Preference for Simplicity *

Rui Tang † Mu Zhang ‡

[Preliminary version. Please contact us before citing]

October 26, 2021

Abstract

Evaluating and comparing the difficulty of problems are playing an important role in decision making. In this paper, we define and axiomatize the preference for simplicity as a subjective measure for simplicity of problems and introduce the concept of *difficulty levels* as endogenous outcomes. A prior and an induced preference over finite lotteries of natural numbers characterize the the preference for simplicity. The decision maker transforms each problem into a prospect of difficulty levels and then rank prospects according to the induced preference over lotteries. We then characterize the the preference for simplicity with more structures like the mixture continuous utility representation and the bounded expected utility representation. We also define the comparison of accuracy and variation aversion for the preference for simplicity and relate our models to existing models on ambiguity.

Keywords: Preference for simplicity; Subjective probability; Probabilistic sophistication; Savage; Expected uncertainty utility

---

*We thank Faruk Gul for the inspiration, guidance and suggestions, especially as the third-year paper advisor for Mu Zhang. We also thank Jay Lu, Wolfgang Pesendorfer and Satoru Takahashi for helpful comments in the discussion, as well as participants in Microeconomic Theory Student Lunch Seminar at Princeton University. All remaining errors are our own.

†Department of Economics, Princeton University, ruit@princeton.edu

‡Department of Economics, Princeton University, muz@princeton.edu

Contents

1 Introduction

A student has accepted a conditional offer from her dream university which requires her to qualify at least one of two English level tests. Ex ante, the student knows that the questions in both tests will be drawn from the same pool and she has solutions to questions in the past years. The only difference is that the proportion of past questions in Test 1 is 1/2, while the proportion of past questions in Test 2 is 1/3. The student chooses to register for Test 1 since with a higher probability, she will encounter questions with known answers. In other words, she believes that *Test 1 is simpler than Test 2*.

Indeed, simplicity plays an important role in decision making. When there are multiple problems and an agent only needs to solve for one, she tends to choose the simplest one given that the payoffs of solving each problem are identical. For instance, students take modules that are relatively easier for them. Managers choose projects that are not hard to finish. Basketball players seek to score in an easy way instead of making tough shots. Despite its prevalence in the real world, simplicity is not yet well-formulated in decision-theoretical literature. In this paper, we formally study the preference over problems where a problem consists of multiple alternatives among which only one will prove to be correct. We propose a model called *the preference for simplicity* as a subjective measure of the difficulty of problems and provide its characterization results, which can be potentially used to explain the prevalence of simple contracts. To illustrate our model, consider the following example.

Example 1. *A job seeker needs to make a choice between two interviews due to time conflict. She is indifferent between the two job positions so that she would like to attend the simpler interview. According to her prior knowledge, there are respectively three potential interviewers for each of the interview who might be in charge. After meeting the interviewer, she could choose to behave aggressively (A), humbly (H) or neutrally (N). Each interviewer likes one strategy and hates the other two. The prior knowledge of the job seeker about the two interviews is listed in the following tables.*

Table 1: Interview 1

Interviewer	Probability in Charge	Likes	Hates
A1	1/3	A	H, N
A2	1/3	Unknown	A
A3	1/3	Unknown	H

Table 2: Interview 2

Interviewer	Probability in Charge	Likes	Hates
B1	1/3	H	A, N
B2	1/3	Unknown	N
B3	1/3	Unknown	Unknown

By choosing interview 1, with probability 1/3 the job seeker knows the exact preference of the interviewer (when A1 is in charge), with probability 2/3 she can exclude one strategy and is uncertain about the rest two (when A2 or A3 are in charge). In contrast, if the job seeker chooses interview 2, with probability 1/3 she knows the preference of the interviewer (when B1 in charge), with probability 1/3 she needs to choose from A and H (when B2 is in charge), and with probability 1/3 she could not eliminate any interviewing strategy (when B3 is in charge). Ex ante, the only difference between the two interviews is that with probability 1/3, the number of possibly optimal interviewing strategies the job seeker is uncertain about, after making use of her knowledge, is larger in interview 2 than in interview 1. Thus, we say that interview 1 is **simpler** *for the job seeker and she will finally choose interview 1.*

We first overview the model for simplicity. Each *problem* is defined as a finite partition of the state space Ω. A block in the partition is called an *alternative*, which proves to be correct ex post if it contains the realized state. Then to solve the problem, the decision maker (DM) needs to guess which alternative is correct. For instance, in example 1, the problem is a partition with three blocks and each block represents the set of states where a particular strategy is optimal, that is, matches the preference of the interviewer. The primitive of the model is a complete and transitive preference relation $\succsim$ over the set of all problems, which denotes the agent's evaluation of relative simplicity for solving the problems.

We propose a model *preference for simplicity* to characterize $\succsim$ with a tuple $(\Sigma, \mu, \succsim^l)$. The σ-algebra Σ denotes the *knowledge algebra* of the DM and her *prior* μ is a complete, convex-ranged and countably additive probability measure on the measurable space (Ω, Σ)[1]. Also, $\succsim^l$ is called the *induced preference* over simple lotteries on positive integers. Then our representation follows a two-step approach: first, the decision maker maps each problem

[1] μ defined on (Ω, Σ) is said to be complete if whenever $A \subset B \in \Sigma$ with $\mu(B) = 0$, we have $\mu(A) = 0$. μ is convex-ranged if for any $a \in [0, 1]$ and any $B \in \Sigma$, there is $A \subset B, A \in \Sigma$ such that $\mu(A) = a\mu(B)$.

into a simple lottery over difficulty levels, which denote the number of potentially correct alternatives which the agent cannot exclude after utilizing the prior knowledge (Σ, μ); in the second step, she just ranks problems by ranking the induced lotteries using $\succsim^l$.

