PRELIMINARY PREPRINT VERSION – DO NOT CITE
This is not the version of record. The final version will appear in the Proceedings of IJCAI-ECAI 2026.

Distance Between CP-theory Preferences

Zihao Qi, **Erik Rauer**, **Samik Basu**
Iowa State University
{zihaoqi, etrauer, sbasu}@iastate.edu

Abstract

Qualitative preferences in CP-theory define a partial order over outcomes based on induced dominance relations. We investigate the problem of measuring divergence between such induced orders. We introduce two formally equivalent distance metrics, each leveraging a strategy based on different types of normalization. Preliminary empirical results support the soundness of the proposed methodologies and offer comparative insights into their computational properties. To the best of our knowledge, this is the first study to compute distance measures between qualitative preferences that jointly accounts for conditional dependencies and relative importance statements.

1 Introduction

Qualitative reasoning with preferences have made significant progress in real-world decision-making processes in recent years [Brafman and Domshlak, 2009; Santhanam *et al.*, 2016]. Such preferences are typically expressed over properties or attributes of outcomes using well-studied graphical languages like CP-nets and TCP-nets [Boutilier *et al.*, 2004; Brafman *et al.*, 2006], which collectively fall under the category of CP-theory languages [Wilson, 2004]. The semantics of these languages induce a partial order over outcomes, enabling the resolution of queries related to dominance (i.e., whether one outcome is better than another) and optimality (i.e., whether an outcome is not dominated by any other).

A key challenge in this context is reasoning with the preferences of multiple stakeholders. One approach is to merge all preferences and attempt to identify an outcome that aligns with everyone's (or majority's) interests [Rossi *et al.*, 2004]. However, in real-world scenarios, achieving consensus on optimality is often impractical. A more feasible strategy is to reach a compromise solution, where each stakeholder adjusts their preferences to reduce contradictions, thereby enabling a mutually acceptable consensus. Yet, this process may fail if stakeholders' preferences are fundamentally incompatible. Consequently, it becomes important to quantify the degree of difference between the preferences of various stakeholders. For example, if stakeholder A's preferences are more conflicting with stakeholder B's preferences than with stakeholder C's, then it is reasonable to infer that A and C are more likely to reach a consensus than A and B.

Driving Problem. The difference between preferences expressed by A and B is quantified in terms of the difference between the partial orders induced by the pair's preferences. For CP-net qualitative preferences, one of the notable works on distance measures is due to Loreggia *et al.* [2018], where the distance is an estimate of distance between partial orders as measured by their Kendall tau (KT) distance. Another relevant measure is the cumulative swap disagreement introduced by Ali *et al.* [2021] in the context of aggregating CP-nets. This function counts the number of swaps (outcome pairs differing in exactly one attribute) on which a candidate consensus network disagrees with the input preferences. Ali et al. utilize the distance as an objective function to be minimized for finding an optimal consensus CP-net.

Our Solution. We address the problem of distance measure in the general setting of CP-theory qualitative preferences. We demonstrate that our distance function realizes a metric space over partial orders induced by CP-theory preferences. We present a semantically equivalent alternate syntactic construct for CP-theory preferences (referred to as $\mathbb{A}$CP-theory language) that facilitates the computation of distance between preferences. We propose different computational strategies for our distance function and experimentally validate the viability of these strategies. To the best of our knowledge, our proposed distance computation is the first effective strategy for computing the distance between qualitative preferences that involve both conditional dependencies and set-based relative importance between attributes.

2 Qualitative Preferences

CP-theory. We present a brief overview of CP-theory preferences. For details we refer to [Santhanam *et al.*, 2016]. Preference statements in CP-theory express the desirability between a pair of values of a (preference attribute) variable optionally conditioned on the values of some other variables and in the context of relative importance between variables. Our focus, in this paper, is on binary domain variables. For instance, $X = x \wedge Y = y : \; Z = z \succ Z = \bar{z} \quad [U, V]$ is a statement that expresses the desirability of value z over $\bar{z}$ when the valuation of variables X and Y are x and y respectively and the variables U and V are relatively less important compared to Z. Semantically, this statement induces a *dominance* between a pair of outcomes o and o'. We will conclude that $o \succ o'$ as per the above preference statement if and only if the following holds:

1. valuations of Z in o and o' are z and $\bar{z}$, respectively (preference variable valuations).
2. valuations of X and Y in both o and o' are x and y, respectively (conditional variable valuations).
3. valuations of all variables other than X, Y, Z, U, V in o and o' are identical (all else being equal). We will refer to these variables as ceteris paribus variables [Boutilier *et al.*, 2004] and denote the set of these variables as $V^{\mathtt{cp}}$.

Observe that, the valuations of U and V can be anything in o and o' as they are relatively less important variables compared to Z. Further observe that, a preference statement can induce dominance between multiple pairs of outcomes.

Notations. The general syntax of CP-theory preference statements over a set of variables $\mathcal{V}$ is

$$\varrho: \ Z = z \succ Z = \bar{z} \quad [\Omega]$$

where ϱ captures the conjunction of conditions over variables in $Var(\varrho)$, and Ω represents the set of variables relatively less important than Z such that $Var(\varrho)$, Ω and $\{Z\}$ are pairwise disjoint sets. Whenever necessary, we use the subscript p to denote the conditions, relatively less important sets and ceteris paribus variables associated with the preference statement p. We denote the valuation of a conditional variable $X \in Var(\varrho)$ as $\varrho(X)$.

If $o \succ o'$ as per p in CP-theory, then we denote $o \succ_p o'$. The dominance relation induced by p is denoted by

$$\succ_{_p} = \{(o, o') \mid o \succ_p o'\}$$

Given a set of preferences P, the dominance relation induced by P is $\succ_{_P} = \bigcup_{p \in P} \succ_{_p}$. The semantics of P is captured by the transitive closure of $\succ_{_P}$, denoted by $\succ_{_P}^*$. The semantics, therefore, induces a partial order over the outcomes. In the proceeding, we will consider the set of preferences to be in irreducible or in minimal format; one where there are no redundancies in the conditionalities and relative importance.

