Risk-perception-aware control design under dynamic spatial risks

Aamodh Suresh and Sonia Martínez

Abstract—This work proposes a novel risk-perception-aware (RPA) control design using non-rational perception of risks associated with uncertain spatial costs. We use Cumulative Prospect Theory (CPT) to model the risk perception of a decision maker (DM) and use it to construct perceived risk functions that transform the uncertain dynamic spatial cost to deterministic perceived risks of a DM. These risks are then used to build safety sets which can represent risk-averse to risk-insensitive perception. We define a notions of “inclusiveness” and “versatility” based on safety sets and use it to compare with other models such as Conditional value at Risk (CVaR) and Expected risk (ER). We theoretically prove that CPT is the most “inclusive” and “versatile” model of the lot in the context of risk-perception-aware controls. We further use the perceived risk function along with ideas from control barrier functions (CBF) to construct a class of perceived risk CBFs. For a class of truncated-Gaussian costs, we find sufficient geometric conditions for the validity of this class of CBFs, thus guaranteeing safety. Then, we generate perceived-safety-critical controls using a Quadratic program (QP) to guide an agent safely according to a given perceived risk model. We present simulations in a 2D environment to illustrate the performance of the proposed controller.

I. INTRODUCTION

Motivation: Safety is a desirable and necessary design constraint for any control system; specially when operated in a shared environment with a decision maker (DM). Arguably, most environments have associated spatial risks, whose source can vary from hard constraints (e.g. moving obstacles) to softer constraints (e.g. wind conditions). Different DMs can perceive these risks differently, leading to notions of perceived risks and perceived safety from these risks.

It is well known from psychophysics [1] and behavioral economics [2] research that humans as DMs have fundamental non-linear perception leading to non-rational decision making in risky situations. In such cases, existing methods assuming perfect knowledge or rational and coherent treatment (as in expected risk and Conditional Value at Risk (CVaR)) of risks may not suffice, which can lead to loss of trust or discomfort among DMs. This motivates the need of richer and more inclusive modeling of risk perception to capture a variety of DMs and use them for safe control design. This work aims to bridge the gap between behavioral decision making and safety using Cumulative Prospect Theory (CPT) as a risk perception model, and Control Barrier Functions (CBFs) for safe control design.

Related Work: Safe control system design has been tackled using various frameworks such as artificial potential functions [3], barrier certificates [4] and, more recently, control barrier functions (CBFs) [5]. CBFs have gained popularity due to their Lyapunov-like properties, rigorous safety guarantees and ease of application. They have been successfully used in optimization [5], stabilization [6] and data-driven control frameworks [7]. CBFs were traditionally used in static scenarios, more recently, they have been used to deal with moving obstacles [8] and multi-agent systems [9].

Uncertainty has been mainly handled using robustness measures [10], stochastic control [11], or chance constraints [12]. Very few works have considered the notion of risk perception explicitly in a control system [13], [14]. All these works use CVaR to quantify risk perception, which only captures linear and rational risk-averse behavior. CPT on the other hand is a more expressive (see [15]), non-linear and non-rational perception theory which is yet to be applied in the context of safety for a control system. Moreover, CPT has been successfully used in engineering applications like path planning [15], traffic routing [16], and network protection [17].

Contributions: We first adapt the notion of non-rational risk perception to the context of safety for control systems. With this, we capture a larger spectrum of DM’s risk profile, extending the existing literature. We support this claim theoretically by defining the notion of “inclusiveness” and proving that CPT is the most inclusive risk perception model out of the other popular models: CVaR and ER. We then use the CPT value function to construct a class of CBFs to guarantee safety according to a DM’s perceived risk and define the notion of perceived safety. Additionally, we find sufficient geometric conditions on the control input to maintain the validity of our proposed RPA CBF and compare them among the three risk perception models (RPMs). Then, we design a QP-based RPA controller to guide an agent to a desired goal safely w.r.t. perceived risks. Thus we extend the literature with more inclusive safe control design. Practically, we consider 2D simulations with moving obstacles and show the effectiveness of the proposed RPA controller along with the practical translation of the inclusiveness hierarchy.

This work provides a framework to incorporate and compare a wide range of RPMs to generate a variety of RPA controls. We would also like to clarify that the validation of CPT models using user studies for typical control scenarios is beyond the scope of this work.
II. RISK PERCEPTION FORMALISM AND PROBLEM SETUP

Here, we introduce some notation and a formal notion of risk perception, starting with a concise description of CPT and CVaR (see [18] and [19] for more details). Later, we describe our problem statement.

Risk Perception: By risk perception, we refer to the notion of attaching a value (risk) to a random cost output. Formally, let \( S \) be a discrete sample space endowed with a probability distribution \( \mathbb{P} \). We model environmental cost via a real-valued, discrete random variable \( c : S \to \mathbb{R}_{\geq 0}, \) taking \( M \) possible values, \( c_i \in \mathbb{R}_{\geq 0}, \) \( i \in \{1, \ldots, M\} \), and such that \( p_i = \mathbb{P}(c = c_i) \), with \( \sum_i p_i = 1 \). Let \( \mathcal{C} \) be the set of such random cost variables and \( R : \mathcal{C} \to \mathbb{R}_{\geq 0} \) a value function which associates a value (risk) to a random cost variable.

A value function \( R \) can be defined in many ways, resulting in different risk perceptions. Here, an RPM is characterized as a parameterized family \( \mathcal{M} = \{R_\theta \mid \theta \in \mathbb{R}^k\} \) of value functions. In what follows, we consider three popular RPMs: Expected Risk (ER) Conditional Value at Risk (CVaR) [19] and Cumulative Prospect Theory (CPT) [2].