Now we briefly describe how the first step works. We assume that ex post, the DM can only discern *non-null* events, that is, measurable events with positive probability. Then at the ex ante stage, for a given problem, the DM could choose any partition of the state space consisting of non-null events such that she is informed of which block of the partition occurs after the state realizes. The partition is called the *information partition* of the DM. Intuitively, refining the information partition will weakly improve the accuracy of the information about the true state and help with the decision. Thus the DM will only choose *exhaustive* information partition where at the ex post stage, either the she can exactly pin down the correct alternative or she cannot conduct further refinement due to the constraint of her knowledge. In both cases, the agent utilizes all the prior knowledge she has to solve the problem. Then we can compute the *difficulty level* as the number of alternatives in the problem whose intersection with the realized information set is either non-null or non-measurable. If the difficulty level is k, then the DM can exclude all other alternatives and reduce the original problem into a k-choice problem. For example, $k = 1$ implies that the agent fully solves the problem by identifying the correct alternative. Given a problem with n blocks, the DM can compute the probability of difficulty level ranging from 1 to n. In this way, each problem is mapped to a simple lottery over positive integers. We show that this *induced lottery* is independent of the chosen exhaustive information partition.

To understand why in the second step the preference over problems $\succsim$ can be represented by the preference over induced lotteries $\succsim^l$, we need to show that two problems with the same induced lottery should be equally simple. Since the agent could rationally expect that she would exhaust her prior knowledge to solve for the problem she chooses, she should believe that there is no additional available information for her to make use of when the original problem reduces to a k-choice problem (i.e. difficulty level is k) ex post. Thus, any two reduced k-choice problems should be taken as identical in terms of simplicity, which justifies the second step. Moreover, it is intuitive that the higher the difficulty level is, the harder the reduced problem is as there are more potentially correct choices. Thus, the induced preference over lotteries should exhibit monotonicity with respect to first order stochastic dominance.

Our model provides predictions and explanations for choice behavior over problems given that the payoffs of solving all problems are identical, which is common in delegation problems where the agent's payoff is constant once a predetermined objective has been achieved. For instance, the mutual fund or pension fund manager might only need to guarantee that the return exceeds some threshold to receive full payment; the worker might only need to make sure the produced items satisfy the minimum quality requirement as her wage just depends on the number of qualified items; the international student only needs to pass one English proficiency test to meet the requirements of Princeton University ... Even when the realistic payoff structure is more complex, we claim that it will only affect the rewards associated with each problem and the agent's subjective assessment of its simplicity remains the same. In this way, our model of preference for simplicity still serves as a milestone in understanding choice behavior among more general problems.

The main result of the paper provides a sufficient and necessary condition for the preference over problems to be a preference for simplicity, i.e. there exists a complete, convex ranged probability space (Ω, Σ, μ) and a monotone preference $\succsim^l$ defined on simple lotteries over positive integers such that the agent evaluates the simplicity of each problem as if she transfers each problem into a lottery according to her knowledge prior (Ω, Σ, μ) and evaluates the lottery using $\succsim^l$. Under the framework of Savage (1954), we can uniquely pin down the probability measure (Σ, μ). This unique identification allows us to interpret (Σ, μ) as the agent's knowledge algebra and prior respectively. Here, we provide a simple illustration for the identification strategy. Suppose that event A is within the knowledge algebra of the agent. Consider the problem (A, A^c) which is a binary partition of the state space. Easy to see that the agent could correctly pin down which block the true state lies in when it is realized, and no other problem is strictly simpler than this one. Thus, whenever A is within the knowledge algebra, (A, A^c) is the simplest problem for the agent. It turns out that the inverse is also true. By this, we fully identify the knowledge algebra by the preference for simplicity. Also, routine arguments from the theory for qualitative probability ensure the uniqueness of the knowledge prior μ.

Since we impose no restrictions other than monotonicity on the induced preference $\succsim^l$, the preference for simplicity may not admit a cardinal representation. Thus, stronger axioms for continuity and independence are introduced to guarantee the existence of either a mixture continuity utility representation or a bounded expected utility representation. The latter representation can be interpreted as the ex-ante subjective assessment of the likelihood that the decision maker can solve the problem correctly. We also conduct comparative statics of the preference for simplicity by utilizing the separation between the perception of simplicity, i.e. the knowledge algebra and prior (Σ, μ), and the attitude towards simplicity, i.e. the induced preference $\succsim^l$. Roughly, the perception (Σ_1, μ_1) is defined to be *more accurate* than (Σ_2, μ_2) if for any problem, the induced lottery under perception (Σ_1, μ_1) is preferred to the induced lottery under perception (Σ_2, μ_2) by any attitude $\succsim^l$. We show that the comparison of accuracy can be characterized by the *extension*, an order defined over probability spaces in probability theory.

We close this section by reviewing some relevant literature. First, our model is built up based on the framework of Savage (1954) in the sense that the objectives are subjectively mapped into lotteries. Also, we follow Machina and Schmeidler (1992) to separate subjective probabilistic sophistication from expected utility hypothesis. However, the preference in our model is defined over problems (finite partitions of the state space) instead of Savage acts and there are no exogenous outcomes. The most related paper with us is Gul and Pesendorfer (2014). Both papers make use of non-measurable events under a similar framework. However, Gul and Pesendorfer (2014) uses non-measurability to capture ambiguity while we use the number of non-measurable blocks to capture the difficulty of the problem. In section 6, we will provide a more detailed comparison between our model and the ambiguity literature.

