A Hypothesis on Black Swan in Unchanging Environments

Hyunin Lee[1]
With Chanwoo Park[2], David Abel[3], Ming Jin[4]
[1]UC Berkeley, [2]MIT, [3]Google DeepMind, [4]Virginia Tech

December 16, 2024

Machine Learning Safety

Black swans: rare but extremely high-risk events [Tal10]

Machine Learning Safety

Black swans: rare but extremely high-risk events [Tal10]

1. The COVID-19 pandemic [Ant20].

Machine Learning Safety

Black swans: rare but extremely high-risk events [Tal10]

1. The COVID-19 pandemic [Ant20].
2. Automated trading systems that overreact to market anomalies [KKST17, Phi21, Sta22].

Machine Learning Safety

Black swans: rare but extremely high-risk events [Tal10]

1. The COVID-19 pandemic [Ant20].
2. Automated trading systems that overreact to market anomalies [KKST17, Phi21, Sta22].
3. Unexpected bankruptcies [WPM14, ABG23].

Machine Learning Safety

Black swans: rare but extremely high-risk events [Tal10]

1. The COVID-19 pandemic [Ant20].
2. Automated trading systems that overreact to market anomalies [KKST17, Phi21, Sta22].
3. Unexpected bankruptcies [WPM14, ABG23].
4. Failures in monitoring hypoglycemic events in healthcare [WLY+23].

Machine Learning Safety

Black swans: rare but extremely high-risk events [Tal10]

1. The COVID-19 pandemic [Ant20].
2. Automated trading systems that overreact to market anomalies [KKST17, Phi21, Sta22].
3. Unexpected bankruptcies [WPM14, ABG23].
4. Failures in monitoring hypoglycemic events in healthcare [WLY+23].

Forecasting black swans are still vulnerable regardless of an algorithm's representation capacity or scalability [Cho19, SN20, HZC21, LQL+23, YZLL24].

A Hypothesis on Black Swan

Current approaches:

- algorithm improvement based on conventional belief that such events primarily arise from **dynamic, time-varying** environments [Pre19, ABK20, DMS21, WDFLP21, BD24, Jin24]

A Hypothesis on Black Swan

Current approaches:

- algorithm improvement based on conventional belief that such events primarily arise from **dynamic, time-varying** environments [Pre19, ABK20, DMS21, WDFLP21, BD24, Jin24]

This talk claims the above approach should be re-examined since conventional belief might be wrong.

A Hypothesis on Black Swan

Current approaches:

- algorithm improvement based on conventional belief that such events primarily arise from **dynamic, time-varying** environments [Pre19, ABK20, DMS21, WDFLP21, BD24, Jin24]

This talk claims the above approach should be re-examined since conventional belief might be wrong.

Hypothesis 1

Black swans can originate from misperceptions of an event's reward and likelihood, even within **static, stationary** *environments.*

Example

Lehman Brothers Bankrupt (2008)

- Most significant black swan event in the financial industry [WPM14]

Example

Lehman Brothers Bankrupt (2008)

- Most significant black swan event in the financial industry [WPM14]
- The firm declared bankruptcy within 72 hours without any precursor [MR09], and the only factor that changed during those three days was investors' perception of the company [Hou23, Maw14, FS14]

Example

Lehman Brothers Bankrupt (2008)

- Most significant black swan event in the financial industry [WPM14]
- The firm declared bankruptcy within 72 hours without any precursor [MR09], and the only factor that changed during those three days was investors' perception of the company [Hou23, Maw14, FS14]
- The bank's loss endurance, evaluated at 11.7% by the U.S. government, stayed *stationary, static* over the 72 hours.

Example

Lehman Brothers Bankrupt (2008)

- Most significant black swan event in the financial industry [WPM14]
- The firm declared bankruptcy within 72 hours without any precursor [MR09], and the only factor that changed during those three days was investors' perception of the company [Hou23, Maw14, FS14]
- The bank's loss endurance, evaluated at 11.7% by the U.S. government, stayed *stationary, static* over the 72 hours.
- Investors making rational decisions on the false market perception which appeared rational at the time but proved irrational by correcting their perception in hindsight

Main Contribution

- Define black swan events in stationary environments as **S-BLACK SWAN**.

> **(Informal)** *An* S-BLACK SWAN *event is a state-action pair where humans misperceive both its likelihood and reward. It is perceived as impossible, despite occurring with small probability, while its reward is overestimated relative to its true value in a stationary environment.*

Main Contribution

- Define black swan events in stationary environments as **S-BLACK SWAN**.

> **(Informal)** *An* S-BLACK SWAN *event is a state-action pair where humans misperceive both its likelihood and reward. It is perceived as impossible, despite occurring with small probability, while its reward is overestimated relative to its true value in a stationary environment.*

- A case study on how S-BLACK SWAN emerge and cause suboptimality gaps in various MDP settings, such as bandit (Theorem 8), small state spaces (Theorem 9), and large state spaces (Theorem 10).

Main Contribution

- Our main finding (Theorem 14) shows that even with zero estimation error, a lower bound on approximating the true optimal policy remains due to **perception error**, influenced by *reward misperception*, the *size of the* S-BLACK SWAN *set*, and their *minimum probability of occurrence*.

Main Contribution

- Our main finding (Theorem 14) shows that even with zero estimation error, a lower bound on approximating the true optimal policy remains due to **perception error**, influenced by *reward misperception*, the *size of the* S-BLACK SWAN *set*, and their *minimum probability of occurrence*.
- Theorem 15 examines S-BLACK SWAN hitting time provides an guide on how often a human should correct their internal perception.

Main Contribution

- Our main finding (Theorem 14) shows that even with zero estimation error, a lower bound on approximating the true optimal policy remains due to **perception error**, influenced by *reward misperception*, the *size of the* S-BLACK SWAN *set*, and their *minimum probability of occurrence*.
- Theorem 15 examines S-BLACK SWAN hitting time provides an guide on how often a human should correct their internal perception.
- Suggestions on design of future safe machine learning algorithms.

Contents

Preliminary

Markov Decision Process.

- **Definition:** $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, P, R, T \rangle$,
  - $\mathcal{S}$: State space, $\mathcal{A}$: action space, T: horizon length.
  - $P = \{P_t\}_{t=1}^{T}, P_t : \mathcal{S} \times \mathcal{A} \to \Delta(\mathcal{S})$ is a transition probability.
  - $R = \{R_t\}_{t=1}^{T}, R_t : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ is a reward function.
- $\boldsymbol{P^{\pi}(s, a)}$: **visitation probability** of (s, a) by planning with π in the world P.
- If $P_{t_1} = P_{t_2}, R_{t_1} = R_{t_2}$ for any $t_1, t_2 \in [T]$, we say stationary environment or otherwise we say non-statinoary environment.

Preliminary

Markov Decision Process.

- **Definition:** $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, P, R, T \rangle$,
  - $\mathcal{S}$: State space, $\mathcal{A}$: action space, T: horizon length.
  - $P = \{P_t\}_{t=1}^{T}, P_t : \mathcal{S} \times \mathcal{A} \to \Delta(\mathcal{S})$ is a transition probability.
  - $R = \{R_t\}_{t=1}^{T}, R_t : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ is a reward function.
- $\boldsymbol{P^{\pi}(s,a)}$: **visitation probability** of (s,a) by planning with π in the world P.
- If $P_{t_1} = P_{t_2}, R_{t_1} = R_{t_2}$ for any $t_1, t_2 \in [T]$, we say stationary environment or otherwise we say non-statinoary environment.
- **Setting:** With policy (decision) $\pi : \mathcal{S} \to \Delta(\mathcal{A})$, the agent gathers a trajectory $\{s_0, a_0, r_1, s_1, a_1, r_2, \cdots, s_{T-1}, a_{T-1}, r_{T-1}, s_T\}$ where $a_t \sim \pi(\cdot|s_t),\ s_{t+1} \sim P_t(\cdot|s_t, a_t),\ r_t = R_t(s_t, a_t)$.