CPT captures non-rational decision making, and was introduced in [18], [20]. In CPT, outcomes are first weighed using a non-linear utility function \( w \) introduced in [18], [20]. In CPT, outcomes are first weighed using a non-linear utility function \( w \) introduced in [18], [20]. In CPT, outcomes are first weighed using a non-linear utility function \( w \) introduced in [18], [20]. In CPT, outcomes are first weighed using a non-linear utility function \( w \) introduced in [18], [20]. In CPT, outcomes are first weighed using a non-linear utility function \( w \) introduced in [18], [20]. In CPT, outcomes are first weighed using a non-linear utility function \( w \) introduced in [18], [20].

In what follows, we consider three popular RPMs: Expected Risk (ER) Conditional Value at Risk (CVaR) [19] and Cumulative Prospect Theory (CPT) [2].

III. PERCEIVED SAFETY USING VARIOUS RPMs

This section compares various RPMs, solving Problem 1 given an uncertain field \( c : \mathcal{X} \to \mathbb{R}_{\geq 0} \), that aims to quantify objectively the (negative) consequences of being at \( x \in \mathcal{X} \) relative to a known risk source at \( y \in \mathcal{X} \). More precisely, \( c(x) \) is a discrete RV which can take \( M \) possible values, \( c_i(x) \in \mathbb{R}_{\geq 0}, \) for \( i \in \{1, \ldots, M\} \). We assume that \( c \) has associated mean and standard deviation functions \( c_\mu : \mathcal{X} \to \mathbb{R}_{\geq 0} \) and \( c_\sigma : \mathcal{X} \to \mathbb{R}_{\geq 0} \), respectively. We assume that \( c_\mu, c_\sigma \) are continuously differentiable in their domains. Given \( c \), an associated spatial-risk function is given by \( R_c : \mathcal{X} \to \mathbb{R}_{\geq 0} \), \( R_c(x) \equiv R(c(x)) \), where \( R \) belongs to any of the previous RPMs defined in (1) above. When clear from the context, we will identify \( R_c \equiv R \in \mathcal{M} \). The larger \( R_c \) is at \( x \), the higher the perceived risk of being at \( x \in \mathcal{X} \).

Dynamic systems: We aim to control an agent modeled as a control-affine dynamic system:

\[
\dot{x} = f_s(x, u) = f(x) + G(x)u, \quad (2)
\]

where \( u \in \mathbb{R}^m, \) \( G : \mathcal{X} \to \mathbb{R}^{n \times m} \), \( f : \mathcal{X} \to \mathbb{R}^n \) and \( f \) and \( G \) are locally Lipschitz. We also consider a dynamic risk

\[ \dot{y} = f_y(y), \quad y \in \mathcal{X}, \quad f_y : \mathcal{X} \to \mathbb{R}^n, \]

with a locally Lipschitz \( f_y \). We focus on moving obstacles as the source of risk, but the approach can be extended to other scenarios. We also assume that an asymptotically stable controller \( k : \mathcal{X} \to \mathbb{R}^m \) has been designed to guide the agent to a goal state \( x^* \in \mathcal{X} \) in the absence of risk sources. We wish to drive the agent to a goal \( x^* \in \mathcal{X} \) safely, while avoiding risky areas. Formally, we define safety considering a perceived spatial risk function \( R_c \) as follows:

Definition 1: (Perceived Safety) An agent moving under (2), and subject to an uncertain cost source \( c \) with dynamics (3), is said to be safe w.r.t. the perceived risk \( R_c \) iff \( R_c(x(t)) \leq \rho, \) \( \forall t \geq 0, \) for some tolerance \( \rho \in \mathbb{R}_{\geq 0} \). We now state the problems we address in this work:

Problem 1: (RPA safe sets) Given a risky environment \( \mathcal{X} \), endowed with an uncertain cost source \( c \), design perceived safety sets considering RPMs from (1). Characterize and constrain the properties of these sets among the three RPMs.

Problem 2: (RPA safe controls) Under previous conditions, design a controller \( u \), nominally deviating from a stable state feedback controller \( k \), such that the agent reaches the goal \( x^* \) safely (Definition 1) and examine feasibility of \( u \).
Fix a model $\mathcal{M}$ and a risk source at $y \in \mathcal{X}$. The total safe set of $\mathcal{M}$ wrt $y$ is given as $\mathcal{Y}_\theta(y, c) \triangleq \bigcup_{b \in \mathcal{M}} \text{Safe}(R, y)$ (resp. the total risky set of $\mathcal{M}$ wrt $y \in \mathcal{X}$ is $\mathcal{Y}_\theta(y, c) \triangleq \bigcup_{b \in \mathcal{M}} \text{Risk}(R, y)$). Thus, given $y \in \mathcal{X}$, the set $\mathcal{Y}_\theta(y, c)$ (resp. $\mathcal{Y}_\theta(y, c)$) covers all the states in $\mathcal{X}$ that are safe (resp. unsafe) according to a RPM $\mathcal{M}$.

