interval. If not, then the model enters a fail state and returns a violation of $\psi_{nocol}$. Finally, note that the braking region for the AEBS is fixed in this interval, so there is a 1-1 correspondence between perception identifying the obstacle and the controller applying a braking action.

First, we need to prove that the overall idea is conservative. Let the model be at the beginning of the $i^{th}$ interval. Thus, the distance $d$ and velocity $v$ are such that $d \geq d_i$ and $v \leq v_i$. But by our MoS assumption, the distance and velocity values which result in the lowest chance of safety are $d = d_i$ and $v = v_i$. So the implicit rounding that occurs between intervals preserves conservatism.

Next, we need to prove that the $K_i$ and $N_i$ chosen by the algorithm do indeed ensure that when at least $K_i$ LEC detections occur in $N_i$ LEC readings, the car will end up in a state with distance $d'_i \geq d_{i+1}$ and $v'_i \leq v_{i+1}$. The algorithm first considers the sequence of braking/no braking actions from the original safe trace. Let $N$ be the total number of control iterations in the interval, of which $K$ are braking actions. It then moves every no braking action to the beginning of the trace and then checks if the resulting distance $d'_i$ and velocity $v'_i$ satisfy $d'_i \geq d_{i+1}$ and $v'_i \leq v_{i+1}$. If so, then the model determines that the LEC must produce $K$ detections out of $N$ readings in this interval. Note that by lemma 1, if the trace of $K$ braking actions out of $N$ control iterations with all $N - K$ no braking actions at the front results in $d'_i \geq d_{i+1}$, then any ordering of $K$ braking actions out of $N$ control iterations results in $d'_i \geq d_{i+1}$.

If $d'_i \geq d_{i+1}$ and $v'_i \leq v_{i+1}$ does not hold, then the algorithm removes no braking actions from the front of the trace until it holds. Let $m$ be the total number of no braking actions removed. Then the model stipulates that the controller must have $K$ braking actions out of $N - m$ control iterations. Once again, with all the braking actions at the front of the trace, our distance and velocity requirement hold for any ordering of $K$ detections out of $N - m$ readings.

So the model $\mathsf{M}_{\text{dr}}$ only accepts traces which ensure that the distance to the obstacle is greater than some intermediate value for each interval in the original safe trace. Consider the last interval. It ends with the car having position distance and 0 velocity. So any LEC trace which is accepted by $\mathsf{M}_{\text{dr}}$ also results in the car have position distance and 0 velocity. So $\mathsf{M}_{\text{dr}}$ only LEC traces which are safe on the original model. Therefore, it is a conservative abstraction.

□

### A.5 AEBS Evaluation Setup Details

We constructed a probabilistic representation $\mathsf{M}_{\text{per}}$ for YoloNet as follows. We ran 500 simulations of a car approaching the obstacle at a constant speed, starting at 200 meters away from the obstacle. This resulted in a dataset of 180500 distance, detection/misdetection pairs. From these data we computed a probabilistic model by conditioning the detection probabilities on the distance to the obstacle and the previous readings, using up to the three previous readings if the amount of data permitted. Otherwise, we limited it to fewer readings.

We used the following algorithm for constructing $\mathsf{M}_{\text{dr}}$ abstractions. Starting with a safe trace ($S_{safe}$) of high-level detections in all the time steps, the set of $m$ intervals is obtained by selecting distances ($d_i$) corresponding to the time step with 0 BP control action followed with a non-zero BP ($B_1$ or $B_2$) control action at the next time step. With the same set of intervals, the $n$ pairs of $(K_i, N_i)$ are generated by varying the starting position for detection (or range of detection) by YoloNet. These positions are generated by equally dividing the distance $= dist - braking\ distance\ for\ B_2$ into $n$ positions. The $(K_i, N_i \mid i = 1...m)$ pairs are generated by the following algorithm for each interval. With $(K_i, N_i \mid i = 1...m)$ initialized from the $S_{safe}$, we shift all the 0 BP actions to the start of the interval. This is the worst case analysis for the controller as it will allow the car to cover maximum distance (with the maximum allowed velocity) in that interval. If the controller still gets $K_i$ (from $S_{safe}$) samples for its braking actions in the interval, then return $(K_i, N_i)$. Otherwise, reduce $N_i$ by 1 and repeat the algorithm. This is worst case analysis for the detector as we are reducing the total number of chances for the detection of the obstacle in the interval.