Preliminary

Markov Decision Process.

- **Definition:** $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, P, R, T \rangle$,
  - $\mathcal{S}$: State space, $\mathcal{A}$: action space, T: horizon length.
  - $P = \{P_t\}_{t=1}^{T}, P_t : \mathcal{S} \times \mathcal{A} \to \Delta(\mathcal{S})$ is a transition probability.
  - $R = \{R_t\}_{t=1}^{T}, R_t : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ is a reward function.
- $\boldsymbol{P^\pi(s,a)}$: **visitation probability** of (s, a) by planning with π in the world P.
- If $P_{t_1} = P_{t_2}, R_{t_1} = R_{t_2}$ for any $t_1, t_2 \in [T]$, we say stationary environment or otherwise we say non-statinoary environment.
- **Setting:** With policy (decision) $\pi : \mathcal{S} \to \Delta(\mathcal{A})$, the agent gathers a trajectory $\{s_0, a_0, r_1, s_1, a_1, r_2, \cdots, s_{T-1}, a_{T-1}, r_{T-1}, s_T\}$ where $a_t \sim \pi(\cdot|s_t),\ s_{t+1} \sim P_t(\cdot|s_t, a_t),\ r_t = R_t(s_t, a_t)$.
- **Goal:** find optimal π^* that maximizes the value function $V^\pi_{\mathcal{M}} := \mathbb{E}_\pi \left[\sum_{t=0}^{T-1} R_t(s_t, a_t) | P_t, \right]$.

Preliminary

Following three theorems lay groundworks for *misperception* in Hypothesis 1.

1. Expected Utility Theory

- Explains the human's rational choice under uncertainty [vN44].
- Outcome space $\mathcal{O} = \{o_1, o_2, \cdots, o_K\}$.
- Utility function $g : \mathcal{O} \to \mathbb{R}$ represents gain or loss of outcomes.
- Choice $c \in \mathcal{C}$: returns outcomes o_i with probabilities $p_i^{(c)}$.
- EUT evalutes the riskiness of choice c as $V(c) = \sum_{i=1}^{K} g(o_i) p_i$, then human choose $c^\star$ that $\max_{c \in \mathcal{C}} V(C)$ [Rab13].

Preliminary

Following three theorems lay groundworks for *misperception* in Hypothesis 1.

1. Expected Utility Theory

- Explains the human's rational choice under uncertainty [vN44].
- Outcome space $\mathcal{O} = \{o_1, o_2, \cdots, o_K\}$.
- Utility function $g : \mathcal{O} \to \mathbb{R}$ represents gain or loss of outcomes.
- Choice $c \in \mathcal{C}$: returns outcomes o_i with probabilities $p_i^{(c)}$.
- EUT evalutes the riskiness of choice c as $V(c) = \sum_{i=1}^{K} g(o_i)p_i$, then human choose $c^\star$ that $\max_{c \in \mathcal{C}} V(C)$ [Rab13].
- (Ex) stock market investment scenario:
  - $\mathcal{O} = \{\text{Economic Boom (EB)}, \text{Economic Recession (ER)}\}$
  - $g(EB) = +100$, $g(ER) = -1000$.
  - $\mathcal{C} = \{\text{invest in stocks, invest in bonds, keep cash}\}$ with different probabilities $(p_1^{(c)}, p_2^{(c)})$.

Preliminary

2. Prospect Theorem

- *EUT* fails to account for empirical observations from psychological experiments [DT16, PAPAB19, WS00, VYD05, vdMKV22] and economic cases [Rog98, WW07, Bet22] that demonstrate human irrationality.

Preliminary

2. Prospect Theorem

- *EUT* fails to account for empirical observations from psychological experiments [DT16, PAPAB19, WS00, VYD05, vdMKV22] and economic cases [Rog98, WW07, Bet22] that demonstrate human irrationality.
- Internally distorts event probability $p_i^{(c)}$ and event value $g(o_i)$ for any $c \in C$, $\forall i \in [k]$.

Preliminary

2. Prospect Theorem

- *EUT* fails to account for empirical observations from psychological experiments [DT16, PAPAB19, WS00, VYD05, vdMKV22] and economic cases [Rog98, WW07, Bet22] that demonstrate human irrationality.
- Internally distorts event probability $p_i^{(c)}$ and event value $g(o_i)$ for any $c \in C$, $\forall i \in [k]$.
- (ex1) $-1M$ or $+1M$. (ex2) buying Powerball lottery.

Preliminary

2. Prospect Theorem

- *EUT* fails to account for empirical observations from psychological experiments [DT16, PAPAB19, WS00, VYD05, vdMKV22] and economic cases [Rog98, WW07, Bet22] that demonstrate human irrationality.
- Internally distorts event probability $p_i^{(c)}$ and event value $g(o_i)$ for any $c \in C$, $\forall i \in [k]$.
- (ex1) $-1M$ or $+1M$. (ex2) buying Powerball lottery.
- Probability distortion function $w : [0,1] \to [0,1]$.
- Value distortion function $u : \mathbb{R} \to \mathbb{R}$.
- *PT* evaluates the choice c as $V(c) = \sum_{i=1}^{K} u(g(o_i))w(p_i^{(c)})$ [KT13, FW97].

Preliminary

3. Cumulative Prospect Theorem

- To enhance mathematical rigor—specifically (ensuring that distorted probabilities still sum to one), *Prospect Theory (PT)* was further revised into *Cumulative Prospect Theory (CPT)*.
- *CPT* distorts the cumulative probability rather than the probability itself.
- *CPT* evaluates the choice c as $V(c) = \sum_{i=1}^{K} u(g(o_i))\Big(w\big(\sum_{j=1}^{i} p_j^{(c)}\big) - w\big(\sum_{j=1}^{i-1} p_j^{(c)}\big)\Big)$.

Preliminary

3. Cumulative Prospect Theorem

Example 1 (Insurance policies)

Consider an example where the probability of an insured risk is 1%, the potential loss is $1,000$, and the insurance premium is 15. According to CPT, most would opt to pay the 15 premium to avoid the larger loss.

Preliminary

3. Cumulative Prospect Theorem

Example 1 (Insurance policies)

Consider an example where the probability of an insured risk is 1%, the potential loss is 1,000, and the insurance premium is 15. According to CPT, most would opt to pay the 15 premium to avoid the larger loss.

- Two-step Markov Decision Process.
- $\mathcal{S} = \left(s_{base}, s_{preminum}, s_{risk}\right) \rightarrow$ outcome space $\mathcal{O}$.
- $\mathcal{A} = \{a_p, a_{np}\} \rightarrow$ choice set $\mathcal{C}$.
- EUT returns
  - $V(a_{np}) = -1000 \cdot 0.01 = -10$
  - $V(a_p) = -15 \cdot 1 = -15$
- But human choose a_p ☺.

Value distortion function

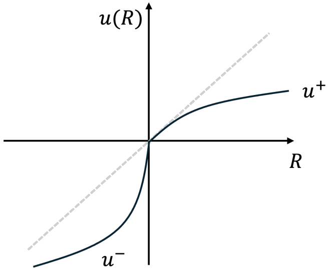


Figure: Value distortion function. Gray line represents $y = x$.

Definition 2 (Value Distortion Function)

The value distortion function u is defined as:

$$u(x) = \begin{cases} u^+(x) & \text{if } x \geq 0, \\ u^-(x) & \text{if } x < 0, \end{cases}$$

where

- $u^+ : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is non-decreasing, concave with $\lim_{h \to 0^+} (u^+)'(h) \leq 1$
- $u^- : \mathbb{R}_{\leq 0} \to \mathbb{R}_{\leq 0}$ is non-decreasing, convex with $\lim_{h \to 0^-} (u^-)'(h) > 1$

Value distortion function

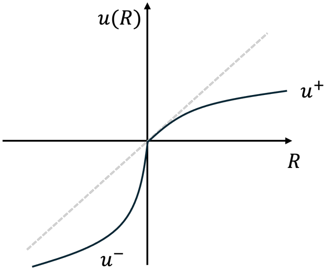


Figure: Value distortion function. Gray line represents $y = x$.