Definition 2: (Inclusiveness and Strict Inclusiveness).
Consider two RPMs $\mathcal{M}_1$ and $\mathcal{M}_2$, a threshold $\rho \in \mathbb{R}_{\geq 0}$, and a risk source at $y \in \mathcal{X}$. Let the sets $\mathcal{Y}_1(y, c), \mathcal{Y}_1(y, c)$ and $\mathcal{Y}_2(y, c), \mathcal{Y}_2(y, c)$ be the total safe and risky sets of $\mathcal{M}_1$ and $\mathcal{M}_2$ wrt $y$ and a spatial cost $c$, respectively. We say that $\mathcal{M}_1$ is more inclusive than $\mathcal{M}_2$ $(\mathcal{M}_1 \triangleright \mathcal{M}_2)$ if either $\mathcal{Y}_2(y, c) \subseteq \mathcal{Y}_1(y, c)$ and $\mathcal{Y}_2(y, c) \subseteq \mathcal{Y}_1(y, c)$ holds, or $\mathcal{Y}_1(y, c) 
subseteq \mathcal{Y}_1(y, c)$ and $\mathcal{Y}_2(y, c) \subseteq \mathcal{Y}_1(y, c)$ holds, for all $y \in \mathcal{X}$ and costs $c : \mathcal{X} \to \mathbb{R}_{\geq 0}$. If $\mathcal{Y}_2(y, c) \subseteq \mathcal{Y}_1(y, c)$ and $\mathcal{Y}_2(y, c) \subseteq \mathcal{Y}_1(y, c)$ both hold, then $\mathcal{M}_1$ is strictly more inclusive than $\mathcal{M}_2$ $(\mathcal{M}_1 \triangleright \mathcal{M}_2)$.

In particular, if $\mathcal{M}_1 \triangleright \mathcal{M}_2$, then $\mathcal{M}_1$ results into a wider range of safety and risky sets for a given environment than $\mathcal{M}_2$.

Now we compare the inclusiveness of CPT, CVaR and ER via their respective value functions. We start by comparing the range space of these RPMs.

Lemma 1: Consider a threshold $\rho \in \mathbb{R}_{\geq 0}$, a risk source at $y \in \mathcal{X}$, and two RPMs $\mathcal{M}_1$ and $\mathcal{M}_2$ with range spaces $\mathcal{R}_1, \mathcal{R}_2$, respectively. If $\mathcal{Y}_2(y, c) \subseteq \mathcal{Y}_1(y, c)$, and if there exists an $R_{1,c} \in \mathcal{M}_1$ such that $R_{1,c} > R_{2,c}$ or $R_{1,c} < R_{2,c}$ for any $R_{2,c} \in \mathcal{M}_2$, and any $c$, $\mathcal{M}_1 \triangleright \mathcal{M}_2$. In addition, if $\mathcal{Y}_1(y, c) \subseteq \mathcal{Y}_1(y, c)$ and $\mathcal{Y}_2(y, c) \subseteq \mathcal{Y}_1(y, c)$, for all $y \in \mathcal{X}$ and costs $c : \mathcal{X} \to \mathbb{R}_{\geq 0}$, then $\mathcal{M}_1 \triangleright \mathcal{M}_2$.

Proof: Fix $c$. Since $\mathcal{Y}_2 \subseteq \mathcal{Y}_1$, $\forall R_2 \in \mathcal{M}_2$, there is $R_1 \in \mathcal{M}_1$ s.t. $R_1(c(y-x)) = R_2(c(y-x)), \forall x \in \mathcal{X}$. Thus, $\mathcal{Y}_2(y, c) \subseteq \mathcal{Y}_1(y, c)$ and $\mathcal{Y}_2(y, c) \subseteq \mathcal{Y}_1(y, c)$. Assume $\exists R_1, R_2 \in \mathcal{M}_1$ s.t. $R_1(c(y-x)) > R_2(c(y-x))$ or $R_1(c(y-x)) < R_2(c(y-x))$ hold for all $R_2 \in \mathcal{M}_2$. This implies either $\mathcal{Y}_2(y, c) \subseteq \mathcal{Y}_1(y, c)$ and $\mathcal{Y}_2(y, c) \subseteq \mathcal{Y}_1(y, c)$ or $\mathcal{Y}_1(y, c) \subseteq \mathcal{Y}_1(y, c)$ and $\mathcal{Y}_2(y, c) \subseteq \mathcal{Y}_1(y, c)$. Inclusiveness follows from Definition 2.

Lemma 2: Consider the CPT, CVaR and ER risk models, with associated range sets $\mathcal{R}_{\text{CPT}}(c), \mathcal{R}_{\text{CVaR}}(c)$, and $\mathcal{R}_{\text{ER}}(c)$. Then, it holds that $\mathcal{R}_{\text{CPT}}(c) \supseteq \mathcal{R}_{\text{CVaR}}(c) \supseteq \mathcal{R}_{\text{ER}}(c), \forall c$.

Proof: Fix $c$. Note that $\mathcal{R}_{\text{CPT}}(c) = \{\mu\}$. By choosing $\mathcal{R}_{\text{CPT}}(c) \in \text{CPT}$ with $\tilde{\theta} = \{1, 1, 1\}$ and $\mathcal{R}_{\text{CVaR}}(c) \in \text{CVaR}$ we have $\mathcal{R}_{\text{CPT}}(c) = \mathcal{R}_{\text{CVaR}}(c) = \{\mu\}$, $\forall c$. Note that only if $c_{\mu} = 0$ then $\mathcal{R}_{\text{CVaR}}(c) = \{\mu\} = \mathcal{R}(c)$ for all $q$. When $c_{\mu} \neq 0$, with any other valid choice of parameters in CPT we obtain $\mathcal{R}_{\text{CVaR}}(c) \neq \mathcal{R}_{\text{ER}}(c)$. We can find $\theta \neq \tilde{\theta}$ such that $\mathcal{R}_{\text{CPT}}(c) \subseteq \mathcal{R}_{\text{ER}}(c), \forall c$. Hence, $\mathcal{R}_{\text{ER}}(c) \subseteq \mathcal{R}_{\text{CVaR}}(c)$ and $\mathcal{R}_{\text{CPT}}(c) \subseteq \mathcal{R}_{\text{ER}}(c)$.