The term "simplicity" is also discussed in other literature such as mechanism design and incomplete contracts. For instance, in mechanism design, Li (2017) considers the mechanism design problem when the agent is lack of contingent thinking, and Borgers and Li (2017) considers simple mechanisms where the agents need not to specify higher order beliefs. In contrast, our model aims to model agent's subjective evaluation of simplicity and keeps silent of how the agent would simplify the problem heuristically. By comparison, Hart and Moore (1999) and Segal (1999) try to provide a foundation for incomplete contracts and explain why simple contracts are prevalent in the real world. Their environments are different from ours and we will briefly discuss how the preference for simplicity might help to understand why agents are biased towards simple contracts.

Another strand of relevant literature models the cost of thinking. Ortoleva (2013) considers an increasing number of alternatives decrease the attractiveness of a menu due to the mental cost. Ergin and Sarver (2010) models the cost of contemplation for decision making. Similarly, we model an agent who needs to collect information before solving the problem. However, it is assumed in our model that there is no cost for information acquisition and the only constraint is her knowledge. Moreover, at the ex post stage, we need to point out that the number of choices does not characterize the agent's mental cost or thinking cost. It should be clear that with no additional available information, the agent makes a choice by guessing the right answer. Thus, no mental cost should occur at that stage. Instead, the number of choices should be interpreted as the agent's subjective evaluation of simplicity.

The outline of the paper is as follows. In section 2, we provide model setup and basic representation. Section 3 contains the axioms and the main characterization result. Section 4 extends the characterization with stronger axioms, and section 5 contains some comparative statics of our model. A discussion of the model is in section 6 and we conclude in section 7.

2 Basic Representation

2.1 Formal Setup

Consider a nonempty set Ω as the state space. A *problem* is a partition of the state space $\hat{P} = \{P_n\}_{n=1}^{\infty}$ such that all sets are mutually disjoint and there is some k with $\forall n \geq k, P_n = \emptyset$ and $\cup_{n=1}^{\infty} P_n = \Omega$. When there is no confusion, we write $\hat{P} = \{P_n\}_{n=1}^{k}$ or $\hat{P} = (P_n)_{n=1}^{k}$, and we may also use Ω to represent the problem $\{\Omega\}$. Denote $\mathbb{P}(\Omega)$ as the set of all problems and $\hat{P}, \hat{Q}, \hat{R}$ as generic elements in $\mathbb{P}(\Omega)$. A *problem* is a finite partition of the state space and we allow the empty set to be an element of the partition just for notational and illustrative convenience. The primitive of our analysis is a complete and transitive preference relation $\succsim$ on $\mathbb{P}(\Omega)$, which represents the decision maker's assessment of the simplicity of all problems. Moreover, a problem $\hat{P}$ is *binary* if $\hat{P} = (A, A^c)$ for some $A \subseteq \Omega$. We also call (A, A^c) as the binary problem of set A.

Intuitively, a problem can be interpreted as a choice problem which asks the decision maker in which block of the partition the true state lies. Multiple choice questions and true-or-false questions in homework or examinations are the most straightforward examples. The decision faced by the interviewee in Section 1 is also a problem as she needs to choose the correct interview strategy (the block containing the true state) after observing who the interviewer is. More generally, a project with different available actions can be regarded as a problem where blocks in the partition are the events in which a particular action is optimal.

2.2 Basic Representation

In this section, we will introduce our basic representation of the binary relation $\succsim$ on the set of problems $\mathbb{P}(\Omega)$– preference for simplicity.

The preference for simplicity is characterized by a tuple $(\Sigma, \mu, \succsim^l)$. Σ is a σ-algebra defined on Ω which we call the *knowledge algebra* of the DM. The *prior* of the DM μ is a complete, convex-ranged and countably additive probability measure on (Ω, Σ). $\succsim^l$ is the *induced preference* over simple lotteries on positive integers. Similar to (subjective) expected utility theory (EU) (Savage, 1954) and expected uncertainty utility theory (EUU) (Gul and Pesendorfer, 2014), the DM adopts a two-step procedure for evaluation by first mapping each object to a lottery and then evaluating the lottery using induced preference. However, the objects in our analysis are problems (partitions) instead of Savage acts and there are no exogenous outcomes so that it is not straightforward where the lotteries should be defined on, which distinguishes our model from the literature. To overcome this obstacle, we introduce the *difficulty level* as an endogenous outcome, which can be interpreted as the number of potentially correct alternatives which the agent cannot exclude after utilizing the prior knowledge.

First we describe how the DM makes use of her prior knowledge. A set $A \subset \Omega$ is *non-null* if $A \in \Sigma$ and $\mu(A) > 0$. By comparison, A is *null* if $A \in \Sigma$ and $\mu(A) = 0$. We assume that the DM can only discern non-null events as she cannot identify a set outside her knowledge algebra and she thinks the null sets will not occur. Thus, at the ex ante stage, for a given problem, the DM could choose any partition of the state space consisting of non-null events such that she is informed of which block of the partition occurs after the state realizes. Such a partition is called the *information partition* of the DM, denoted by $\mathcal{I} = \{I_m\}_{m=1}^{\infty}$.[2] Each block in the information partition is called an *information set* and specifically, we denote $I(w) \in \mathcal{I}$ as the information set that includes w. Thus, when state w realizes, the DM is informed that $I(w)$ occurs.