Definition 3 (Value Distortion Function)

The value distortion function u is defined as:

$$u(x) = \begin{cases} u^+(x) & \text{if } x \geq 0, \\ u^-(x) & \text{if } x < 0, \end{cases}$$

where

- $u^+ : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is non-decreasing, concave with $\lim_{h \to 0^+} (u^+)'(h) \leq 1$
- $u^- : \mathbb{R}_{\leq 0} \to \mathbb{R}_{\leq 0}$ is non-decreasing, convex with $\lim_{h \to 0^-} (u^-)'(h) > 1$

Value distortion function

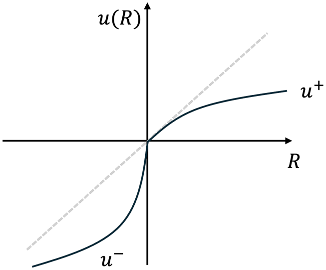


Figure: Value distortion function. Gray line represents $y = x$.

Definition 4 (Value Distortion Function)

The value distortion function u is defined as:

$$u(x) = \begin{cases} u^+(x) & \text{if } x \geq 0, \\ u^-(x) & \text{if } x < 0, \end{cases}$$

where

- $u^+ : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is non-decreasing, concave with $\lim_{h \to 0^+} (u^+)'(h) \leq 1$
- $u^- : \mathbb{R}_{\leq 0} \to \mathbb{R}_{\leq 0}$ is non-decreasing, convex with $\lim_{h \to 0^-} (u^-)'(h) > 1$

Probability distortion function

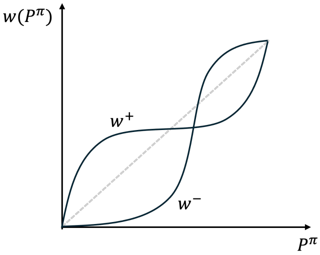


Figure: Probability distortion function. Gray line represents $y = x$.

Definition 5 (Probability Distortion Function)

The probability distortion function w is defined as:

$$w(p_i) = \begin{cases} w^+(p_i) & \text{if } g(x_i) \geq 0, \\ w^-(p_i) & \text{if } g(x_i) < 0, \end{cases}$$

where $w^+, w^- : [0,1] \to [0,1]$ satisfy:

- $w^+(0) = w^-(0) = 0, \; w^+(1) = w^-(1) = 1$
- $w^+(a) = a$ and $w^-(b) = b$ for some $a, b \in (0,1)$
- $(w^+)'(x)$ is decreasing on $[0,a)$ and increasing on $(a,1]$ and $(w^-)'(x)$ is increasing on $[0,b)$ and decreasing on $(b,1]$.

Probability distortion function

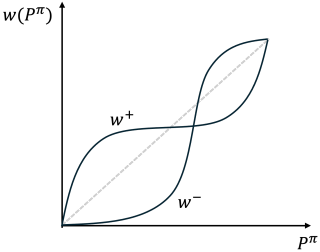


Figure: Probability distortion function. Gray line represents $y = x$.

Definition 6 (Probability Distortion Function)

The probability distortion function w is defined as:

$$w(p_i) = \begin{cases} w^+(p_i) & \text{if } g(x_i) \geq 0, \\ w^-(p_i) & \text{if } g(x_i) < 0, \end{cases}$$

where $w^+, w^- : [0,1] \to [0,1]$ satisfy:

- $w^+(0) = w^-(0) = 0$, $w^+(1) = w^-(1) = 1$
- $w^+(a) = a$ and $w^-(b) = b$ for some $a, b \in (0,1)$
- $(w^+)'(x)$ is decreasing on $[0,a)$ and increasing on $(a,1]$ and $(w^-)'(x)$ is increasing on $[0,b)$ and decreasing on $(b,1]$.

Probability distortion function

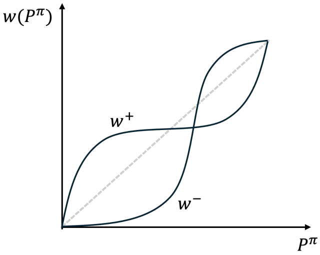


Figure: Probability distortion function. Gray line represents $y = x$.

Definition 7 (Probability Distortion Function)

The probability distortion function w is defined as:

$$w(p_i) = \begin{cases} w^+(p_i) & \text{if } g(x_i) \geq 0, \\ w^-(p_i) & \text{if } g(x_i) < 0, \end{cases}$$

where $w^+, w^- : [0,1] \to [0,1]$ satisfy:

- $w^+(0) = w^-(0) = 0,\ w^+(1) = w^-(1) = 1$
- $w^+(a) = a$ and $w^-(b) = b$ for some $a, b \in (0,1)$
- $(w^+)'(x)$ is decreasing on $[0, a)$ and increasing on $(a, 1]$ and $(w^-)'(x)$ is increasing on $[0, b)$ and decreasing on $(b, 1]$.

Emergence of S-BLACK SWAN in Sequential Decision Making

Case studies to substantiate Hypothesis 1. The goal of this section is to see how policy deviates due to misperception.

- Function u distorts the reward $R(s, a)$
- Function w distorts the transition probabilities $\{P(s'|s, a)\}_{\forall s' \in \mathcal{S}}$ where s' is the next state.
- Let $\mathcal{M}$ represent the real-world, and define the distorted MDP $\mathcal{M}_d := \langle \mathcal{S}, \mathcal{A}, w(P), u(R), \gamma \rangle$, where u and w introduce distortions in the R and P of $\mathcal{M}$.

Emergence of S-BLACK SWAN in Sequential Decision Making

Case 1. Contextual Bandit ($T = 1$) [LS20]

Theorem 8 (One-Step Optimality Deviation)

If $T = 1$, then the optimal policy in the MDP $\mathcal{M}$ is identical to the optimal policy in the distorted MDP $\mathcal{M}_d$.

Emergence of S-BLACK SWAN in Sequential Decision Making

Case 1. Contextual Bandit ($T = 1$) [LS20]

Theorem 8 (One-Step Optimality Deviation)

If $T = 1$, then the optimal policy in the MDP $\mathcal{M}$ is identical to the optimal policy in the distorted MDP $\mathcal{M}_d$.

- Somewhat counterintuitive. Recall Example 1.

Emergence of S-BLACK SWAN in Sequential Decision Making

Case 1. Contextual Bandit ($T = 1$) [LS20]

Theorem 8 (One-Step Optimality Deviation)

If $T = 1$, then the optimal policy in the MDP $\mathcal{M}$ is identical to the optimal policy in the distorted MDP $\mathcal{M}_d$.

- Somewhat counterintuitive. Recall Example 1.
- Reason: u preserves ordering, i.e.,
  $r(s_{loss}) < r(s_{premium}) < r(s_{base}) \to u^-(r(s_{loss})) < u^-(r(s_{premium})) < u^-(r(s_{base}))$,
  since u^- is non-decreasing.

Emergence of S-BLACK SWAN in Sequential Decision Making

Case 1. Contextual Bandit ($T = 1$) [LS20]

Theorem 8 (One-Step Optimality Deviation)

If $T = 1$, then the optimal policy in the MDP $\mathcal{M}$ is identical to the optimal policy in the distorted MDP $\mathcal{M}_d$.

- Somewhat counterintuitive. Recall Example 1.
- Reason: u preserves ordering, i.e., $r(s_{loss}) < r(s_{premium}) < r(s_{base}) \to u^{-}(r(s_{loss})) < u^{-}(r(s_{premium})) < u^{-}(r(s_{base}))$, since u^{-} is non-decreasing.
- Suggests that a *short* decision horizon may *reduce* the impact of human irrationality.