For CVaR, $\mathcal{R}_{\text{CPT}}(c) = \{\mu\}$ and $\mathcal{R}_{\text{CVaR}}(c) \subseteq \{\mu\}$, where $b \in \mathbb{R}$ is the worst-case outcome of $c$. Since $\mathcal{R}_{\text{CPT}}(c) > \mathcal{R}_{\text{CVaR}}(c)$ increases in $q$, $\mathcal{R}_{\text{CVaR}}(c) \subseteq \{\mu\}$, $b$. Choosing $\theta = \{1, 1, 1\}$, for $\lambda \geq 1$, leads to $\mathcal{R}_{\text{CPT}}(c) = \lambda \sum c_i p_i = \lambda c_{\mu}$. Taking $\lambda \in [1, b]$, with $b > \frac{1}{c_{\mu}}$, we get $\mathcal{R}_{\text{CPT}}(c) \supseteq \{\mu\}$; hence, $\mathcal{R}_{\text{CPT}}(c) \supseteq \mathcal{R}_{\text{CVaR}}(c)$.

The previous results now lead to the following.

Theorem 1: Let $c$ be a discrete random field cost. Consider the ER, CVaR and CPT risk perception models with

When clear from the context, we will just denote $\mathcal{Y}_\theta(y, c) \equiv \mathcal{S}_\theta(y, c)$.
theorem \( \exists \lambda \) s.t. \( R^\text{P} \in [c_\mu, c_{\text{max}}] \), and a \( \gamma \) s.t. \( R^\text{P} \in [c_{\text{min}}, c_\mu] \). Hence, from Lemma 3, CPT is most versatile in \( \mathcal{Y} \). \( \blacksquare \)

**Lemma 4:** Under the assumptions of Lemma 3 with \( T_1 = [c_\mu, c_{\text{max}}] \) and \( T_2 = [c_\mu, c_\mu] \), CVA is \( T_1 \)-versatile and ER is \( T_2 \)-versatile. Hence neither are most versatile RPMs.

**Proof:** This result trivially follows from the range spaces \( \mathcal{R}_{\text{CVA}}(c) \) and \( \mathcal{R}_{\text{ER}}(c) \) in the proof of Lemma 2. \( \blacksquare \)

### IV. Control Design with Risk-Perception-Aware-CBFs

Here, we address Problem 2 and design controls \( u \) for an agent subject to (2), to ensure perceived safety (Definition 1).

**Definition 4 (RPA-CBF):** Consider an agent subject to (2), a dynamic source of risk (3), and a perceived risk \( R \) model. A \( \mathcal{C} \) function \( h_R \triangleq h \circ R_\gamma : \mathcal{X} \to \mathbb{R} \) is an RPA-CBF for this system, if there is an extended class \( \mathcal{X}_\omega \) function \( \eta \) such that the control set \( K_R \) defined as

\[
K_R(R_c) = \{ u \in \mathcal{U} | h_R^*(\xi) \geq -\eta_1(h_R(\xi)) \}
\]

is non-empty for all \( \xi \in \mathcal{X} \).

The existence of \( h_R \) according to Definition 4 implies that the superlevel set \( \{ x \in \mathcal{X} | h_R(\xi) > 0 \} \) is forward invariant under (2). We specify \( h_R = h \circ R_c \) via \( h \) given as

\[
h(\xi) \triangleq \eta_2(\rho - R_c(\xi)),
\]

where \( \eta_2 : \mathbb{R} \to \mathbb{R} \) is a \( \mathcal{C} \) extended class \( \mathcal{X}_\omega \) function. Since \( \eta_2 \) is non-decreasing, \( h(R_c(\xi)) \geq 0 \) implies \( R_c(\xi) \leq \rho \) and from (4), \( \mathcal{X}_\omega \) safe \( R_c \) is indicated \( \eta \). Thus, \( h(R_c(\xi)) > 0 \) indicates that \( x \) is perceived as safe w.r.t. \( R_c \).

The RPA control input \( u \) can now be computed via:

\[
u(x) = \arg \min_u [u - k(x)]^2 \tag{7a}
\]

\[
d\eta_2 \left( \frac{\partial R_c}{\partial \xi} \right) \cdot (f_c(y) - f_c(x)) \geq -\eta_1(h_R(\xi)) \tag{7b}
\]

The above problem captures the notion of minimally modifying a stable controller to ensure safety of the system. Next we will analyze the feasibility conditions for the proposed controller \( u \) and compare it across the proposed models.