Intuitively, refining the information partition will weakly improve the accuracy of the information about the true state and thus make the DM weakly better off. Thus, she will only choose information partition such that refinement cannot help with collection of alternatives. For each problem $\hat{P} = (P_1, ..., P_K) \in \mathbb{P}(\Omega)$ and each $n = 1, ..., K$, denote

[2] An information partition has at most countable number of blocks as the probability measure is finite and each block has strictly positive measure.

$\mathcal{D}_n(\hat{P})$ as the collection of sets derived by the union of n nonempty alternatives in $\hat{P}$, i.e. $A \in \mathcal{D}_n(\hat{P})$ if and only if $A = \cup_{i=1}^{n} P_{k_i}$ for some $\{k_1, ..., k_n\} \subset \{1, 2, ..., K\}$.Then we call information partition $\mathcal{I}(\hat{P})$ is *exhaustive* for problem $\hat{P}$ if

$$\forall A \in \cup_{n=1}^{K} \mathcal{D}_n(\hat{P}), \quad \forall I_k \in \mathcal{I}(\hat{P}), \quad \mu_*(A \cap I_k) = 0 \text{ or } \mu(I_k)$$

where $\mu_*(B)$ denotes the *inner measure* of B: $\mu_*(B) = \sup_{E \subseteq B: E \in \Sigma} \mu(E)$. For $A = \cup_{i=1}^{n} P_{k_i} \in \mathcal{D}_n(\hat{P})$, if $\mu_*(A \cap I_k) = 0$, then there is no non-null event in $A \cap I_k$ and the DM cannot further refine within this event due to the constraint of her knowledge, although this could be beneficial. If instead, $\mu_*(A \cap I_k) = \mu(I_k)$, then I_k is contained in A except for a measure 0 set (null set) and thus the DM can essentially know whether the correct alternative is among $\{P_{k_i}\}_{i=1}^{n}$ or not. We require the expression to hold for all sets in $\cup_{n=1}^{K} \mathcal{D}_n(\hat{P})$, which means the agent utilizes all the prior knowledge she has to discern whether the correct alternative is among any collection of alternatives in the problem. That explains why the DM would choose exhaustive information partition.

Now we can formally define the difficulty level. Given a problem $\hat{P} = (P_1, ..., P_K)$ and an exhaustive information partition $\mathcal{I}(\hat{P})$, the *difficulty level* at state w is

$$K_{\hat{P}, \mathcal{I}(\hat{P})}(w) := |\{P_n | I(w) \cap P_n \notin \mathcal{N}\}| \in \{1, ..., K\}$$

where $\mathcal{N} = \{A \in \Sigma : \mu(A) = 0\}$ is the set of null events. That is, after the DM is informed of $I(w)$, she cannot exclude alternative P_n if the intersection of the information set and the alternative is either non-measurable or non-null. In other words, if the difficulty level at w is k, then the DM can reduce the original problem into a k-alternative problem after w is realized. For example, $k = 1$ implies that the agent fully solves the problem by identifying the correct alternative. It is clear that the set of difficulty levels is exactly the set of positive integers $\mathbb{N} := \{1, 2, 3, ...\}$. Since all prior knowledge has been utilized, different reduced problems with the same number of remaining alternatives should be perceived as equally simple, which makes the difficulty level an appropriate measure of the endogenous outcome in terms of simplicity at each state.

After utilizing the information, each problem is mapped to a lottery over difficulty levels. Denote $\mathcal{L}(\mathbb{N})$ as the space of simple lotteries over $\mathbb{N}$ where $\hat{p}, \hat{q}, \hat{r}$ are generic elements. For given problem $\hat{P} = (P_1, ..., P_K)$ and exhaustive information partition $\mathcal{I}(\hat{P})$, we denote $\hat{p}(\mathcal{I}(\hat{P})) = (p_1, ..., p_K) \in \mathcal{L}(\mathbb{N})$ as the induced lottery, where p_n is the probability of difficulty level n:

$$p_n := \mu(\{w \in \Omega : K_{\hat{P},\mathcal{I}(\hat{P})}(w) = n\}), \forall n \in \mathbb{N}$$

To see why p_n is well-defined, notice that states in the same information set share the same difficulty level, which implies that the set of states with difficulty level n is a finite or countable union of information sets and is thus measurable.

By definition, the difficulty level of each state should depend on the chosen exhaustive information partition. However, we will show that the induced lottery over difficulty levels remains the same. Before stating the result, we need another notation for illustrative simplicity. For any $A = \cup_{i=1}^{n} P_{k_i} \in \mathcal{D}_n(\hat{P})$, we denote $A^i = A \backslash P_{k_i}$ for $i = 1, ..., n$ as the union of some collection of $n-1$ alternatives contained in A. The following Lemma 1 establishes the independence of the induced lottery from exhaustive information partitions by providing its explicit formulas.

Lemma 1. [3] *Given a problem $\hat{P} = (P_1, ..., P_K)$, there exists $\hat{p} \in \mathcal{L}(\mathbb{N})$ such that for any exhaustive information partition $\mathcal{I}(\hat{P})$, the induced lottery $\hat{p}(\mathcal{I}(\hat{P})) = \hat{p}$. Moreover, for any $n > K$, $p_n = 0$ and for any $n \leq K$,*

$$p_n = \sum_{A \in \mathcal{D}_n(\hat{P})} \left(\mu_*(A) - \sup_{\forall i, E^i \subseteq A^i : E^i \in \Sigma} \mu(\cup_{i=1}^{n} E^i) \right) \tag{1}$$

A direct corollary of this lemma is that there exists a well-defined operator mapping each problem into a simple lottery on difficulty levels and we denote it as $\Gamma_\mu : \mathbb{P}(\Omega) \to \mathcal{L}(\mathbb{N})$ such that $\Gamma_\mu(\hat{P}) = \hat{p}$ where $\hat{p}$ is defined in Lemma 1.