Emergence of S-BLACK SWAN in Sequential Decision Making

Case 2. $|\mathcal{S}| = 2$ **when** $T > 1$

Theorem 9 (Multi-step Optimality Deviation with $|\mathcal{S}| = 2$)

If $|\mathcal{S}| = 2$, then the optimal policy from the MDP $\mathcal{M}$ is also identical to the optimal policy of the distorted MDP $\mathcal{M}_d$ for all $t \in [T]$.

Emergence of S-BLACK SWAN in Sequential Decision Making

Case 2. $|\mathcal{S}| = 2$ **when** $T > 1$

Theorem 9 (Multi-step Optimality Deviation with $|\mathcal{S}| = 2$)

If $|\mathcal{S}| = 2$, then the optimal policy from the MDP $\mathcal{M}$ is also identical to the optimal policy of the distorted MDP $\mathcal{M}_d$ for all $t \in [T]$.

- Counterintuitive due to model-error propagation [JFZL19].

Emergence of S-BLACK SWAN in Sequential Decision Making

Case 2. $|\mathcal{S}| = 2$ **when** $T > 1$

Theorem 9 (Multi-step Optimality Deviation with $|\mathcal{S}| = 2$)

If $|\mathcal{S}| = 2$, then the optimal policy from the MDP $\mathcal{M}$ is also identical to the optimal policy of the distorted MDP $\mathcal{M}_d$ for all $t \in [T]$.

- Counterintuitive due to model-error propagation [JFZL19].
- Reason: w preserves ordering, i.e., if $P(s_1|s,a) > P(s_2|s,a)$, then $w(P(s_1|s,a)) > w(P(s_2|s,a))$, where $\mathcal{S} = \{s_1, s_2\}$.

Emergence of S-BLACK SWAN in Sequential Decision Making

Case 2. $|\mathcal{S}| = 2$ **when** $T > 1$

Theorem 9 (Multi-step Optimality Deviation with $|\mathcal{S}| = 2$)

If $|\mathcal{S}| = 2$, then the optimal policy from the MDP $\mathcal{M}$ is also identical to the optimal policy of the distorted MDP $\mathcal{M}_d$ for all $t \in [T]$.

- Counterintuitive due to model-error propagation [JFZL19].
- Reason: w preserves ordering, i.e., if $P(s_1|s,a) > P(s_2|s,a)$, then $w(P(s_1|s,a)) > w(P(s_2|s,a))$, where $\mathcal{S} = \{s_1, s_2\}$.
- Suggests that a *small* state space requires relatively *low* informational complexity to determine the real-world optimal action.

Emergence of S-BLACK SWAN in Sequential Decision Making

Case 3. $|\mathcal{S}| = 3$ **with unbiased reward perception**

Theorem 10 (Two-step Optimality Deviation with $|\mathcal{S}| = 3$)

If $|\mathcal{S}| = 3$ and $T = 2$, there exists a transition probability P and a reward function R such that the optimal policy of the MDP $\mathcal{M}$ differs from that of the distorted MDP $\mathcal{M}_d$.

- Now aligns with the empirical observation in model-based reinforcement learning; increasing suboptimality is caused by model error propagation [JFZL19]

Emergence of S-BLACK SWAN in Sequential Decision Making

Case 3. $|\mathcal{S}| = 3$ **with unbiased reward perception**

Theorem 10 (Two-step Optimality Deviation with $|\mathcal{S}| = 3$)

If $|\mathcal{S}| = 3$ and $T = 2$, there exists a transition probability P and a reward function R such that the optimal policy of the MDP $\mathcal{M}$ differs from that of the distorted MDP $\mathcal{M}_d$.

- Now aligns with the empirical observation in model-based reinforcement learning; increasing suboptimality is caused by model error propagation [JFZL19]

Summary.

Theorems 8, 9, and 10 demonstrate that the discrepancy between $\pi^{\dagger,\star}$ and $\pi^{\star}$ increases as the complexity of the environment ($\mathcal{S}$) or the horizon length (T) increases.

Ground MDP, Human MDP, Human-Estimation MDP

To explore Hypothesis 1, we propose three different MDPs.

1. Ground MDP

- *stationary* ground MDP (GMDP) $\mathcal{M}$ is an abstraction of real-world environments without information loss.
- mathematically identical with $\mathcal{M}$ defintion.

Ground MDP, Human MDP, Human-Estimation MDP

2. Human MDP

- $\mathcal{M}^{\dagger} := \langle \mathcal{S}, \mathcal{A}, P^{\dagger}, R^{\dagger}, \gamma, T \rangle$
  - $P^{\dagger,\pi}$: misperceived visitation probability $P^{\pi}(s, a)$ through the function w.
  - $R^{\dagger}$: misperceived reward function $R(s, a)$ through the function u.
- Internal assumption: $\mathcal{S}^{\dagger} = \mathcal{S}$ and $\mathcal{A}^{\dagger} = \mathcal{A}$.

Ground MDP, Human MDP, Human-Estimation MDP

2. Human MDP

- $\mathcal{M}^\dagger := \langle \mathcal{S}, \mathcal{A}, P^\dagger, R^\dagger, \gamma, T \rangle$
  - $P^{\dagger,\pi}$: misperceived visitation probability $P^\pi(s, a)$ through the function w.
  - $R^\dagger$: misperceived reward function $R(s, a)$ through the function u.
- Internal assumption: $\mathcal{S}^\dagger = \mathcal{S}$ and $\mathcal{A}^\dagger = \mathcal{A}$.
- $\mathcal{M}^\dagger$ perceives $\mathcal{M}$ by u, w:

$$\int P^{\dagger,\pi}(s,a) = \begin{cases} w^+\left(\int P^\pi(s,a)\right) \text{ if } R(s,a) \geq 0 \\ w^-\left(\int P^\pi(s,a)\right) \text{ if } R(s,a) < 0 \end{cases} \tag{1}$$

$$R^\dagger(s,a) = \begin{cases} u^+(R(s,a)) \text{ if } R(s,a) \geq 0 \\ u^-(R(s,a)) \text{ if } R(s,a) < 0 \end{cases} \tag{2}$$

- $V^\pi_{\mathcal{M}^\dagger}(s) := \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t R^\dagger(s_t, a_t) \middle| P^\dagger, \pi, s_0 = s\right]$.

Ground MDP, Human MDP, Human-Estimation MDP

2. Human Estimation MDP

- why distortions occur in visitation probability (P^{π}) rather than transition probability (P).

Ground MDP, Human MDP, Human-Estimation MDP

2. Human Estimation MDP

- why distortions occur in visitation probability (P^{π}) rather than transition probability (P).
- reason: (s, a) is an event unit, and a distortion in transition probability implies a distortion in the state space for a given previous state and action pair.

Ground MDP, Human MDP, Human-Estimation MDP

2. Human Estimation MDP

- why distortions occur in visitation probability (P^{π}) rather than transition probability (P).
- reason: (s, a) is an event unit, and a distortion in transition probability implies a distortion in the state space for a given previous state and action pair.
- The central question is how distortions in visitation probability relate directly to data collection.

Lemma 11

For a given $\mathcal{M}$, there always exists a function $h : \mathcal{S} \to \mathcal{S}$ such that $w\left(\int P^{\pi}(s, a)\right) = \int P^{\pi}(h(s), a)$ holds for any function w. That is $\mathcal{D}^{\dagger} = \{h(s_t), a_t, u(r_t), h(s_{t+1})\}_{t=0}^{T-1}$ is sampled from $\mathcal{M}^{\dagger}$

Ground MDP, Human MDP, Human-Estimation MDP

3. Human-Estimation MDP

- $\widehat{\mathcal{M}}^{\dagger} = \langle \mathcal{S}, \mathcal{A}, \widehat{P}^{\dagger}, \widehat{R}^{\dagger}, \gamma, T \rangle$.
    - $\widehat{P}^{\dagger,\pi}$: Estimated visitation probability of $P^{\dagger,\pi}$ by dataset $\mathcal{D}^{\dagger}$.
    - $\widehat{R}^{\dagger}$: Estimated reward of $R^{\dagger}$ by by dataset $\mathcal{D}^{\dagger}$.