**Feasibility analysis and comparison:** We first describe a construction of finite outcomes of \( c \) from \( c_\mu \) and \( c_\sigma \) called “truncated-Gaussian cost” which will be used for analysis. Assume that \( c(\xi) \) is distributed as a truncated Gaussian \( \mathcal{N} \left( \mu, \sigma^2 \right) \) around \( \mu(\xi) \) and \( \sigma^2(\xi) \). Then, given \( M \in \mathbb{N} \), we approximate \( c \) by means of \( M \) discrete values \( c_i \), \( i \in \{1, \ldots, M\} \), with probability calculated from the CDF \( F_c \) of \( c \) at each \( c_i \). That is, \( p_i = F(c_i) \), and \( p_i = F(c_i) - F(c_{i-1}) \), \( i \in \{2, \ldots, M\} \). Now, we show conditions on \( u \) for the set \( K_{\text{CBF}} \) to be non-empty for a given risk function \( R_c \). We define a few constants and variables to help us compare the feasibility conditions of the three RPMs. Let \( \phi_j \in [-\pi, \pi] \) be the relative angle \( 4 \) between \( \frac{\partial R_c}{\partial \xi} \) and \( \xi(u(x), y) \), and \( \phi_j = \frac{\partial u}{\partial c} \) and \( \phi_j = \frac{\partial u}{\partial c} \). Now define \( k^* (\xi) = (\eta_1 \circ R^\text{ER}(\xi))/\frac{d\eta_1}{dR^c} \) and \( k^* (\xi) = (\eta_1 \circ R^\text{SV}(\xi))/\frac{d\eta_1}{dR^c} \) and \( k^* (\xi) = (\eta_1 \circ R^\text{CV}(\xi))/\frac{d\eta_1}{dR^c} \) and \( k^* (\xi) = (\eta_1 \circ R^\text{CV}(\xi))/\frac{d\eta_1}{dR^c} \). Also we define constants \( k_0^*(\xi) = (\eta_1 \circ R^\text{ER}(\xi))/\frac{d\eta_1}{dR^c} \) and \( k_0^*(\xi) = (\eta_1 \circ R^\text{SV}(\xi))/\frac{d\eta_1}{dR^c} \) and \( k_0^*(\xi) = (\eta_1 \circ R^\text{CV}(\xi))/\frac{d\eta_1}{dR^c} \) and \( k_0^*(\xi) = (\eta_1 \circ R^\text{CV}(\xi))/\frac{d\eta_1}{dR^c} \).

\[
\eta_1 (h_R(\xi)) \geq -\eta_1 (h_R(\xi)), \tag{8}
\]

then \( h_R \) defined according to (9) is a valid RPA-CBF for any \( \eta_2 \) and (7) is feasible. Specifically, with \( \xi = (\xi(u(x), y)) \) the RHS of the above inequality reduces to \(-\eta^*, -\eta^*, \) and \(-\eta^* \) for ER, CVA, and CPT, respectively.

**Proof:** For first part, rearranging terms in (8) we get:

\[
l_k \left( \left| \frac{\partial R_c}{\partial \xi} \right| \cdot \left| f_c(\xi) - f_c(x) \right| \cos(\phi_j) \right) \geq 0, \tag{9}
\]

Then \( h_R \) defined according to (9) is a valid RPA-CBF for any \( \eta_2 \) and (7) is feasible. Specifically, with \( \xi = (\xi(u(x), y)) \) the RHS of the above inequality reduces to \(-\eta^*, -\eta^*, \) and \(-\eta^* \) for ER, CVA, and CPT, respectively.

**Proof:** For first part, rearranging terms in (8) we get:

\[
\eta (h_R(\xi)) \geq -\eta_1 (h_R(\xi)) \tag{10}
\]

Using the chain rule, we get the time derivative of \( h_R(\xi) \):

\[

\frac{d\eta_1}{dR^c} \left( \frac{\partial R_c}{\partial \xi} \right) \cdot \left| f_c(\xi) - f_c(x) \right| \cos(\phi_j) \right) \geq 0, \tag{11}
\]

For the last part, the expressions are obtained by substituting the respective risk functions and evaluating the partial derivatives \( \frac{\partial R_c}{\partial \xi} \) and \( \frac{\partial R_c}{\partial \xi} \) (part of \( \frac{\partial R_c}{\partial \xi} \)). Thus we need to show the following hold true:

\[
\xi \geq \left( \frac{k^* (\xi)}{\left| \frac{\partial \xi}{\partial \xi} \right|} \right) \tag{12a}
\]

\[
\xi \geq \left( \frac{k_0^* (\xi)}{\left| \frac{\partial \xi}{\partial \xi} \right|} \right) \tag{12b}
\]

\[
\xi \geq \left( \frac{k_0^* (\xi)}{\left| \frac{\partial \xi}{\partial \xi} \right|} \right) \tag{12c}
\]

For ER we get \( \frac{\partial R_c}{\partial \xi} = 1 \) and \( \frac{\partial R_c}{\partial \xi} = 0 \). For CVA, since \( c \) is assumed to belong to a truncated Gaussian distribution,
we can use the closed form expression of CVaR \((13)\) for a Gaussian distribution to calculate the partials \(\frac{\partial R_q^p}{\partial c}\), and \(\frac{\partial R_q^p}{\partial c}\).

\[
R_q^p = c_{\mu} + c_{\sigma} \left( \frac{F^{-1}(q)}{q} \right).
\]

(13)

From \((13)\), it is easy to see that CVaR is linear in \(c_{\mu}\) and \(c_{\sigma}\). With this, we get \(\frac{\partial R}{\partial c_{\mu}} = 1\) and \(\frac{\partial R}{\partial c_{\sigma}} = \frac{F^{-1}(q)}{q}\). Substituting these derivatives in \((11)\) correspondingly we obtain the results.