Example 1. *Consider a CP-theory preference statement p:*

$$Z = z \succ Z = \bar{z} \quad [X, Y]$$

defined over the variable set $\{X, Y, Z\}$. There are 16 different dominance pairs (o, o') induced by the statement. The valuations of Z in o and o' are z and $\bar{z}$, respectively, with no restrictions on the valuations of X and Y.

3 Distance Between CP-theory Preferences

The distance between a pair of CP-theory preferences can be measured as the difference between the partial order induced by each element in the pair. For instance, KT-distance [Kendall, 1938] is measured as the cardinality of the set $\succ_{_{P_1}}^* \Delta \succ_{_{P_2}}^*$. That is,

$$d_{KT}(P_1, P_2) = |\succ_{_{P_1}}^* \Delta \succ_{_{P_2}}^*|$$

Here Δ denotes symmetric difference: for any sets A and B, $A \Delta B = (A \setminus B) \cup (B \setminus A)$.

Definition 1. *A metric space is a pair (M, d), where M is a set of elements and $dist$ is a function (distance or metric) $dist : M \times M \to \mathbb{R}$, which satisfies the following axioms for all $A, B, C \in M$:*

1. *$dist(A, B) \geq 0$;*
2. *$dist(A, B) = 0$ if and only if $A = B$;*
3. *$dist(A, B) = dist(B, A)$;*
4. *$dist(A, B) \leq dist(A, C) + dist(C, B)$.*

We present a distance function over CP-theory preferences which computes the difference between the dominance relations induced by preference statements P_1 and P_2. While this characterization of distance does not align with the difference between partial orders (transitive closure is not considered), it still realizes a metric space over the CP-theory preferences.

Theorem 1. *For any two CP-theory preference-sets A and B, a distance function defined as*

$$d_D(A, B) = |\succ_{_A} \Delta \succ_{_B}| \tag{1}$$

realizes a metric space $(\mathcal{P}, d_D)$, where any CP-theory preference set belongs to $\mathcal{P}$.

Proof. It is immediate that $\forall A, B \in \mathcal{P}$

$$\begin{array}{l} d_D(A, B) \geq 0 \\ d_D(A, B) = 0 \quad \Leftrightarrow \quad \succ_{_A} = \succ_{_B} \quad \Rightarrow \quad \succ_{_A}^* = \succ_{_B}^* \\ d_D(A, B) = d_D(B, A) \end{array}$$

Next consider the relationship $d_D(A, B) \leq d_D(A, C) + d_D(C, B)$. We prove that for any pair $(o, o') \in \succ_{_A} \Delta \succ_{_B}$, the pair either belongs to $\succ_{_B} \Delta \succ_{_C}$ or $\succ_{_A} \Delta \succ_{_C}$.

WLOG, let $(o, o') \in \succ_{_A} \setminus \succ_{_B}$. That is, (o, o') contributes to the $d_D(A, B)$. If $(o, o') \in \succ_{_C}$ then $(o, o') \in \succ_{_C} \setminus \succ_{_B}$. On the other hand, if $(o, o') \notin \succ_{_C}$ then $(o, o') \in \succ_{_A} \setminus \succ_{_C}$. Thus (o, o') contributes to the either $d_D(C, B)$ or $d_D(A, C)$, which implies $d_D(A, B) \leq d_D(B, C) + d_D(A, C)$. □

Computation of d_D requires generation of the dominance relations induced by the CP-theory preferences. We present an overestimate of d_D that can be computed from the syntactic structure of CP-theory statements. We conjecture (following [Loreggia *et al.*, 2018]) that distance functions defined using the syntactic difference between CP-theory preferences incur less execution time overhead compared to the computation of distance functions defined over the underlying dominance relations induced by the preferences. We experimentally validate this conjecture. To facilitate such a distance computation, we propose an alternate syntactic representation of CP-theory preferences.

$\mathbb{A}$CP-theory. Recall that the CP-theory preference statement: $\varrho: \ Z = z \succ Z = z' \quad [\Omega]$ induces dominance between pairs of outcomes o and o', where there is no relationship between the valuation of elements of Ω in o and o' (these variables are relatively less important compared to Z). Our alternate syntactic form for CP-theory preference statements explicitly captures this *lack of relationship between valuations of variables in* Ω. We refer to this syntactic form as $\mathbb{A}$CP-theory. Specifically, $\varrho: \ Z = z \succ Z = \bar{z} \quad (\Gamma)$ captures the dominance between o and o' where the o has the preferred valuation of Z, the conditions expressed in ϱ are satisfied in o and o', the valuations of variables $\in \Gamma$ differ in o and o' and all other variable valuations are identical in o and o'. We refer to the variables in Γ as the *difference variables*. As before, we use the subscript p to denote the conditions, the preference variable, and the difference variables associated to the preference statement p.

For a preference statement p in $\mathtt{A}$CP-theory, we denote the dominance relation induced by p as $\succ_p^a$ and for a set A of preferences in $\mathtt{A}$CP-theory, we denote the dominance relation induced by A as $\succ_A^a$.

Proposition 1. *Given a CP-theory statement p of the form $\varrho : \ Z = z \succ Z = \bar{z} \quad [\Omega]$, the set of statements in $\mathtt{A}$CP-theory that are semantically equivalent to p is*

$$\{\varrho : \ Z = z \succ Z = \bar{z} \quad (\Gamma) \mid \forall \Gamma \subseteq \Omega\}$$

For any set A of CP-theory preferences, there is a set A' of $\mathtt{A}$CP-theory preferences such that A' is the union of semantically equivalent $\mathtt{A}$CP-theory statements corresponding to each $p \in A$, thus ensuring $\succ_A = \succ_{A'}^a$.

We denote the function that transforms CP-theory preferences in A to $\mathtt{A}$CP-theory preferences as $\mathtt{C}_a(A)$.