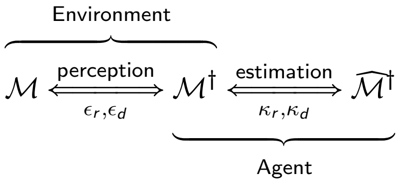


Figure: The agent and environment intersect with perception.

Ground MDP, Human MDP, Human-Estimation MDP

3. **Human-Estimation MDP**
   - $\widehat{\mathcal{M}}^{\dagger} = \langle \mathcal{S}, \mathcal{A}, \widehat{P}^{\dagger}, \widehat{R}^{\dagger}, \gamma, T \rangle$.
     - $\widehat{P}^{\dagger,\pi}$: Estimated visitation probability of $P^{\dagger,\pi}$ by dataset $\mathcal{D}^{\dagger}$.
     - $\widehat{R}^{\dagger}$: Estimated reward of $R^{\dagger}$ by by dataset $\mathcal{D}^{\dagger}$.
   - Estimation process is the same as estimation of the generative model in model-based reinforcement learning [GAMK13, SWW+18, AKY20, Kak03].
   - $V^{\pi}_{\widehat{\mathcal{M}}^{\dagger}}(s) :=$ $\mathbb{E}\left[\sum_{t=0}^{T} \gamma^t \widehat{R}^{\dagger}(s_t, a_t) \middle| \widehat{P}^{\dagger}, \pi, s_0 = s\right]$.

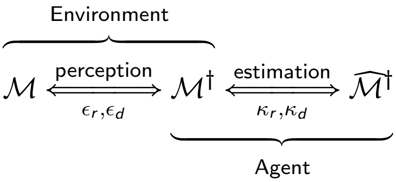


Figure: The agent and environment intersect with perception.

S-BLACK SWAN

1. **Discrete state and action space**
   - Order statistics : $R_{[1]} \leq R_{[2]} \leq \cdots \leq R_{[|\mathcal{S}||\mathcal{A}|]}$ and $P^{\pi}_{[1]} \leq P^{\pi}_{[2]} \leq \cdots \leq P^{\pi}_{[|\mathcal{S}||\mathcal{A}|]}$.
   - $R_{[I_r(s,a)]} = R(s,a)$ and $P^{\pi}_{[I_p(s,a)]} = P^{\pi}(s,a)$

S-BLACK SWAN

1. Discrete state and action space

- Order statistics : $R_{[1]} \leq R_{[2]} \leq \cdots \leq R_{[|\mathcal{S}||\mathcal{A}|]}$ and $P^{\pi}_{[1]} \leq P^{\pi}_{[2]} \leq \cdots \leq P^{\pi}_{[|\mathcal{S}||\mathcal{A}|]}$.
- $R_{[I_r(s,a)]} = R(s,a)$ and $P^{\pi}_{[I_p(s,a)]} = P^{\pi}(s,a)$

Definition 12 (S-BLACK SWAN - Discrete State and Action Space)

Given distortion functions u, w and constants $C_{bs} \gg 0$ and $\epsilon_{bs} > 0$, if (s,a) satisfies:

1. (High-risk): $R_{[I_r(s,a)]} - u^{-}(R_{[I_r(s,a)]}) < -C_{bs}$.
2. (Rare): $w^{-}\left(\sum_{j=1}^{I_p(s,a)} P^{\pi}_{[j]}\right) = w^{-}\left(\sum_{j=1}^{I_p(s,a)-1} P^{\pi}_{[j]}\right)$, yet $0 < P^{\pi}_{[I_p(s,a)]} < \epsilon_{bs}$.

then we define (s,a) as S-BLACK SWAN .

S-BLACK SWAN

Let's dig into the definition.

1. **(High-risk):** $R_{[I_r(s,a)]} - u^{-}(R_{[I_r(s,a)]}) < -C_{bs}$

S-BLACK SWAN

Let's dig into the definition.

1. **(High-risk):** $\boldsymbol{R_{[I_r(s,a)]} - u^{-}(R_{[I_r(s,a)]}) < -C_{bs}}$
   - $R_{[I_r(s,a)]}$: Ground truth reward of an event (s, a).

S-BLACK SWAN

Let's dig into the definition.

1. **(High-risk):** $\boldsymbol{R_{[I_r(s,a)]} - u^{-}(R_{[I_r(s,a)]}) < -C_{bs}}$
   - $R_{[I_r(s,a)]}$: Ground truth reward of an event (s,a).
   - $u^{-}(R_{[I_r(s,a)]})$: Perceived reward by agent.

S-BLACK SWAN

Let's dig into the definition.

1. **(High-risk):** $\boldsymbol{R_{[I_r(s,a)]} - u^-(R_{[I_r(s,a)]}) < -C_{bs}}$
   - $R_{[I_r(s,a)]}$: Ground truth reward of an event (s, a).
   - $u^-(R_{[I_r(s,a)]})$: Perceived reward by agent.
   - $R_{[I_r(s,a)]} + C_{bs} < u^-(R_{[I_r(s,a)]})$: Overestimation (optimistic perception) of an event's loss.
   - C_{bs} is given constant that quantifies distortion of u^-.

S-BLACK SWAN

Let's dig into the definition.

2. **(Rare):** $\boldsymbol{w^{-}\left(\sum_{j=1}^{I_p(s,a)} P^{\pi}_{[j]}\right) = w^{-}\left(\sum_{j=1}^{I_p(s,a)-1} P^{\pi}_{[j]}\right)}$, **yet** $\boldsymbol{0 < P^{\pi}_{[I_p(s,a)]} < \epsilon_{bs}}$

S-BLACK SWAN

Let's dig into the definition.

2. **(Rare):** $\boldsymbol{w^{-}\left(\sum_{j=1}^{I_p(s,a)} P^{\pi}_{[j]}\right) = w^{-}\left(\sum_{j=1}^{I_p(s,a)-1} P^{\pi}_{[j]}\right)}$, **yet** $\boldsymbol{0 < P^{\pi}_{[I_p(s,a)]} < \epsilon_{bs}}$
   - $P^{\pi}_{[I_p(s,a)]}$: Ground truth visitation probability of an event (s, a).

S-BLACK SWAN

Let's dig into the definition.

2. **(Rare):** $\boldsymbol{w^{-}\left(\sum_{j=1}^{I_p(s,a)} P^{\pi}_{[j]}\right) = w^{-}\left(\sum_{j=1}^{I_p(s,a)-1} P^{\pi}_{[j]}\right)}$, **yet** $\boldsymbol{0 < P^{\pi}_{[I_p(s,a)]} < \epsilon_{bs}}$
   - $P^{\pi}_{[I_p(s,a)]}$: Ground truth visitation probability of an event (s, a).
   - $\sum_{j=1}^{I_p(s,a)} P^{\pi}_{[j]}$: Ground truth cumulative visitation probability.

S-BLACK SWAN

Let's dig into the definition.

2. (Rare): $\boldsymbol{w^{-}\left(\sum_{j=1}^{l_p(s,a)} P^{\pi}_{[j]}\right) = w^{-}\left(\sum_{j=1}^{l_p(s,a)-1} P^{\pi}_{[j]}\right)}$, **yet** $\boldsymbol{0 < P^{\pi}_{[l_p(s,a)]} < \epsilon_{bs}}$

- $P^{\pi}_{[l_p(s,a)]}$: Ground truth visitation probability of an event (s, a).
- $\sum_{j=1}^{l_p(s,a)} P^{\pi}_{[j]}$: Ground truth cumulative visitation probability.
- $w^{-}(\sum_{j=1}^{l_p(s,a)} P^{\pi}_{[j]})$: Perceived cumulative visitation probability by agent.

S-BLACK SWAN

Let's dig into the definition.