For CPT, the expression is obtained by substituting the CPT risk function \(R_q^{\text{CPT}}\) and evaluating the partial derivatives \(\frac{\partial R_q^{\text{CPT}}}{\partial c_{\mu}}\) and \(\frac{\partial R_q^{\text{CPT}}}{\partial c_{\sigma}}\). Constructing truncated Gaussian costs \(c\) from \(c_{\mu}\) and \(c_{\sigma}\), we get outcomes \(\{c_1, \ldots, c_M\}\) and corresponding probabilities \(\{p_1, \ldots, p_M\}\) resulting in constant \(\Pi\) throughout. In this way, from \((1c)\), the CPT value of a random cost \(c\) with mean \(c_{\mu}\) and \(c_{\sigma}\) is given by:

\[
R_q^{\text{CPT}}(c_{\mu}, c_{\sigma}) = \frac{M}{\gamma} \left( (c_{\mu} + c_{\sigma} - 3M) \right) \Pi,
\]

(14)

With this expression, we can proceed to calculate the partial derivatives \(\frac{\partial R_q^{\text{CPT}}}{\partial c_{\mu}}\) and \(\frac{\partial R_q^{\text{CPT}}}{\partial c_{\sigma}}\). From \((14)\), we get

\[
\frac{\partial R_q^{\text{CPT}}}{\partial c_{\mu}}(c_{\mu}, c_{\sigma}) = \lambda \sum_{i=1}^{M} \left( (c_{\mu} + c_{\sigma} - 3M) \right) \Pi,
\]

(15a)

\[
\frac{\partial R_q^{\text{CPT}}}{\partial c_{\sigma}}(c_{\mu}, c_{\sigma}) = \lambda \sum_{i=1}^{M} \left( (c_{\mu} + c_{\sigma} - 3M) \right) \Pi.
\]

(15b)

We have \(\frac{\partial R_q^{\text{CPT}}}{\partial c_{\mu}}(c_{\mu}, c_{\sigma}) = k_{\mu}^c\) and \(\frac{\partial R_q^{\text{CPT}}}{\partial c_{\sigma}}(c_{\mu}, c_{\sigma}) = k_{\sigma}^c\). Substituting \(k_{\mu}^c\), \(k_{\mu}^c\) and \(k_{\sigma}^c\) in \((8)\), we obtain \((12b)\).

From \((8)\), observe that the RHS is independent of \(u\) and the LHS is independent of \(R\) and the RPM. This separation makes it easier to compare various RPMs and their associated feasibility conditions.

Next, we remark on the uncertainty perception of each RPM, which will be used in the subsequent proposition to compare the size of control sets \(K_{\text{CBF}}\) respectively generated by each of the RPMs.

Remark 1 (Uncertainty perception among RPMs): The ER model is insensitive to uncertainty as \(\frac{\partial R_{ER}^p}{\partial c_{\sigma}} = 0\). In this way, CVar is averse to uncertainty as \(\frac{\partial R_{\text{CVar}}^p}{\partial c_{\sigma}} \geq 0\) for all \(q\). With CPT, \(\theta\) can be tuned to get both uncertainty insensitive and uncertainty averse behavior, additionally, it can also produce uncertainty liking behavior (when \(\frac{\partial R_{\text{CPT}}^p}{\partial c_{\sigma}} \leq 0\)).

We finally compare the the flexiblity provided by each model via the corresponding control sets \(K\).

**Proposition 3:** Assume the conditions of Proposition 2 hold. Then, the feasibility sets defined according to \((5)\) for the three RPMs satisfy \(K_{\text{ER}} \subseteq K_{\text{CPT}}\) and \(K_{\text{CVar}} \subseteq K_{\text{CPT}}\).

**Proof:** In order to compare the feasibility of the sets \(K_{\text{CBF}}\) from \((5)\) for the three RPMs, we can compare their respective feasibility conditions \((8)\). Consider \(\eta^\varepsilon = \frac{\xi^\varepsilon}{\left| \xi^\varepsilon \right|^2}\) and \(\eta^\varepsilon = \frac{k_\mu^c(\xi)}{|k_\mu^c(\xi) + k_\sigma^c(\xi)|}\). Since the LHS in \((8)\) remains the same for any RPM and its parameter choice, to prove the proposition, it is sufficient to show that \(\eta^\varepsilon \leq \eta^\varepsilon\) and \(\eta^\varepsilon \leq \eta^\varepsilon\). These inequalities follow from the choice of \(\theta = 1\) in Theorem 1 and CPT’s more adaptable uncertainty perception from Remark 1.

It is interesting to note that although CVar is more inclusive than ER as proved in Theorem 1, it does not immediately translate into CVar having a larger control feasibility set. We provide more insight in the following remark.

**Remark 2:** Consider the control feasibility sets \(K_{\text{ER}}\) and \(K_{\text{CVar}}\) respectively for ER and CVar, defined according to \((5)\). Then, depending on the choice of \(\theta\) and construction of \(c_{\sigma}\), we can obtain either \(K_{\text{ER}} \subseteq K_{\text{CVar}}\) or \(K_{\text{CVar}} \subseteq K_{\text{ER}}\). Looking at the RHS of inequalities \((12b)\) and \((12a)\), although we have \(k_\mu^c(\xi) > k_\mu^c(\xi)\) from Theorem 1 there isn’t conclusive proof to suggest \(K_{\text{ER}} \subseteq K_{\text{CVar}}\) due to the additional \(k_\sigma^c(\xi)\) term in the denominator of \((12b)\).

Stability analysis: Next, let us look at the stability properties of the proposed controller \(u\) in \((7)\). It is clear that if the nominal controller \(k(x)\) also satisfies the safety constraint \((7b)\), then \(u = k(x)\) and the stability properties of \(k(x)\) transfer over to \(u\). To analyze stability, first we look into the RPMs and determine how they affect the deviation from \(k(x)\). Later, we treat the controller \(u\) as a perturbed version of \(k(x)\) and analyze accordingly.