To see why p_n given in equation (1) denotes the probability of difficulty level n for all n, $p_1 = \sum_{k=1}^{K} \mu_*(P_k)$ is the sum of inner measures of all blocks in the partition and thus means the maximum probability that the DM can pin down the exact block where the true state lies. For $1 < n \leq K$, $A \in \mathcal{D}_n(\hat{P})$ is the union of n different nonempty alternatives $\{P_{k_i}\}_{i=1}^{n}$, which implies that $\mu_*(A)$ denotes the probability that the DM knows that the correct alternative is among $\{P_{k_i}\}_{i=1}^{n}$. However, this does not necessarily mean that the states in A have difficulty level n as the DM might be able to exclude more alternatives. In this sense, to get probability of difficulty level n associated with alternatives $\{P_{k_i}\}_{i=1}^{n}$, we need to subtract the probability that the the DM can exclude at least one more alternative in $\{P_{k_i}\}_{i=1}^{n}$, which is exactly $\sup_{\forall i, E^i \subseteq A^i : E^i \in \Sigma} \mu(\cup_{i=1}^{n} E^i)$. Summing over all collections of n

[3] All proofs are in Appendix B

alternatives, we can derive the probability of difficulty level n as is given in equation (1). For $n > K$, as there are only K nonempty alternatives in the problem, the difficulty level will never reach n and thus the corresponding probability is 0.

By transforming the problems into simple lotteries over difficulty levels, we use $\succsim^l$ to characterize the decision maker's subjective ranking over induced lotteries.

Not every preference over $\mathcal{L}(\mathbb{N})$ is consistent with some preference over problems. Recall that the difficulty level n denotes the number of alternatives that the DM is uncertain about after fully utilizing the knowledge. Intuitively, the larger n is, the more difficult the problem is perceived by the DM. This natural ranking in difficulty levels further implies a monotonicity requirement for the induced preference on lotteries. In words, a lottery with more weights on lower difficulty levels should be simpler and thus preferred.

Formally, for any n, define $F_n(\hat{p})$ as $F_n(\hat{p}) = \sum_{i=1}^{n} p_i$ which is the cumulative function. Consider a preference $\succsim^l$ defined on $\mathcal{L}(\mathbb{N})$, we say $\succsim^l$ is *monotone*[4] if for any $\hat{p}, \hat{q}$ we have

$$\forall n, F_n(\hat{p}) \geq F_n(\hat{q}) \Rightarrow \hat{p} \succsim^l \hat{q}$$

and

$$\forall n, F_n(\hat{p}) \geq F_n(\hat{q});\ \exists k, F_k(\hat{p}) > F_k(\hat{q}) \Rightarrow \hat{p} \succ^l \hat{q}$$

Then we define the preference of simplicity as our basic representation.

Definition 1. *For given Ω, a preference $\succsim$ over $\mathbb{P}(\Omega)$ is a* ***preference for simplicity*** *if there is a σ-algebra Σ, a prior μ over (Ω, Σ) and a monotone preference $\succsim^l$ defined over $\mathcal{L}(\mathbb{N})$ such that*

$$\hat{P} \succsim \hat{Q} \Leftrightarrow \Gamma_\mu(\hat{P}) \succsim^l \Gamma_\mu(\hat{Q})$$

Besides monotonicity, we do not impose any further requirement for the induced preference over lotteries in our representation. However, monotonicity can be quite strong if we restrict our attention to binary problems. Actually, for binary problems, only probability of difficulty level 1 matters and monotonicity requires that the $(A, A^c) \succsim (B, B^c)$ if and only if $\mu_*(A) + \mu^*(A^c) \geq \mu_*(B) + \mu^*(B^c)$.

[4]Note that if we rank the problems according to their difficulty (instead of simplicity), then the monotonicity requirement is exactly respecting first order stochastic dominance (FOSD). Or we can say the preference for simplicity respects **reverse FOSD**.

3 Axioms and Representation Theorem

In this section, we will characterize the preference for simplicity representation by axioms. First, Ω should the simplest problem as Ω is trivially correct no matter which state realizes. We state it as Axiom 0.

Axiom 0. (Simplest Problem) $\Omega \succsim \hat{P}$ for all $\hat{P} \in \mathbb{P}(\Omega)$.

Ω is not the unique simplest problem. Recall that (A, A^c) is the binary problem of set A. We call an event A *trivial* if its binary problem is the simplest, that is, $(A, A^c) \sim \Omega$. Denote $\mathcal{T}$ as the set of all trivial events. If A is trivial, (A, A^c) is as simple as the simplest problem Ω. This implies that whenever a state w realizes, the DM knows exactly whether A or A^c is true. Actually we will later see that an event is trivial if and only if it is within the knowledge algebra of the DM.

Axiom 1 is standard as it requires the binary relation over problems to be a weak order.

Axiom 1. (Weak Order) $\succsim$ is complete and transitive.

In contrast with trivial events, there are also binary problems induced by events which are the most difficult among all binary problems. We define $D \subseteq \Omega$ as an *intractable event* if for any $(A, A^c) \in \mathbb{P}(\Omega)$, we have $(A, A^c) \succsim (D, D^c)$. Let $\mathcal{D}$ be the set of intractable events. The following axiom asserts the existence of intractable events and strict comparisons.