3.1 Distance by Overestimating Dominance Difference

We apply normalization to preference statements in $\mathtt{A}$CP-theory in a fashion similar to that discussed by Loreggia *et al.* [2018]. Normalization amounts to augmenting each preference statement p with additional conditions over variables in $\mathcal{V} \setminus (Z_p \cup \Gamma_p)$ without altering the semantics of p. In short, all possible valuations of ceteris paribus variables $\in V_p^{\mathtt{cp}}$ are added as conditions resulting in $2^{|V_p^{\mathtt{cp}}|}$ different preference statements corresponding to p. For instance, consider the following preference statement p over $\mathcal{V} = \{X, Y, Z, W\}$

$$X = x : \ Z = z \succ Z = \bar{z} \quad (Y)$$

where the $V_p^{\mathtt{cp}} = \{W\}$. By normalization process, we obtain

$$X = x \wedge W = w : \ Z = z \succ Z = \bar{z} \quad (Y)$$
$$X = x \wedge W = \bar{w} : \ Z = z \succ Z = \bar{z} \quad (Y)$$

The process of normalization (adding redundant conditions) results in preference statements such that there are no Ceteris Paribus variables associated with them. We will refer to this normalization operation as *standard normalization* and denote its application on an $\mathtt{A}$CP-theory preference set as the function $\mathtt{N}_s(A)$. We denote the dominance induced by a normalized $\mathtt{A}$CP-theory preference statement p by $\succ_p^{an}$.

In order to estimate the distance, we first compute $|\succ_p^{an}|$. The valuation depends on the cardinality of Γ_p as the pair of outcomes related by the $\succ_p^{an}$ must differ in the valuations of each variable in Γ_p. It is immediate that

$$|\succ_p^{an}| = 2^{|\Gamma_p|} \tag{2}$$

Proposition 2. *Given a set A of CP-theory statements, let $A' = \mathtt{N}_s(\mathtt{C}_a(A))$ be the set of normalized $\mathtt{A}$CP-theory statements. Then,*

$$\bigcup_{p \in A} \succ_p = \bigcup_{q \in A'} \succ_q^{an} \quad \textit{hence,} \quad \succ_A = \succ_{A'}^{an}$$

Therefore, for any two sets A and B of CP-theory statements, if $A' = \mathtt{N}_s(\mathtt{C}_a(A))$ and $B' = \mathtt{N}_s(\mathtt{C}_a(B))$ then

$$d_D(A, B) = |\succ_A \Delta \succ_B| = |\succ_{A'}^{an} \Delta \succ_B^{an}| = d_D(A', B')$$

Example 2. *For the CP-theory statement in Example 1, the corresponding $\mathtt{A}$CP-theory preferences are*

$$\begin{array}{llllll} p_1 & Z = z \succ Z = \bar{z} & () & p_2 & Z = z \succ Z = \bar{z} & (X) \\ p_3 & Z = z \succ Z = \bar{z} & (Y) & p_4 & Z = z \succ Z = \bar{z} & (X, Y) \end{array}$$

Subsequent normalization results in the preferences

$$\begin{array}{llll} p_{11} & X = x \wedge Y = y & : Z = z \succ Z = \bar{z} & () \\ p_{12} & X = x \wedge Y = \bar{y} & : Z = z \succ Z = \bar{z} & () \\ p_{13} & X = \bar{x} \wedge Y = y & : Z = z \succ Z = \bar{z} & () \\ p_{14} & X = \bar{x} \wedge Y = \bar{y} & : Z = z \succ Z = \bar{z} & () \\ p_{21} & Y = y & : Z = z \succ Z = \bar{z} & (X) \\ p_{22} & Y = \bar{y} & : Z = z \succ Z = \bar{z} & (X) \\ p_{31} & X = x & : Z = z \succ Z = \bar{z} & (Y) \\ p_{32} & X = \bar{x} & : Z = z \succ Z = \bar{z} & (Y) \\ p_{41} & & : Z = z \succ Z = \bar{z} & (X, Y) \end{array}$$

Observe that the preference statement $p_{1,i}$ induces $2^0 = 1$ dominance relationship for each i. Preference statement p_{2i} as well as p_{3i} induces $2^1 = 2$ dominance relationships. Finally, statement p_{4i} induces $2^2 = 4$ relations.

Definition 2. *Given sets A and B of CP-theory statements, let $A' = \mathtt{N}_s(\mathtt{C}_a(A))$ and $B' = \mathtt{N}_s(\mathtt{C}_a(B))$. We define*

$$d(A, B) = \sum_{p \in A' \Delta B'} |\succ_p^{an}|$$

We establish the following properties of the distance measure d.

Theorem 2. *Given sets A and B of CP-theory statements, $d_D(A, B) \leq d(A, B)$.*

Proof. Let $A' = \mathtt{N}_s(\mathtt{C}_a(A))$ and $B' = \mathtt{N}_s(\mathtt{C}_a(B))$ be the sets of normalized $\mathtt{A}$CP-theory statements corresponding to CP-theory preferences in sets A and B. Note that, $d_D(A, B) = d_D(A', B')$ due to Proposition 2. For any $(o, o') \in \succ_{A'}^{an} \setminus \succ_{B'}^{an}$, there exists some $p \in A' \setminus B'$ such that $(o, o') \in \succ_p^{an}$. Therefore,

$$\succ_{A'}^{an} \setminus \succ_{B'}^{an} \quad \subseteq \quad \bigcup_{p \in A' \setminus B'} \succ_p^{an}$$

Proceeding further,

$$|\succ_{A'}^{an} \setminus \succ_{B'}^{an}| \quad \leq \quad \left| \bigcup_{p \in A' \setminus B'} \succ_p^{an} \right| \leq \sum_{p \in A' \setminus B'} |\succ_p^{an}| \tag{3}$$

Simliarly, we have $|\succ_{B'}^{an} \setminus \succ_{A'}^{an}| \leq \sum_{p \in B' \setminus A'} |\succ_p^{an}|$. Therefore,

$$\begin{aligned} d_D(A, B) &= d_D(A', B') \quad \text{due to Proposition 2} \\ &= |\succ_{A'}^{an} \Delta \succ_{B'}^{an}| \quad \text{by Equation 1} \\ &= |\succ_{A'}^{an} \setminus \succ_{B'}^{an}| + |\succ_{B'}^{an} \setminus \succ_{A'}^{an}| \\ &\leq \textstyle\sum_{p \in A' \setminus B'} |\succ_p^{an}| + \sum_{p \in B' \setminus A'} |\succ_p^{an}| \\ &= \textstyle\sum_{p \in A' \Delta B'} |\succ_p^{an}| \\ &= d(A, B) \quad \text{by Definition 2} \end{aligned}$$

$\square$

Theorem 3. *The function $d(A, B)$ realizes a metric space $(\mathcal{P}, d)$ where any CP-theory preference set belongs to $\mathcal{P}$.*

Proof. $\forall A, B \in \mathcal{P}$ the following properties hold.