Let \(\delta = k(x) - u(x)\) be the perturbation to the nominal controller \(k(x)\) and \(\delta_{ER}^\varepsilon\), \(\delta_{CVar}^\varepsilon\) and \(\delta_{CPT}^\varepsilon\) be the respective perturbations of ER, CVar and CPT with corresponding parameter choices. Then we have the following:

**Proposition 4:** Under the assumptions of Proposition 2 choose \(u\) as in \((7)\). Assume \(\|\delta_{ER}^\varepsilon\|\), \(\|\delta_{CVar}^\varepsilon\|\) and \(\|\delta_{CPT}^\varepsilon\|\) are bounded. Then for any given states \(x, y\), and choice of \(\varepsilon\), there exists a \(\theta\) such that:

1. \(\|\delta_{ER}^\varepsilon\| \leq \|\delta_{ER}^\varepsilon\|\) and \(\|\delta_{CVar}^\varepsilon\| \leq \|\delta_{CVar}^\varepsilon\|\).

2. The agent stabilizes inside \(B_{\varepsilon}(x^*)\) asymptotically for all RPMs and their respective \(\varepsilon\) follow \(\varepsilon_{CPT}^\varepsilon < \varepsilon_{CVar}^\varepsilon < \varepsilon_{ER}^\varepsilon\).

**Proof:** For 1), apply Proposition 3 and the fact that \(K_{\text{ER}} \subseteq K_{\text{CPT}}\) and \(K_{\text{CVar}} \subseteq K_{\text{CPT}}\).

For 2), employ an ISS argument to construct the \(B_{\varepsilon}(x^*)\) for each RPM considering the unforced system with \(u = k(x)\) in \((2)\) and \(P(x) = G(x)\delta\) being the forcing term after applying RPA controls \(u\) from \((7)\). From ISS, since the radius of the stability ball is proportional to the upper bound on \(\|P(x)\|\), the result immediately follows from the first part.

**Proposition 4** implies that, with an appropriate \(\theta\), CPT can not only produce the least perturbation among the three RPMs, but can also stabilize to the smallest ball around \(x^*\).
In Fig. 2d, we observe the opposite, where a very high threshold proceeds to construct the uncertain cost outcomes according to (7), with Proposition 1. The level set $R^\text{ER}(\xi) = \rho$ is shown in Fig. 2e as the inner most ellipse. We see that CPT is able to capture a more risk averse (larger) as well as more risk insensitive (smaller) perception than CVaR (and ER). This verifies the claims of Theorem 1 visually.

**RPA-CBF controller:** We consider a single agent with unicycle dynamics and a single obstacle whose dynamics evolve in the space $\mathcal{D} \subset \mathbb{R}^2$. We use the costs defined in the previous paragraph with $\tau = 0.5$ and $k_1, k_2 = 0.01$. The agent starts at $x(0) = (5, 2)^T$ (green dot) and its goal is $x^* = (10, 10)^T$ (motion up) while the obstacle moves from $(13, 13)^T$ (rhino in red ellipse) to $(2, 3)^T$ (motion down). If obstacle and vehicle follow along straight paths, a collision would occur and safety would be violated. To handle unicycle dynamics we use the projected point method to control a virtual point $p \in \mathbb{R}^2$, a distance $l$ along the direction of its heading. That is, $p = x + ld$, where $d$ is the direction vector corresponding to the agents heading $\phi$ and $x \in \mathbb{R}^2$ is the planar coordinates of the agent. With this we get the reverse transformation for the control inputs:

$$
\begin{bmatrix}
  u_1 \\
  u_2
\end{bmatrix} = \begin{bmatrix}
  \cos(\phi) & \sin(\phi) \\
  -\sin(\phi) & \cos(\phi)
\end{bmatrix} l.
$$

Where $u_1, u_2 \in \mathbb{R}$ are the linear and angular velocity inputs of the unicycle model and $u \in \mathbb{R}^2$ is the optimized input generated from (7) considering the $p$ dynamics $\dot{p} = u$. We use a standard proportional controller for $k(x)$ with a constant $(0.6, 0.6)$. We note that one can always appropriately tune the reference value $\rho$ by $l$ units to ensure safety w.r.t. $x$. The results of varying $\lambda$, $\gamma$ and $k$ are shown in Fig. 2f. For all the settings, the agent will collide with the obstacle (red ellipse) if it follows the nominal path (black line) from applying controls $k(x)$, thus making it unsafe. By using the controller $u$ from (7), the agent is able to swerve away from the obstacle in time and consider itself safe to reach the goal. We see that from Fig. 2d - Fig. 2f, the cfr $h_R$ remains positive throughout the execution, thus indicating that perceived safety is maintained irrespective of model and parameter choice. Next, we notice that by using CPT-CBF controller $u$ (Fig. 2d and Fig. 2e), the deviations from the nominal path correspondingly get more pronounced as the perceived risk increases (higher $\gamma$ and $\gamma$). Whereas, for CVaR-CBF this deviation (Fig. 2f) is comparatively minimal across its parameter spectrum. This is in accordance with our claims that CVaR is less inclusive (Theorem 1) and versatile (Proposition 1) than CPT, causing only minor changes in trajectories in comparison with CPT-based CBF controller.

From (Fig. 3), we see that the agent is able to reach the goal while maintaining $h \geq 0$ throughout, implying that perceived safety is maintained according to Definition 1. Furthermore, as before, we see that CPT is able to generate
a wider range of paths by tuning the risk aversion and risk sensitivity parameter than CVaR, thus capturing a greater variety of risk perception, which follows the theoretical arguments from Theorem 3 and Proposition 4. We also see that the agent also reaches the goal owing to the inherent stability properties of the nominal controller $k(x)$. 