Axiom 2. (Strictness) $\mathcal{D}$ is not empty and $\Omega \succ (D, D^c)$ for $D \in \mathcal{D}$.

An event $N \subseteq \Omega$ is called a *null event* if $(N, N^c) \sim \Omega$, and $(N \cup D, N^c \cap D^c) \sim (D, D^c)$ for some $D \in \mathcal{D}$. Let $\mathcal{N}$ be the set of all null events and N, N' be the generic elements. Intuitively, N is a null event if the binary problem of itself is the simplest, but the binary problem of its union with some intractable event becomes the most difficult binary problem. This implies that N matters little for the simplicity of the problem. Actually it turns out that a null event is exactly a set with 0 probability in the DM's prior. Let $\mathcal{T}_+ \subset \mathcal{T}$ denote the set of non-null trivial events. Similar to Gul and Pesendorfer (2014), we assume that the agents will use elements in $\mathcal{T}_+$ to quantify the likelihood of an event.

The following axiom states that a modification for a problem by gathering more states from a trivial event into the same alternative of the new partition will weakly simplify the problem.

Axiom 3. (Triviality Improvement) $B \in \mathcal{T}$ and $\hat{P} = (P_1, ..., P_n)$, $\hat{Q} = (P_1 \cup (B \cap P_2), P_2 \backslash B, P_3, ..., P_n) \Rightarrow \hat{Q} \succsim \hat{P}$.

Briefly, problem $\hat{Q}$ is derived from $\hat{P}$ by moving the intersection of B and P_2 to the alternative P_1. To see why this can simplify the problem, we take the binary problem (A, A^c) as an example. Suppose B is trivial, then in the new problem $(A \cup B, (A \cup B)^c)$, when a state in B is realized, the DM knows for sure that the option $A \cup B$ is correct, while in the problem (A, A^c) she might still be uncertain about the two options. This implies that the set of states with difficulty level 1 of $(A \cup B, (A \cup B)^c)$ is a superset of that of (A, A^c) and thus $(A \cup B, (A \cup B)^c)$ would be simpler and more preferred than (A, A^c). For more general problem, the same intuition applies as the concentration of states within $B \in \mathcal{T}$ in the same block of the new problem will expand the set of states with lower difficulty levels and shrink the set of states with higher difficulty levels, which simplifies the original problem.

The Axiom 4 below, is similar with Axiom P4* in Machina and Schmeidler (1992), which strengthens the weak comparative probability axiom and guarantees probabilistic sophistication in the absence of sure thing principle. Recall the intuition in Machina and Schmeidler (1992): whenever a probabilistically sophisticated individual exhibits the ranking revealing that A is at least as likely as B in the case where the relevant exogenous outcomes are $x^* \succ x$ and the complementary event $(A \cup B)^c$ yields the subact g, then she must continue to reveal that A is at least as likely as B given other outcomes $y^* \succ y$ and subact h on $(A \cup B)^c$. The idea also applies here, except for the fact that the objectives in our analysis are partitions instead of acts and the outcomes are endogenous.

We call an event $V \subseteq \Omega$ **unverifiable** if $B \subseteq V, B \in \mathcal{T} \Rightarrow B \in \mathcal{N}$. Denote the set of unverifiable events as $\mathcal{V}$. Equivalently, V is unverifiable if and only if it does not contain any non-null trivial event. However, unlike null events, the finite union of unverifiable events might be in $\mathcal{T}_+$. This leads to the following important notion:

Definition 2. *A partition $\{A_i\}_{i=1}^n$ is said to be a* ***nowhere verifiable n-partition*** *of $A \in \mathcal{T}$ if for any proper subcollection $\{A_{k_i}\}_{i=1}^m$ of $\{A_i\}_{i=1}^n$, $m < n$, we have $\cup_{i=1}^m A_{k_i} \in \mathcal{V}$.*

Here are several quick remarks related to the definition. First, $\{A\}$ is a nowhere verifiable 1-partition of A. For $\emptyset$, it could have a nowhere verifiable k-partition for arbitrary large k since $\emptyset = \cup_{i=1}^k A_k$ with $A_k = \emptyset$. Second, suppose $A \in \mathcal{T}$ admits a nowhere verifiable n-partition which includes an emptyset, then A must be unverifiable and thus null. In this way, the partition restricted to A will not affect the simplicity of the overall problem. Finally, consider a problem $\hat{P} = \{P_1, ...P_n\}$ such that $\{P_1, ...P_k\}, k \leq n$ forms a nowhere verifiable n-partition for $A \in \mathcal{T}_+$, then the problem restricted to A is exactly the partition $\{P_1, ...P_k\}$.

To see why this definition is important, suppose that $A \in \mathcal{T}_+$ admits a nowhere verifiable n-partition $\{A_i\}_{i=1}^n$, we claim that states in A have difficulty level n. Briefly, $A \in \mathcal{T}_+$ implies that A has positive probability in DM's prior. Then for any exhaustive information partition $\mathcal{I}(\hat{P})$, it is without loss of generality that $I(w) \subseteq A$ for all $w \in A$. As the union of any proper subcollection of $\{A_i\}_{i=1}^n$ is unverifiable, it does not contain any non-null set, which implies that for any $w \in A$, $I(w)$ is not essentially contained in the union any $n-1$ alternatives A^i, $i = 1, \ldots, n$. Thus, the difficulty level at w is n. Another way to see this is by observing that $\mu(A)$ is contained in the probability of difficulty level n in equation (1).