$d(A, B) \geq 0$ as $d_D(A, B) \geq 0$ (due to Theorem 2) and $d_D(A, B) \leq d(A, B)$.

Next consider, $d_D(A, B) = 0$. Due to Theorem 1, $\succ_{A} = \succ_{B}$. Let $A' = \mathtt{N}_s(\mathtt{C}_a(A))$ and $B' = \mathtt{N}_s(\mathtt{C}_a(B))$. Translation and normalization do not alter the semantics of dominance. Hence $\succ^{an}_{A'} = \succ^{an}_{B'}$ and there exists no $p \in \succ^{an}_{A'} \Delta \succ^{an}_{B'}$. Therefore $d_D(A, B) = 0$ implies $d(A, B) = 0$. The converse $d(A, B) = 0$ implies $d_D(A, B) = 0$ is immediate as $d(A, B) \geq d_D(A, B)$ due to Theorem 2.

The symmetric relationship $d(A, B) = d(B, A)$ follows from the definition d.

Finally, let A, B, C be three sets of CP-theory statements and let A', B', C' be the corresponding normalized sets in $\mathbb{A}$CP-theory. Let $p \in A' \Delta B'$ and WLOG consider $p \in A' \setminus B'$. Therefore, $| \succ^{an}_{p} |$ contributes to $d(A, B)$.

If $p \in C'$, then $p \in C' \setminus B'$ and hence $| \succ^{an}_{p} |$ contributes to $d(C, B)$. If instead $p \notin C'$, then $p \in A' \setminus C'$ and $| \succ^{an}_{p} |$ contributes to $d(A, C)$. Therefore, in either case, if $| \succ^{an}_{p} |$ contributes to $d(A, B)$, then it also contributes to $d(A, C) + d(C, B)$. It follows that, $d(A, B) \leq d(A, C) + d(C, B)$. $\square$

Example 3. *Consider a set A*

$$\begin{array}{ll} a_1 & X_1 = x_1 : X_4 = x_4 \succ X_4 = \bar{x}_4 \ (X_5) \\ a_2 & X_2 = x_2 : X_4 = x_4 \succ X_4 = \bar{x}_4 \ (X_1) \end{array}$$

and set B

$$\begin{array}{ll} b & X_3 = x_3 : X_4 = x_4 \succ X_4 = \bar{x}_4 \ (X_5) \end{array}$$

of $\mathbb{A}$CP-theory preferences over variables $\{X_1, X_2, X_3, X_4, X_5\}$. Normalizing the preference statements results in the following statements for the set A

$$\begin{array}{ll} a_{11} & X_1 = x_1 \wedge X_2 = x_2 \wedge X_3 = x_3 : \ X_4 = x_4 \succ X_4 = \bar{x}_4 \ (X_5) \\ a_{12} & X_1 = x_1 \wedge X_2 = \bar{x}_2 \wedge X_3 = x_3 : \ X_4 = x_4 \succ X_4 = \bar{x}_4 \ (X_5) \\ a_{13} & X_1 = x_1 \wedge X_2 = x_2 \wedge X_3 = \bar{x}_3 : \ X_4 = x_4 \succ X_4 = \bar{x}_4 \ (X_5) \\ a_{14} & X_1 = x_1 \wedge X_2 = \bar{x}_2 \wedge X_3 = \bar{x}_3 : \ X_4 = x_4 \succ X_4 = \bar{x}_4 \ (X_5) \\ a_{21} & X_2 = x_2 \wedge X_3 = x_3 \wedge X_5 = x_5 : \ X_4 = x_4 \succ X_4 = \bar{x}_4 \ (X_1) \\ a_{22} & X_2 = x_2 \wedge X_3 = \bar{x}_3 \wedge X_5 = x_5 : \ X_4 = x_4 \succ X_4 = \bar{x}_4 \ (X_1) \\ a_{23} & X_2 = x_2 \wedge X_3 = x_3 \wedge X_5 = \bar{x}_5 : \ X_4 = x_4 \succ X_4 = \bar{x}_4 \ (X_1) \\ a_{24} & X_2 = x_2 \wedge X_3 = \bar{x}_3 \wedge X_5 = \bar{x}_5 : \ X_4 = x_4 \succ X_4 = \bar{x}_4 \ (X_1) \end{array}$$

Normalizing the preference statements results in the following statements for the set B

$$\begin{array}{ll} b_1 & X_1 = x_1 \wedge X_2 = x_2 \wedge X_3 = x_3 : \ X_4 = x_4 \succ X_4 = \bar{x}_4 \ (X_5) \\ b_2 & X_1 = \bar{x}_1 \wedge X_2 = x_2 \wedge X_3 = x_3 : \ X_4 = x_4 \succ X_4 = \bar{x}_4 \ (X_5) \\ b_3 & X_1 = x_1 \wedge X_2 = \bar{x}_2 \wedge X_3 = x_3 : \ X_4 = x_4 \succ X_4 = \bar{x}_4 \ (X_5) \\ b_4 & X_1 = \bar{x}_1 \wedge X_2 = \bar{x}_2 \wedge X_3 = x_3 : \ X_4 = x_4 \succ X_4 = \bar{x}_4 \ (X_5) \end{array}$$

The symmetric difference set of the above two sets of preference statements is $\{a_{13}, a_{14}, a_{21}, a_{22}, a_{23}, a_{24}, b_2, b_4\}$, so the distance is $d(A, B) = 16$.