Next, we consider an environment where there are three moving obstacles present and a single agent. We use the composition approach proposed in [9] to construct the barrier function to handle multiple obstacles. In this approach, the worst case (closest) obstacle is dealt with first using the min operator on the barrier functions generated by the corresponding obstacles. Here, the agent has to go from $(-15, -15)$ to $(15, 15)$, while the obstacles’ starting and goal points are respectively $(-17, 0), (0, 14), (10, -10)$ and $(17, 0), (0, -14), (-10, 10)$. Nominal controller $k(p)$ is generated with proportional constant $[1, 6, 1.6]$. The uncertainty radius is $r = 2.5$ and other cost constants are identical to the single agent setting. The results of varying the risk aversion $\lambda$, risk sensitivity $\gamma$ of CPT and $\kappa$ of CVaR are shown in Fig. 4.

Similar to the previous case (Fig. 3), we see that the agent is able to reach the goal while maintaining $h \geq 0$ throughout, implying that perceived safety is maintained.
VI. CONCLUSION AND FUTURE WORK

In this work, we have proposed a novel integration of CPT (a non-rational decision making model) into a safety-critical control scheme, to generate risk-perception-aware (RPA) controls (according to a DM’s risk profile) in an environment embedded with uncertain costs. Thus, opening new avenues to incorporate behavioral decision theory into safety-critical controls. Future directions include the design of learning frameworks to determine the risk profile of an observed agent and handling unknown obstacle dynamics.

REFERENCES

[1] S. S. Stevens, “Neural events and the psychophysical law,” Science, vol. 170, no. 3962, pp. 1043–1050, 1970.
[2] A. Tversky and D. Kahneman, “Advances in prospect theory: Cumulative representation of uncertainty,” Journal of Risk and Uncertainty, vol. 5, no. 4, pp. 297–323, 1992.
[3] O. Khatib, Real-Time Obstacle Avoidance for Manipulators and Mobile Robots. Springer New York, 1990, pp. 396–404.
[4] S. Prajna and A. Jadababaie, “Safety Verification of Hybrid Systems using Barrier Certificates,” in Hybrid Systems: Computation and Control, 2004, pp. 477–492.
[5] A. D. Ames, S. Coogan, M. Egerstedt, G. Notomista, K. Sreenath, and P. Tabuada, “Control barrier functions: Theory and applications,” in European Control Conference, 2019, pp. 3420–3431.
[6] P. Ong and J. Cortés, “Universal formula for smooth safe stabilization,” in IEEE Int. Conf. on Decision and Control, Nice, France, Dec. 2019, pp. 2373–2378.
[7] B. T. Lopez, J. J. E. Slotine, and J. P. How, “Robust adaptive control barrier functions: An adaptive and data-driven approach to safety,” IEEE Control Systems Letters, vol. 5, no. 3, pp. 1031–1036, 2021.
[8] Y. Chen, H. Peng, and J. Grizzle, “Obstacle avoidance for low-speed autonomous vehicles with barrier function,” IEEE Transactions on Control Systems Technology, vol. 26, no. 1, pp. 194–206, 2018.
[9] P. Glotfelter, J. Cortés, and M. Egerstedt, “Non-smooth approach to controller synthesis for Boolean specifications,” IEEE Transactions on Automatic Control, vol. 66, no. 11, 2021, to appear.
[10] S. Prajna, A. Jadababaie, and G. J. Pappas, “A framework for worst-case and stochastic safety verification using barrier certificates,” IEEE Transactions on Automatic Control, vol. 52, no. 8, pp. 1415–1428, 2007.
[11] A. Clark, “Control Barrier Functions for Complete and Incomplete Information Stochastic Systems,” in American Control Conference, 2019, pp. 2928–2935.
[12] M. J. Khojasteh, V. Dhiman, M. Franceschetti, and N. Atanasov, “Probabilistic safety constraints for learned high relative degree system dynamics,” in Proceedings of Learning for Dynamics and Control, 2020, pp. 781–792.
[13] S. Samuelson and I. Yang, “Safety-aware optimal control of stochastic systems using conditional value-at-risk,” in American Control Conference, 2018, pp. 6285–6290.
[14] M. Ahmadi, X. Xiong, and A. D. Ames, “Risk-averse control via cvar barrier functions: Application to bipedal robot locomotion,” IEEE Control Systems Letters, vol. 6, no. 1, pp. 878–883, 2021.
[15] A. Suresh and S. Martínez, “Planning under non-rational perception of uncertain spatial costs,” IEEE Robotics and Automation Letters, vol. 6, no. 2, pp. 4133–4140, 2021.
[16] S. Gao, E. Freijinger, and M. Ben-Akiva, “Adaptive route choices in risky traffic networks: A prospect theory approach,” Transportation Research Part C: Emerging Technologies, vol. 18, no. 5, pp. 727–740, 2010.
[17] A. R. Hota and S. Sundaram, “Game-Theoretic Protection against Networked sis Epidemics by Human Decision-Makers,” IFAC Papers Online, vol. 51, no. 34, pp. 145–150, 2019.
[18] S. Dhami, The Foundations of Behavioral Economic Analysis. Oxford University press, 2016.
[19] R. T. Rockafellar and S. Uryasev, “Optimization of conditional value-at-risk,” Journal of risk, vol. 2, no. 1, pp. 21–42, 2000.
[20] D. Prelec, “The probability weighting function,” Econometrica, vol. 66, no. 3, pp. 497–527, 1998.