2. **(Rare):** $\boldsymbol{w^{-}\left(\sum_{j=1}^{I_p(s,a)} P^{\pi}_{[j]}\right) = w^{-}\left(\sum_{j=1}^{I_p(s,a)-1} P^{\pi}_{[j]}\right)}$, **yet** $\boldsymbol{0 < P^{\pi}_{[I_p(s,a)]} < \epsilon_{bs}}$
   - $P^{\pi}_{[I_p(s,a)]}$: Ground truth visitation probability of an event (s, a).
   - $\sum_{j=1}^{I_p(s,a)} P^{\pi}_{[j]}$: Ground truth cumulative visitation probability.
   - $w^{-}(\sum_{j=1}^{I_p(s,a)} P^{\pi}_{[j]})$: Perceived cumulative visitation probability by agent.
   - $w^{-}(\sum_{j=1}^{I_p(s,a)} P^{\pi}_{[j]}) = w^{-}(\sum_{j=1}^{I_p(s,a)-1} P^{\pi}_{[j]})$: The agent perceives the event (s, a) as infeasible.

S-BLACK SWAN

Let's dig into the definition.

2. **(Rare):** $\boldsymbol{w^-\left(\sum_{j=1}^{I_p(s,a)} P^\pi_{[j]}\right) = w^-\left(\sum_{j=1}^{I_p(s,a)-1} P^\pi_{[j]}\right)}$, **yet** $\boldsymbol{0 < P^\pi_{[I_p(s,a)]} < \epsilon_{bs}}$
   - $P^\pi_{[I_p(s,a)]}$: Ground truth visitation probability of an event (s, a).
   - $\sum_{j=1}^{I_p(s,a)} P^\pi_{[j]}$: Ground truth cumulative visitation probability.
   - $w^-(\sum_{j=1}^{I_p(s,a)} P^\pi_{[j]})$: Perceived cumulative visitation probability by agent.
   - $w^-(\sum_{j=1}^{I_p(s,a)} P^\pi_{[j]}) = w^-(\sum_{j=1}^{I_p(s,a)-1} P^\pi_{[j]})$: The agent perceives the event (s, a) as infeasible.
   - $0 < P^\pi_{[I_p(s,a)]} < \epsilon_{bs}$: Feasible but with a small probability.
   - ϵ_{bs} is a given constant that also quantifies distortion of w^-.

S-BLACK SWAN

2. Continuous state and action space

- Suppose $R : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ is bijective.
- Recall $P^\pi : \mathcal{S} \times \mathcal{A} \to [0, 1]$.
- The probability $\mathbb{P}_r := R^{-1} \circ P^\pi : \mathbb{R} \to [0, 1]$ denotes the probability of a feasible reward induced by policy π.

Definition 13 (S-BLACK SWAN - Continuous State and Action Space)

Given distortion functions u, w and constants $C_{bs} \gg 0$ and $\epsilon_{bs} > 0$, if (s, a) satisfies:

1. $R(s, a) - u^-(R(s, a)) < -C_{bs}$.
2. $\frac{dw^-(x)}{dx}\Big|_{x=F(R(s,a))} \cdot \mathbb{P}_r(r = R(s, a)) = 0$, yet $0 < \mathbb{P}_r(r = R(s, a)) < \epsilon_{bs}$,

where $F(r) := \int_{-\infty}^{r} d\mathbb{P}_r$ is the cumulative distribution of $\mathbb{P}_r$, then we define (s, a) as S-BLACK SWAN .

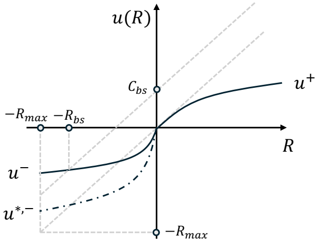
S-BLACK SWAN

1. The role of C_{bs} and ϵ_{bs}.

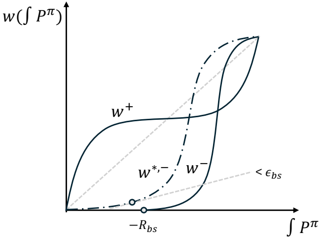
- magnitude of distortion
- the threshold of "safe perception"

S-BLACK SWAN

1. **The role of C_{bs} and ϵ_{bs}.**
   - magnitude of distortion
   - the threshold of "safe perception"
     - $\mathcal{B}$: the collection of all S-BLACK SWAN .
     - If $u^-(r) < r + C_{bs}$ for $\forall r$ then $\mathcal{B} = \emptyset$.
     - If $0 < w^-(p) < \epsilon_{bs}$ for $\forall p$ then $\mathcal{B} = \emptyset$.
   - $w_\star^-$ and $u_\star^-$: w^- and u^- that result in $\mathcal{B} = \emptyset \rightarrow$ safe perception



(a) distortion functions $u, u^\star$.



(b) distortion functions $w, w^\star$

S-BLACK SWAN

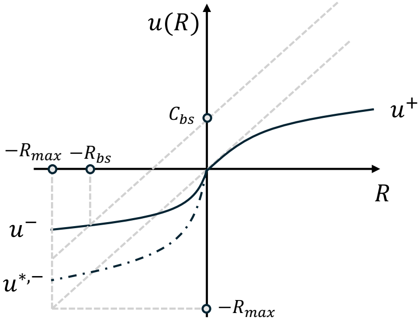


Figure: distortion functions $u, u^{\star}$.

2. The role of $-R_{bs}$.

- intersection between $u^{-}(r)$ and $r + C_{bs}$.
- $[-R_{max}, -R_{bs}]$ is feasible black swan candidates.
- $-R_{bs}$ controls the size of feasible black swan set.

S-BLACK SWAN

Theorem 14 (Convergence of value estimation gap but lower bound on value perception gap)

The asymptotic convergence of the value function estimation holds as follows,

$$V^{\pi}_{\widehat{\mathcal{M}}^{\dagger}}(s) \to V^{\pi}_{\mathcal{M}^{\dagger}}(s) \quad a.s. \quad as \quad T \to \infty, \ \forall s, \pi \in \mathcal{S} \times \Pi.$$

However, under specific conditions on $\epsilon_{bs}, \epsilon_{bs}^{\min}, R_{bs}$, *the lower bound of value perception gap as follows.*

$$|V^{\pi}_{\mathcal{M}^{\dagger}}(s) - V^{\pi}_{\mathcal{M}}(s)| = \Omega\left(\frac{\left((R_{\max} - R_{bs})\,\epsilon_{bs}^{\min} - R_{bs}\epsilon_{bs}\right)(R_{\max} - R_{bs})\,C_{bs}}{R_{\max}^{2}}\right)$$

S-BLACKSWAN

Two Takeaways of Theorem 14

Takeaway 1: how theorem matches with our intuition

$$V^{\pi}_{\widehat{\mathcal{M}}^{\dagger}}(s) \to V^{\pi}_{\mathcal{M}^{\dagger}}(s) \quad \text{a.s.} \quad \text{as} \quad T \to \infty$$

- the value estimation error converges to zero as the agent rolls out longer trajectories.

$$\left|V^{\pi}_{\mathcal{M}^{\dagger}}(s) - V^{\pi}_{\mathcal{M}}(s)\right| = \Omega\left(\frac{\left((R_{\max}-R_{bs})\epsilon_{bs}^{\min}-R_{bs}\epsilon_{bs}\right)(R_{\max}-R_{bs})C_{bs}}{R_{\max}^2}\right)$$

- the value perception gap has a non-zero lower bound, regardless of the horizon length.
- if $u^{-}(x) \to u^{-}_{\star}(x)$ and $w^{-}(x) \to w^{-}_{\star}(x)$, then $R_{bs} \to R_{\max}$ and $\epsilon_{bs} \to 0$, then $\mathcal{B} \to \varnothing$

S-BLACK SWAN

Takeaway 2: three factors influence suboptimality gap

$$|V^{\pi}_{\mathcal{M}^{\dagger}}(s) - V^{\pi}_{\mathcal{M}}(s)| = \Omega\left(\frac{\left((R_{\max}-R_{bs})\epsilon_{bs}^{\min}-R_{bs}\epsilon_{bs}\right)(R_{\max}-R_{bs})C_{bs}}{R_{\max}^2}\right)$$