3.2 Distance Overestimation with Context-Uniform Normalization

Recall that we have applied standard normalization where each preference statement p is normalized with respect to $V^{\mathtt{cp}}_p$. We present here another normalization strategy. As before, the preferences are translated to $\mathbb{A}$CP-theory, resulting in new sets A' and B' corresponding to CP-theory preference sets A and B. Let $Pa_{A'}(X)$ be the set of conditional variables that appear in some preference statement in A' which expresses preferences over the variable X. Each $\mathbb{A}$CP-theory preference statement p is then normalized with respect to $Pa_{A'}(Z_p) \setminus (Var(\varrho_p) \cup \Gamma_p)$. We refer to this normalization as *Context-uniform normalization*, and the dominance relation induced by preference p post normalization is denoted by $\succ^{an_{cu}}_{p}$.

We denote the function realizing the normalization operation as $\mathtt{N}_{cu}(.)$, which takes as input a set of $\mathbb{A}$CP-theory preferences and outputs a resultant set where preferences are context-uniform normalized. Before presenting the distance computation strategy using context-uniform normalization, we introduce the concept of compatible preferences–preferences that induce some common set of dominance relationship. Formally,

Definition 3. *Given two context-uniform normalized $\mathbb{A}$CP-theory preference statements $p \in A'$ and $q \in B'$, we say that they are compatible, denoted by $p \bowtie_{A'B'} q$ (or $(p, q) \in \bowtie_{A'B'}$), iff the following conditions hold:*

1. $\begin{array}{ll} p & \varrho_p : \ Z = z \succ Z = \bar{z} \ (\Gamma) \\ q & \varrho_q : \ Z = z \succ Z = \bar{z} \ (\Gamma) \end{array}$
2. $\forall X \in Var(\varrho_p) \cap Var(\varrho_q), \ \varrho_p(X) = \varrho_q(X).$

Definition 4. *Given sets A and B of CP-theory statements, let $A' = \mathtt{N}_{cu}(\mathtt{C}_a(A))$ and $B' = \mathtt{N}_{cu}(\mathtt{C}_a(B))$. We define*

$$d_{cu}(A, B) = \sum_{p \in A'} 2^{|\mathcal{V}|-1-|Var(\varrho_p)|} + \sum_{p \in B'} 2^{|\mathcal{V}|-1-|Var(\varrho_p)|} - 2 \times \sum_{(p,q) \in \bowtie_{A'B'}} 2^{|\mathcal{V}|-1-|Var(\varrho_p \wedge \varrho_q)|}$$

Example 4. *Consider the same sets of $\mathbb{A}$CP-theory preferences A and B in Example 3. The context-uniform normalization of A results in the following set A' of statements*

$$\begin{array}{lll} a_{11} & X_1 = x_1 \wedge X_2 = x_2 : & X_4 = x_4 \succ X_4 = \bar{x}_4 \ (X_5) \\ a_{12} & X_1 = x_1 \wedge X_2 = \bar{x}_2 : & X_4 = x_4 \succ X_4 = \bar{x}_4 \ (X_5) \\ a_2 & X_2 = x_2 : & X_4 = x_4 \succ X_4 = \bar{x}_4 \ (X_1) \end{array}$$

The context-uniform normalization for the set B is $B' = B$. Note that $\bowtie_{A'B'} = \{(a_{11}, b), (a_{12}, b)\}$ and

$$\begin{array}{l} Var(\varrho_{a_{11}} \wedge \varrho_b) = (X_1 = x_1 \wedge X_2 = x_2 \wedge X_3 = x_3) \\ Var(\varrho_{a_{12}} \wedge \varrho_b) = (X_1 = x_1 \wedge X_2 = \bar{x}_2 \wedge X_3 = x_3) \end{array}$$

So we have

$$\begin{array}{l} \sum_{p \in A'} 2^{|\mathcal{V}|-1-|Var(\varrho_p)|} = 2^2 + 2^2 + 2^3 = 16 \\ \sum_{p \in B'} 2^{|\mathcal{V}|-1-|Var(\varrho_p)|} = 2^3 = 8 \\ \sum_{(p,q) \in \bowtie_{A'B'}} 2^{|\mathcal{V}|-1-|Var(\varrho_p \wedge \varrho_q)|} = 2 + 2 = 4 \end{array}$$

Hence the distance is $d_{cu} = 16 + 8 - 2 \times 4 = 16$.

Theorem 4. *For any pair A and B sets of CP-theory statements, $d(A, B) = d_{cu}(A, B)$.*

Proof. Given sets A and B, let A', B' be their context-uniform forms and A_{std}, B_{std} be their standard normalized forms. We define the weight of a set of standard statements

S as $w(S) = \sum_{s \in S} | \succ_s^{an} |$. From Definition 2, $d(A, B) = w(A_{std} \Delta B_{std}) = w(A_{std}) + w(B_{std}) - 2w(A_{std} \cap B_{std})$.

First, consider A_{std}. It partitions into disjoint sets $S(q)$ for each $q \in A'$, where $S(q)$ contains all standard statements derived from q. The set $S(q)$ consists of all statements matching q's structure (same $Z, z, \bar{z}, \Gamma$) but with full assignments to variables not in $\{Z\} \cup Var(\varrho_q) \cup \Gamma_q$. The number of such statements is $2^{|\mathcal{V}|-1-|Var(\varrho_q)|-|\Gamma_q|}$, and each has dominance size $2^{|\Gamma_q|}$. Thus, the weight is $w(S(q)) = 2^{|\mathcal{V}|-1-|Var(\varrho_q)|}$. Summing over A', we get $w(A_{std}) = \sum_{q \in A'} 2^{|\mathcal{V}|-1-|Var(\varrho_q)|}$. Similarly for B'.