For positive integers m and n, we denote $\hat{P} \ =^{m,n} \ (A, B, C)$ if $\{A, B, C\}$ is a partition of Ω with $A, B, C \in \mathcal{T}$, $\{P_1, ..., P_m\}$ is a nowhere verifiable m-partition of A and $\{P_{m+1}, ..., P_{m+n}\}$ is a nowhere verifiable n-partition of B. Recall the interpretation for the nowhere verifiable partition, $\hat{P} =^{m,n} (A, B, C)$ means that the difficulty level on set A is m and the difficulty level on set B is n. We also write $\hat{P} \approx^n \hat{Q}$ if for $k \geq 1$, $P_{n+k} = Q_{n+k}$, that is, problems $\hat{P}$ and $\hat{Q}$ agree on alternatives with subscript larger than n, whose union is set C. Then the Axiom 4 is stated as follows:

Axiom 4. (Monotone Independence) Take $m > n, v > t$, suppose that

$$\hat{P} =^{m,n} (A_1, B_1, C), \hat{R} =^{v,t} (A_1, B_1, C), \hat{Q} =^{m,n} (A_2, B_2, C), \hat{S} =^{v,t} (A_2, B_2, C)$$

and

$$\hat{P} \approx^{m+n} \hat{Q}, \hat{R} \approx^{v+t} \hat{S}$$

then we have $\hat{P} \succ \hat{Q} \Rightarrow \hat{R} \succ \hat{S}$.

Intuitively, Axiom 4 can be divided into two parts. The first part is independence. Notice that $\hat{P}$ and $\hat{Q}$ only differ in the $m+n$ alternatives contained in C^c, independence implies that only those alternatives matter. Similar arguments hold for $\hat{R}$ and $\hat{S}$. The second part is standard comparison of probability. On A_1 , $\hat{P}$ and $\hat{R}$ have difficulty level m while on B_1 they have difficulty level n. Similarly, on A_2, $\hat{Q}$ and $\hat{S}$ have difficulty level w while on B_2 they have difficulty level t. Since $m > n$ and $w > t$, the problem reduced to A_1 (A_2) is harder than that reduced to B_1 (B_2). Then $\hat{P} \succ \hat{Q}$ and $\hat{R} \succ \hat{S}$ both mean that the DM thinks B_1 is more likely than B_2. Thus Axiom 4 states that the DM has a consistent subjective measure of probability.

Two more technical axioms are necessary to get our representation. Axiom 5 is a weaker version of the axiom for small event continuity in Savage theory, as it solely applies to binary problems, and it serves to guarantee the the existence of a convex-valued probability measure. It states that for binary problems, the state space Ω can be partitioned into small enough trivial events such that improving the worse binary problem with just one of these events in the form of Axiom 3 is not enough to reverse the original ranking

Axiom 5. (Triviality Continuity) If $(A, A^c) \succ (B, B^c)$, then there is a finite partition $\{F_n\}_{n=1}^N$ of Ω such that $F_n \in \mathcal{T}$ and $(A, A^c) \succ (B \cup F_n, B^c \cap F_n^c)$ for each n.

Axiom 6 below imposes another technical requirement on continuity based on convergence of sets. For an increasing sequence of events A_n, we will write $\lim_n A_n$ to represent the limit event $\cup_n A_n$. Axiom 6 states that if the binary problem of the limit event is strictly better (or worse) than some other binary problem, then there exists an event in the sequence whose binary problem maintains this strict comparison. It is used to prove the closure under countable union of the σ-algebra and countable additivity of the probability measure.

Axiom 6. (Dominance Continuity) Suppose that $A_n \in \mathcal{T}$ for $n \geq 1$, and $A_{n-1} \subseteq A_n$, for $E, B \subseteq \Omega$, then

(1) $(B, B^c) \succ (E \cup \lim_n A_n, E^c \cap \lim_n A_n^c) \Rightarrow (B, B^c) \succ (A_k \cup E, A_k^c \cap E^c)$ for some k;

(2) $(E \cup \lim_n A_n, E^c \cap \lim_n A_n^c) \succ (B, B^c) \Rightarrow (A_k \cup E, A_k^c \cap E^c) \succ (B, B^c)$ for some k.

Then our main result is stated as Theorem 1, which establishes the equivalence of the above seven axioms to the existence and uniqueness of the preference for simplicity representation.

Theorem 1. $\succsim$ *over* $\mathbb{P}(\Omega)$ *satisfies Axioms 0-6 if and only if it is a preference for simplicity. Moreover, the representation* $(\Sigma, \mu, \succsim^l)$ *is unique.*

We end this section by outlining the intuition of the proof. The detailed proof can be found in the Appendix B.

For sufficiency, given a preference for simplicity representation $(\Sigma, \mu, \succsim^l)$, we first characterize trivial, intractable, null and unverifiable events with the prior knowledge. Specifically, an event is trivial if and only if it is measurable and it is null if further the probability measure is 0. By comparison, an unverifiable event is one with inner measure 0 and an intractable event is diffuse in Ω, that is, both the event and its complement have inner measure 0. Then we call $\hat{Q}$ as a *refinement* of $\hat{P}$ if for any $Q \in \hat{Q}$, there exists some $P \in \hat{P}$ such that $Q \subseteq P$. It is clear that a refinement of some problem always makes it weakly harder as more alternatives are introduced. However, there exists a type of refinement which maintains the simplicity of the original problem. For any $\hat{P} = (P_1, ..., P_n)$, $B \in \mathcal{T}$, we define the *refined problem induced by* B as $\Phi(\hat{P}, B) = (P_1 \backslash B, ..., P_n \backslash B, B \cap P_1, ..., B \cap P_n)$ and show that $\hat{P} \sim \Phi(\hat{P}, B)$. With those properties, we can directly check that all axioms hold.