- Three factors that increase lower bounds
  - **Greater distortion** in reward perception (i.e., larger C_{bs})
  - **Larger feasible set** of S-BLACK SWAN (i.e., larger $(R_{\max} - R_{bs})$)
  - **Higher minimum probability** of S-BLACK SWAN occurrence (i.e., larger $\epsilon_{bs}^{\min}$)

S-BLACK SWAN

Takeaway 2: three factors influence suboptimality gap

$$\left|V^{\pi}_{\mathcal{M}^{\dagger}}(s) - V^{\pi}_{\mathcal{M}}(s)\right| = \Omega\left(\frac{\left((R_{\max}-R_{bs})\epsilon^{\min}_{bs}-R_{bs}\epsilon_{bs}\right)(R_{\max}-R_{bs})C_{bs}}{R^2_{\max}}\right)$$

- Three factors that increase lower bounds
  - **Greater distortion** in reward perception (i.e., larger C_{bs})
  - **Larger feasible set** of S-BLACK SWAN (i.e., larger $(R_{\max} - R_{bs})$)
  - **Higher minimum probability** of S-BLACK SWAN occurrence (i.e., larger $\epsilon^{\min}_{bs}$)

Summary.

Theorem 14 concludes that even with zero estimation error, a lower bound on approximating the true value function remains, and this lower bound increases as above **three factors** become more pronounced.

S-BLACK SWAN

Next natural question: how to *decrease* the lower bound?

S-BLACK SWAN

Next natural question: how to *decrease* the lower bound?

- how can an agent can learn to *self-correct* toward a safe perception, i.e., $u^- \to u^-_\star$ and $w^- \to w^-_\star$.

S-BLACK SWAN

Next natural question: how to *decrease* the lower bound?

- how can an agent can learn to *self-correct* toward a safe perception, i.e., $u^- \to u^-_\star$ and $w^- \to w^-_\star$.
- we answer when to update the perception?
- may refined to: *What is the probability of encountering* S-BLACK SWAN *if the agent takes t steps?*

S-BLACK SWAN

Next natural question: how to *decrease* the lower bound?

- how can an agent can learn to *self-correct* toward a safe perception, i.e., $u^- \to u^-_\star$ and $w^- \to w^-_\star$.
- we answer when to update the perception?
- may refined to: *What is the probability of encountering* S-BLACK SWAN *if the agent takes t steps?*

Theorem 15 (S-BLACK SWAN hitting time)

Assume $\mathbb{P}_{\pi^\star}(s' \mid s) > 0$ for any $s, s' \in \mathcal{S}$, indicating that the one-step state reachability equipped with optimal policy is non-zero, and consider that one step corresponds to a unit time. Then, if the agent **takes *t* steps** *such that $t \geq \log\left(\frac{\delta}{p_{\min}}\right) / \log(1 - p_{\max}) + 1$, where $p_{\min} = \frac{R_{\max} - R_{bs}}{2R_{\max}} \epsilon_{bs}^{\min}$ and $p_{\max} = \frac{R_{\max} - R_{bs}}{2R_{\max}} \epsilon_{bs}$, it will* **encounter** S-BLACK SWAN **with at least probability $\boldsymbol{\delta \in (0, 1]}$**.

S-BLACK SWAN

Takeaway: How often a human should correct their internal perception

- A **large perception gap** $(R_{\max} - R_{bs})$ and **higher minimum probability** of black swan events $(\epsilon_{bs}^{\min})$ require more frequent execution of the self-perception correction algorithm.

Suggestions for safety algorithm design

1. **Emphasize robustness in data, rather than algorithms**
    - Practically, $E \in \mathcal{D}_{\text{real world}}$ and $\mathcal{D}_{\text{real world}} \rightarrow \mathcal{D}_{\text{train}}$, then $E \notin \mathcal{D}_{\text{train}}$.

Suggestions for safety algorithm design

1. **Emphasize robustness in data, rather than algorithms**
   - Practically, $E \in \mathcal{D}_{\text{real world}}$ and $\mathcal{D}_{\text{real world}} \to \mathcal{D}_{\text{train}}$, then $E \notin \mathcal{D}_{\text{train}}$.
   - When generalizing $p_{\text{train}} \to p_{\text{real world}}$, it is advisable to overestimate the likelihood of events considered to have zero probability in p_{train} that could pose high risk.

Suggestions for safety algorithm design

1. **Emphasize robustness in data, rather than algorithms**
   - Practically, $E \in \mathcal{D}_{\text{real world}}$ and $\mathcal{D}_{\text{real world}} \to \mathcal{D}_{\text{train}}$, then $E \notin \mathcal{D}_{\text{train}}$.
   - When generalizing $p_{\text{train}} \to p_{\text{real world}}$, it is advisable to overestimate the likelihood of events considered to have zero probability in p_{train} that could pose high risk.
   - While foundation models offer a strong baseline, it's essential to modify the generative process to focus on potential "zero probability events" that could pose high risks.

2. Make your policy sparse

- **Antifragility**: The ability to gain from small uncertainties to prevent larger, unforeseen uncertainties in the future.

2. Make your policy sparse

- **Antifragility**: The ability to gain from small uncertainties to prevent larger, unforeseen uncertainties in the future.
- How to *benefit* from uncertainty, rather than merely *avoiding* it.

2. Make your policy sparse

- **Antifragility**: The ability to gain from small uncertainties to prevent larger, unforeseen uncertainties in the future.
- How to *benefit* from uncertainty, rather than merely *avoiding* it.
- Enhancing robustness against environmental changes:

$$\min_{\pi} \left| V^{\pi}_{\mathcal{M}_{k+1}} - V^{\pi}_{\mathcal{M}_k} \right|$$

2. Make your policy sparse

- **Antifragility**: The ability to gain from small uncertainties to prevent larger, unforeseen uncertainties in the future.
- How to *benefit* from uncertainty, rather than merely *avoiding* it.
- Enhancing robustness against environmental changes:

$$\min_{\pi} \left| V^{\pi}_{\mathcal{M}_{k+1}} - V^{\pi}_{\mathcal{M}_k} \right|$$

- Benefit from environmental changes :

Definition 16 (Optimization problem: benefits from uncertainty)

We define an optimization problem that benefits the environmental changes for fixed policy as

$$\max_{\pi} \left(V^{\pi}_{\mathcal{M}_{k+1}} - V^{\pi}_{\mathcal{M}_k} \right) \text{ such that } V^{\pi}_{\mathcal{M}_{k+1}} \geq V^{\pi}_{\mathcal{M}_k} \tag{3}$$

Prevent Blackswan by antifragility

Theorem 17 (Short Horizon requires sparse policy)

For $H = 1$, the policy π satisfies (1)-sparse policy, i.e **zero-hot vector**, *is the unique solution.*

Prevent Blackswan by antifragility

Theorem 17 (Short Horizon requires sparse policy)

For $H = 1$, the policy π satisfies (1)-sparse policy, i.e **zero-hot vector**, *is the unique solution.*

Definition 18 (Sparse Policy)

Let the action space be $\mathcal{A} = \{a^{(1)}, a^{(2)}, \ldots, a^{(|\mathcal{A}|)}\}$. A policy π is called *n-sparse* at state s if it assigns positive probability to n actions. Formally, let $\mathcal{I} = \{i \mid \pi(a^{(i)}|s) > 0\}$. Then π is (n)-sparse at s if $|\mathcal{I}| = n$.

Prevent Blackswan by antifragility

Theorem 17 (Short Horizon requires sparse policy)

For $H = 1$, the policy π satisfies (1)-sparse policy, i.e **zero-hot vector**, *is the unique solution.*

Definition 18 (Sparse Policy)

Let the action space be $\mathcal{A} = \{a^{(1)}, a^{(2)}, \dots, a^{(|\mathcal{A}|)}\}$. A policy π is called *n-sparse* at state s if it assigns positive probability to n actions. Formally, let $\mathcal{I} = \{i \mid \pi(a^{(i)}|s) > 0\}$. Then π is (n)-sparse at s if $|\mathcal{I}| = n$.