Next, consider the intersection $A_{std} \cap B_{std}$. A standard statement s is in the intersection if and only if it is derived from some $p \in A'$ and $q \in B'$. This implies that p and q must share the same preference structure and their conditions must be consistent. Thus, (p, q) must be a compatible pair. Thus, the intersection partitions into disjoint sets $S(p) \cap S(q)$ for each compatible pair $(p, q) \in \bowtie_{A'B'}$. The set $S(p) \cap S(q)$ contains standard statements consistent with $\varrho_p \wedge \varrho_q$. Using the same counting argument as before, the number of such statements is $2^{|\mathcal{V}|-1-|Var(\varrho_p \wedge \varrho_q)|-|\Gamma|}$, and the weight is:

$$w(S(p) \cap S(q)) = 2^{|\mathcal{V}|-1-|Var(\varrho_p \wedge \varrho_q)|}.$$

Substituting these terms into the distance equation yields $d_{cu}(A, B)$. □

Corollary 1. *Given any sets A and B of CP-theory preferences, $d_D(A, B) \leq d(A, B) = d_{cu}(A, B)$.*

Corollary 2. *The function $d_{cu}(A, B)$ realizes a metric space $(\mathcal{P}, d_{cu})$ where any CP-theory preference set belongs to $\mathcal{P}$.*

Complexity analysis. The computations are polynomial in the size of the normalized ACP-theory representation, although this representation may be exponentially larger than the compact CP-theory input in the worst case. This is due to the normalization over multiple difference-variables resulting from relatively less important variables. The benefit of d_{cu} is therefore not the elimination of this worst-case expansion, but the avoidance of explicit construction of the outcome-level dominance graph, which may also contain exponentially many outcome pairs. Moreover, context-uniform normalization is performed independently for each preference set, so normalized representations can be precomputed once and reused across all pairwise comparisons. Given these precomputed representations, evaluating d_{cu} reduces to linear summations over the normalized sets and compatibility checks between statements with matching preference structure.

4 Experimental Evaluation

Data Set. We generate a collection of 100 CP-theory sets and compute the distance measures between all pairs of these sets. We first generate CP-net statements, by adopting the strategy outlined in [Allen *et al.*, 2017]. We then randomly introduce relatively less important variables into selected statements. These variables are chosen from the set V_p^{cp}, which denotes the ceteris paribus attributes for the original preference statement p. Each set of CP-theory statements generated in this way is parameterized by the following: (a) the number

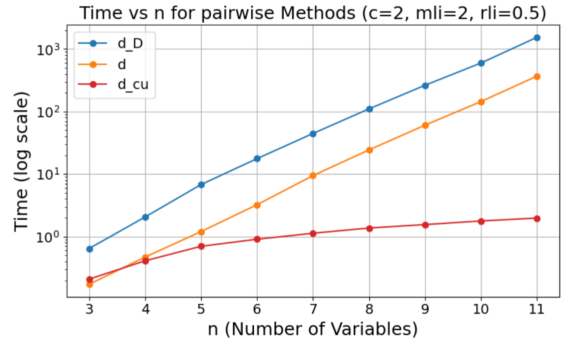

Figure 1: Runtime comparison for pairwise methods.

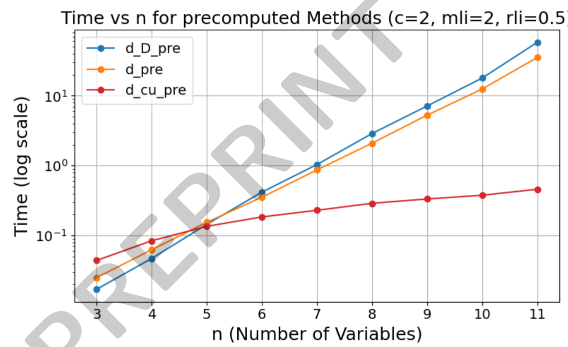

Figure 2: Runtime comparison for precomputed methods.

($\mathbf{n}$) of attributes, (b) the maximum number ($\mathbf{c}$) of conditional attributes per statement, (c) the maximum number (**mli**) of relatively less important attributes per statement, and (d) the probability (**rli**) that a statement includes relatively less important attributes.

Evaluation Objective. The primary objective of our experimental study is to assess the execution time of the distance measure and quantitatively evaluate the differences between d_D and the proposed distance measure.

All experiments are conducted on an Apple M2 CPU with 16 GB memory running macOS 15.5 and Python 3.13.5.

4.1 Runtime Performance

All execution time plots report time in seconds. Figure 1 and Figure 2 show how runtime changes with n under pairwise and precomputed implementations. In the pairwise approach, normalization and computation are performed individually for each pair of CP-theories. In contrast, the precomputed approach first normalizes all CP-theories in the dataset and then performs pairwise comparisons based on these normalized representations. This strategy substantially reduces redundancy in normalization and improves efficiency, especially for large datasets. Notably, pre-computing the context-uniform normalization (Figure 2) makes d_{cu} extremely efficient. With 10 binary preference variables describing 1024 outcomes, precomputed d_{cu} completes $\binom{100}{2} = 4950$ pairwise distance computations within a few seconds.

Figures 3–5 show the effects of c, mli, and rli. As c ap-

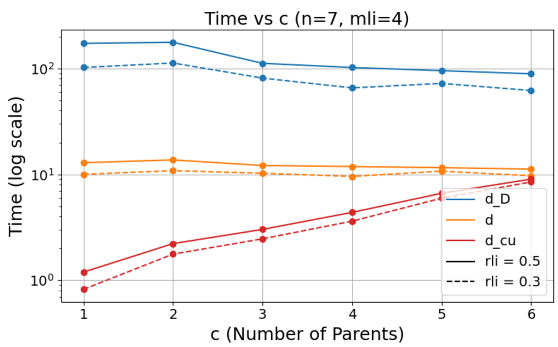

Figure 3: Runtime comparison as a function of c ($n = 7$).

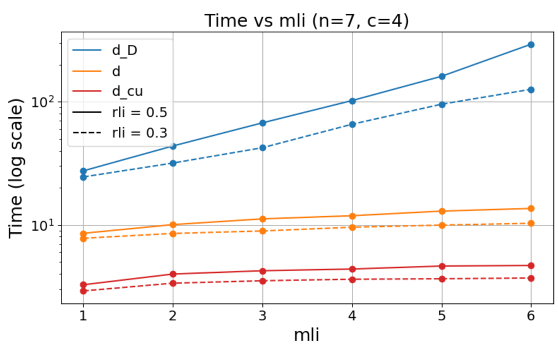

Figure 4: Runtime comparison as a function of mli ($n = 7$).

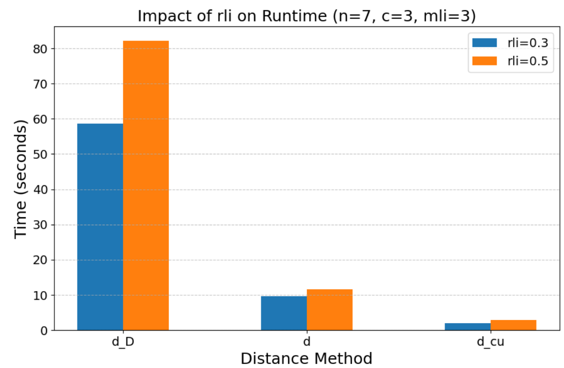

Figure 5: Impact of rli on runtime across distance measures.

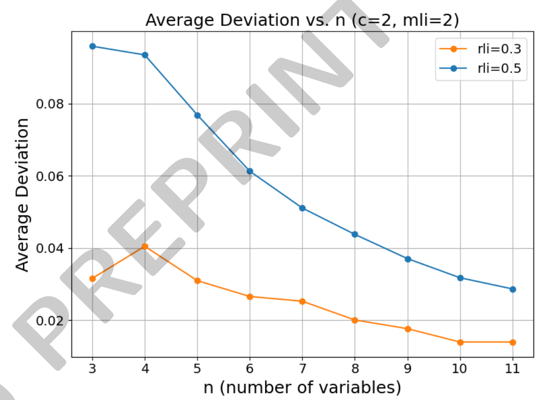

Figure 6: Average deviation between d and d_D as a function of n.

proaches n, the overhead gap between d and d_{cu} narrows. The reason is that less conditional attributes in statements result in less normalized statements for d_{cu}, which leads to less computation overhead. Larger mli and rli generally increase runtime for all measures.

4.2 Approximation Quality

We have proposed an overestimating distance metric d that is computed efficiently from the syntactic structure of preferences rather than the induced preference graphs. We evaluate the quality of this approximation relative to the reference metric d_D. For our dataset, the average relative deviation, $\frac{d-d_D}{d_D}$, is 5.46%. Figure 6 shows how the average deviation between d and d_D changes as the number of variables n increases, with fixed values of c and mli. As n grows, the deviation decreases for both values of rli, indicating that the approximation d becomes more aligned with the reference distance d_D in larger CP-theories. The higher deviation observed for $rli = 0.5$ suggests that a greater proportion of relatively less important variables amplifies the mismatch between d and d_D.

Figure 7 and Figure 8 plot the average relative mismatch between d and d_D as mli increases, for fixed n and $rli = 0.3$ and 0.5, respectively. Each line represents a different value of c. The results show that the deviation of the overestimate from the reference value increases with both mli and c. Moreover, deviation levels are consistently higher for $rli = 0.5$ than for $rli = 0.3$, consistent with earlier observations that a greater presence of relatively less important variables amplifies the mismatch between d and d_D.

4.3 Comparison with Existing Measures

We compare the computational performance of our d_{cu} implementation with I-CPD, a distance measure for CP-nets defined by [Loreggia *et al.*, 2018]. To handle the potential exponential complexity of preference comparison, our library employs a high-performance bitmasked evaluation architecture. This architecture provides a unified foundation for both metrics, ensuring the experimental comparison is fair and focuses on the inherent efficiency of the algorithms themselves.

The core idea shared by both metrics is the need to rapidly identify overlapping contexts between preference statements. While they leverage this information differently, with d_{cu} using it to find agreements for the intersection term $|\succ_A \cap \succ_B|$ and I-CPD using it to find disagreements for counting swaps, the underlying computational bottleneck is identical. To handle this efficiently, we represent the fixed variables (parents) of each statement as a mask m (where the i-th bit is 1 if the i-th variable is fixed) and their assignments as a value vector v (representing 0 or 1 valuations). We can then determine the compatibility of any two statements s_1 and s_2 using a highly efficient bitwise expression:

$$\text{is\_compatible} = \neg((m_1 \wedge m_2) \wedge (v_1 \oplus v_2))$$

where $\wedge$ is bitwise AND, and $\oplus$ is bitwise XOR. If compatible, the size of the overlapping context (the number of flip-

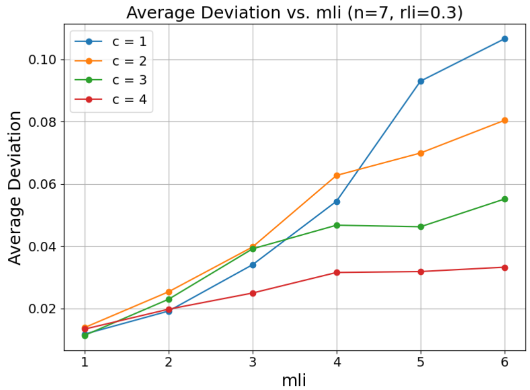

Figure 7: Average deviation between d and d_D as a function of mli.

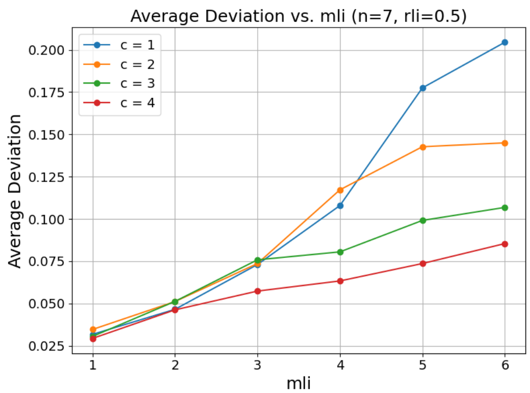

Figure 8: Average deviation between d and d_D as a function of mli.