Theorem 19 (Longer Horizon also requires sparse policy)

For $H \le \mathcal{O}(log(|\mathcal{S}||\mathcal{A}|))$, the policy π satisfies (1)-sparse policy

Why sparse?:

- Probability is not important, the event count is important.

Md Akhtaruzzaman, Sabri Boubaker, and John W Goodell.
Did the collapse of silicon valley bank catalyze financial contagion?
Finance Research Letters, 56:104082, 2023.

Michail Artemenko, Vladimir Budanov, and Nicoly Korenevskiy.
Self-organizing algorithm for pilot modeling the reaction of society to the phenomenon of the black swan.
In *2020 IEEE 14th International Conference on Application of Information and Communication Technologies (AICT)*, pages 1–7. IEEE, 2020.

Alekh Agarwal, Sham Kakade, and Lin F Yang.
Model-based reinforcement learning with a generative model is minimax optimal.
In *Conference on Learning Theory*, pages 67–83. PMLR, 2020.

Tatiana Antipova.
Coronavirus pandemic as black swan event.
In *International conference on integrated science*, pages 356–366. Springer, 2020.

Samit Bhanja and Abhishek Das.

A black swan event-based hybrid model for indian stock markets' trends prediction.
Innovations in Systems and Software Engineering, 20(2):121–135, 2024.

BetterUp.
The availability heuristic.
https://www.betterup.com/blog/the-availability-heuristic, 2022.
Accessed: 2024-05-12.

François Chollet.
On the measure of intelligence.
arXiv preprint arXiv:1911.01547, 2019.

Jinil Persis Devarajan, Arunmozhi Manimuthu, and V Raja Sreedharan.
Healthcare operations and black swan event for covid-19 pandemic: A predictive analytics.
IEEE Transactions on Engineering Management, 70(9):3229–3243, 2021.

Stavros A Drakopoulos and Ioannis Theodossiou.
Workers' risk underestimation and occupational health and safety regulation.
European Journal of Law and Economics, 41:641–656, 2016.

Michael J Fleming and Asani Sarkar.
The failure resolution of lehman brothers.
Economic Policy Review, Forthcoming, 2014.

Hein Fennema and Peter Wakker.
Original and cumulative prospect theory: A discussion of empirical differences.
Journal of Behavioral Decision Making, 10(1):53–64, 1997.

Mohammad Gheshlaghi Azar, Rémi Munos, and Hilbert J Kappen.
Minimax pac bounds on the sample complexity of reinforcement learning with a generative model.
Machine learning, 91:325–349, 2013.

Morgan Housel.
Penguin, 2023.

Xin He, Kaiyong Zhao, and Xiaowen Chu.
Automl: A survey of the state-of-the-art.
Knowledge-based systems, 212:106622, 2021.

Michael Janner, Justin Fu, Marvin Zhang, and Sergey Levine.
When to trust your model: Model-based policy optimization.
Advances in neural information processing systems, 32, 2019.

Ming Jin.
Preparing for black swans: The antifragility imperative for machine learning.
arXiv preprint arXiv:2405.11397, 2024.

Sham Machandranath Kakade.
On the sample complexity of reinforcement learning.
University of London, University College London (United Kingdom), 2003.

Andrei Kirilenko, Albert S Kyle, Mehrdad Samadi, and Tugkan Tuzun.
The flash crash: High-frequency trading in an electronic market.
The Journal of Finance, 72(3):967–998, 2017.

Daniel Kahneman and Amos Tversky.
Prospect theory: An analysis of decision under risk.

In *Handbook of the fundamentals of financial decision making: Part I*, pages 99–127. World Scientific, 2013.

Bo Li, Peng Qi, Bo Liu, Shuai Di, Jingen Liu, Jiquan Pei, Jinfeng Yi, and Bowen Zhou.
Trustworthy ai: From principles to practices.
ACM Computing Surveys, 55(9):1–46, 2023.

Tor Lattimore and Csaba Szepesvári.
Bandit algorithms.
Cambridge University Press, 2020.

John Kwaku Mensah Mawutor.
The failure of lehman brothers: causes, preventive measures and recommendations.
Research Journal of Finance and Accounting, 5(4), 2014.

Larry McDonald and Patrick Robinson.
A colossal failure of common sense: The incredible inside story of the collapse of Lehman Brothers.
Random House, 2009.

Bhavana Pandit, Alex Albert, Yashwardhan Patil, and Ahmed Jalil Al-Bayati.
Impact of safety climate on hazard recognition and safety risk perception.
Safety science, 113:44–53, 2019.

Matt Phillips.
Gamestop's wild stock ride: How ai and social media drove a short squeeze.
The New York Times Business, 2021.
Accessed: 2024-08-19.

SD Prestwich.
Tuning forecasting algorithms for black swans.
IFAC-PapersOnLine, 52(13):1496–1501, 2019.

Matihew Rabin.
Risk aversion and expected-utility theory: A calibration theorem.
In *Handbook of the fundamentals of financial decision making: Part I*, pages 241–252. World Scientific, 2013.

Paul Rogers.
The cognitive psychology of lottery gambling: A theoretical review.

Journal of gambling studies, 14(2):111–134, 1998.

Samuel Henrique Silva and Peyman Najafirad.
Opportunities and challenges in deep learning adversarial robustness: A survey.
arXiv preprint arXiv:2007.00753, 2020.

Philip Stafford.
Citadel securities trading algorithm triggers market volatility.
Financial Times Online, 2022.
Accessed: 2024-08-19.

Aaron Sidford, Mengdi Wang, Xian Wu, Lin F Yang, and Yinyu Ye.
Near-optimal time and sample complexities for solving discounted markov decision process with a generative model.
arXiv preprint arXiv:1806.01492, 2018.

Nassim Nicholas Taleb.
The Black Swan:: The Impact of the Highly Improbable: With a new section:" On Robustness and Fragility", volume 2.
Random house trade paperbacks, 2010.

Toni GLA van der Meer, Anne C Kroon, and Rens Vliegenthart.
Do news media kill? how a biased news reality can overshadow real societal risks, the case of aviation and road traffic accidents.
Social forces, 101(1):506–530, 2022.

Morgenstern von Neumann.
Theory of games and economic behaviour, 1944.

Peter Vasterman, C Joris Yzermans, and Anja JE Dirkzwager.
The role of the media and media hypes in the aftermath of disasters.
Epidemiologic reviews, 27(1):107–114, 2005.

Maxime Wabartha, Audrey Durand, Vincent Francois-Lavet, and Joelle Pineau.
Handling black swan events in deep learning with diversely extrapolated neural networks.
In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pages 2140–2147, 2021.

Guangyu Wang, Xiaohong Liu, Zhen Ying, Guoxing Yang, Zhiwei Chen, Zhiwen Liu, Min Zhang, Hongmei Yan, Yuxing Lu, Yuanxu Gao, et al.
Optimized glycemic control of type 2 diabetes with reinforcement learning: a proof-of-concept trial.
Nature Medicine, 29(10):2633–2642, 2023.

Rosalind Wiggins, Thomas Piontek, and Andrew Metrick.
The lehman brothers bankruptcy a: overview.
Yale program on financial stability case study, 2014.

Anders AF Wahlberg and Lennart Sjoberg.
Risk perception and the media.
Journal of risk research, 3(1):31–50, 2000.

Gregory Wheeler and G Wheeler.
A review of the lottery paradox.
Probability and inference: Essays in honour of Henry E. Kyburg, Jr, pages 1–31, 2007.

Jingkang Yang, Kaiyang Zhou, Yixuan Li, and Ziwei Liu.

Generalized out-of-distribution detection: A survey.
International Journal of Computer Vision, pages 1–28, 2024.