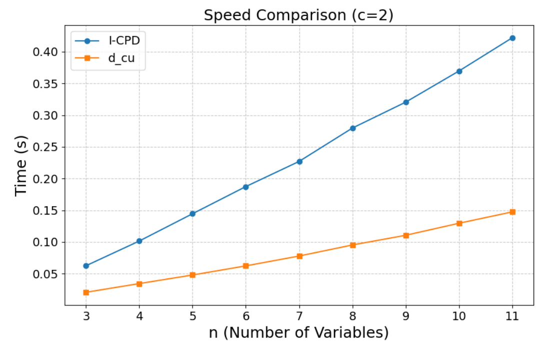

Figure 9: Comparison between our bitmasked d_{cu} implementation and the I-CPD approach ($c = 2$).

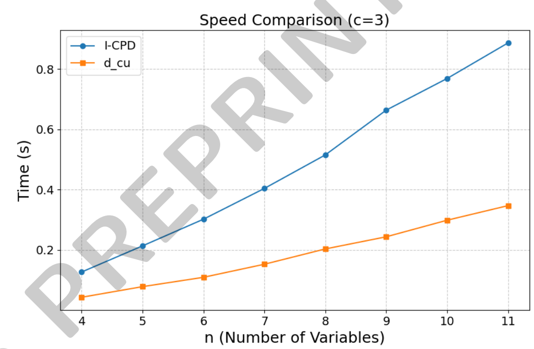

Figure 10: Comparison between our bitmasked d_{cu} implementation and the I-CPD approach ($c = 3$).

pable edges affected) is computed as $2^{n-1-|m_1 \cup m_2|}$ using a population count (popcount) operation. This value serves as the base weight; it is incorporated directly into the calculation of d_{cu}, whereas for I-CPD, it is adjusted by metric-specific scaling factors to determine the final contribution to the distance. As illustrated in Figures 9 and 10, this bitmasked approach allows our library to process thousands of statement comparisons in a fraction of a second. By isolating implementation efficiency within a shared optimization framework, our results accurately reflect the comparative scaling characteristics of d_{cu} and I-CPD.

5 Related Work

The problem of quantifying similarity or distance between partial orders arises naturally in various contexts such as rank aggregation, social choice theory, and hierarchy comparison. Classical distance measures such as Spearman's footrule and Kendall's tau [Kendall, 1948; Diaconis and Graham, 1977] were initially developed for total orders and have since been extended to accommodate partial rankings.

Related to our context is the distance computation between CP-nets proposed by [Loreggia *et al.*, 2018]. They define the distance as a function of the Kendall-tau distance between the induced partial orders, an extension of the classical metric that counts the number of pairwise disagreements between orders. Given the exponential size of the outcome space, directly computing Kendall-tau distances is inefficient. To address this, the authors propose two polynomial-time approximations: the O-legal CP-net Distance (O-CPD) and the Induced CP-net Distance (I-CPD). O-CPD assumes a shared variable ordering between CP-nets, enabling efficient comparison via conditional preference tables. I-CPD generalizes this by removing the ordering constraint, offering broader applicability while retaining computational tractability. In contrast, our proposed distance measure strategies is applicable in more generalized settings, preferences expressed in CP-theory languages.

6 Conclusion

This paper introduces a new framework for measuring distance between outcome orders induced by CP-theories that account for both conditional dependencies and relative importance, thus addressing a gap in existing preference comparison approaches. Our preliminary experiments highlight the validity and show its practical viability. As part of future work, we plan to apply the distance measures to strategize effective aggregation of multi-agent preferences.

Acknowledgments

This work is supported in part by NSF grant IIS 2225823.

References

[Ali *et al.*, 2021] Abu Mohammad Hammad Ali, Howard J Hamilton, Elizabeth Rayner, Boting Yang, and Sandra Zilles. Aggregating preferences represented by conditional preference networks. In *International Conference on Algorithmic Decision Theory*, pages 3–18. Springer, 2021.

[Allen *et al.*, 2017] Thomas E. Allen, Judy Goldsmith, Hayden Elizabeth Justice, Nicholas Mattei, and Kayla Raines. Uniform random generation and dominance testing for cp-nets. *Journal of AI Research*, 2017.

[Boutilier *et al.*, 2004] C. Boutilier, R. I. Brafman, C. Domshlak, H. H. Hoos, and D. Poole. CP-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements. *Journal of Artificial Intelligence Research*, 21:135–191, 2004.

[Brafman and Domshlak, 2009] Ronen Brafman and Carmel Domshlak. Preference handling-an introductory tutorial. *AI magazine*, 30(1):58–58, 2009.

[Brafman *et al.*, 2006] R. I. Brafman, C. Domshlak, and S. E. Shimony. On graphical modeling of preference and importance. *Journal of Artificial Intelligence Research*, 25:389–424, 2006.

[Diaconis and Graham, 1977] Persi Diaconis and Ronald L Graham. Spearman's footrule as a measure of disarray. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 39(2):262–268, 1977.

[Kendall, 1938] Maurice G Kendall. A new measure of rank correlation. *Biometrika*, 30(1-2):81–93, 1938.

[Kendall, 1948] Maurice George Kendall. *Rank correlation methods*. Griffin, 1948.

[Loreggia *et al.*, 2018] Andrea Loreggia, Nicholas Mattei, Francesca Rossi, and K. Brent Venable. On the distance between cp-nets. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '18, page 955–963, Richland, SC, 2018. International Foundation for Autonomous Agents and Multiagent Systems.

[Rossi *et al.*, 2004] F. Rossi, K. B. Venable, and T. Walsh. mcp nets: Representing and reasoning with preferences of multiple agents. In *Proceedings of the 19th National Conference on Artificial Intelligence*, AAAI'04, pages 729–734. AAAI Press, 2004.

[Santhanam *et al.*, 2016] Ganesh Ram Santhanam, Samik Basu, and Vasant Honavar. *Representing and reasoning with qualitative preferences: Tools and applications*, volume 10 of *Synthesis Lectures on Artificial Intelligence and Machine Learning*. Morgan and Claypool Publishers, United States, 2016.

[Wilson, 2004] Nic Wilson. Extending CP-nets with stronger conditional preference statements. In *Proceedings of the 19th National Conference on Artifical Intelligence*, AAAI'04, page 735–741. AAAI Press, 2004.