For necessity, the first step is to identify the knowledge algebra Σ and the prior μ. Concretely, Σ is defined to be the set of all trivial events $\mathcal{T}$, which can be shown to be a σ-algebra. Then we define a binary relation $\succsim^*$ over Σ as

$$A \succsim^* B \Leftrightarrow (A \cup D, A^c \cap D^c) \succsim (B \cup D, B^c \cap D^c), \forall D \in \mathcal{D}$$

We then show that $\succsim^*$ is a qualitative probability that satisfies all the conditions required for the existence of a unique convex ranged finitely additive probability measure μ defined over Σ which represents $\succsim^*$. Also μ is countably additive and complete, and thus a prior. For the second step, we show that the preference restricted to binary problems (A, A^c) can be represented by the sum of inner measures of A and A^c (that is, the probability of difficulty level 1). Finally we extend the result to general problems and show that, by transforming each problem into its equivalent fundamental representation, the preference over induced lotteries of difficulty levels is enough to represent the preference over problems.

The uniqueness of the knowledge algebra $\Sigma = \mathcal{T}$ comes from the definition of trivial events: $A \in \mathcal{T} \Leftrightarrow (A, A^c) \sim \{\Omega\}$. The prior μ is unique as a direct implication of the the representation theorem for qualitative probability for $\succsim^*$. Finally, as Γ_μ is onto, that is, any simple lottery on difficulty levels can be induced by some problem, for any $\hat{p}, \hat{q} \in \mathcal{L}(\mathbb{N})$, we can find $\hat{P}, \hat{Q}$ such that $\Gamma_\mu(\hat{P}) = p, \Gamma_\mu(\hat{Q}) = q$ and $p \succsim^l q \Leftrightarrow \hat{P} \succsim \hat{Q}$. This guarantees the uniqueness of the induced preference $\succsim^l$.

4 Extensions on Representation Theorem

4.1 Mixture Continuous Utility Representation

Recall that in the definition of preference for simplicity, we just require that the induced preference over lotteries $\succsim^l$ to be monotone and thus $\succsim$ does not necessarily admits a cardinal representation. In this section, we will impose a stronger continuity axiom to guarantee a cardinal representation for $\succsim$.

Following the conventional notion of continuity for expected utility preferences over probability distributions, we introduce mixture continuity for preferences and functions as follows. With the standard mixture operator over lotteries $h_\lambda(\hat{p},\hat{q}) = \lambda\hat{p} + (1-\lambda)\hat{q}$ such that the n-th element of $h_\lambda(\hat{p},\hat{q})$ is $\lambda p_n + (1-\lambda)q_n$, we have:

Definition 3. *A preference $\succsim^l$ on $\mathcal{L}(\mathbb{N})$ is said to be **mixture continuous** if for any $\hat{p},\hat{q},\hat{r} \in \mathcal{L}(\mathbb{N})$, the sets $\{\lambda \in [0,1] : \lambda\hat{p} + (1-\lambda)\hat{q} \succsim^l \hat{r}\}$ and $\{\lambda \in [0,1] : \hat{r} \succsim^l \lambda\hat{p} + (1-\lambda)\hat{q}\}$ are closed.*

Definition 4. *A function $U(\cdot) : \mathcal{L}(\mathbb{N}) \to \mathbb{R}$ is said to be **mixture continuous** if for any $\hat{p},\hat{q},\hat{r} \in \mathcal{L}(\mathbb{N})$, the sets $\{\lambda \in [0,1] : U(\lambda\hat{p} + (1-\lambda)\hat{q}) \geq U(\hat{r})\}$ and $\{\lambda \in [0,1] : U(\lambda\hat{p} + (1-\lambda)\hat{q}) \leq U(\hat{r})\}$ are closed.*

We also denote that a function respects monotonicity if it can represent some monotone preference over lotteries. Formally,

Definition 5. *A function $U(\cdot) : \mathcal{L}(\mathbb{N}) \to \mathbb{R}$ **respects monotonicity** if for any $p, q \in \mathcal{L}(\mathbb{N})$, $F_n(p) \geq F_n(q), \forall n \Rightarrow U(p) \geq U(q)$ and $F_n(p) \geq F_n(q), \forall n, F_k(p) > F_k(q), \exists k \Rightarrow U(p) > U(q)$.*

Then we can define the mixture continuous utility representation for the preference for simplicity. Briefly, it means that the induced preference over lotteries $\succsim^l$ can be represented by a mixture continuous utility function.

Definition 6. *The preference $\succsim$ admits a **mixture continuous utility representation** if there exists (Σ, μ, U) such that i) μ is a prior over (Ω, Σ); ii) $U : \mathcal{L}(\mathbb{N}) \to \mathbb{R}$ is non-constant, mixture continuous and respects monotonicity; and iii) $\succsim$ is represented by $V : \mathbb{P}(\Omega) \to \mathbb{R}$ where $V(\hat{P}) = U(\Gamma_\mu(\hat{P}))$.*

Our goal is to find a mixture continuous utility representation for $\succsim$. As was shown in Machina and Schmeidler (1992), the key to guarantee mixture continuity in Savage expected utility theory is the axiom called small event continuity. Recall that Axiom 5 only applies to binary problems, [5] we need to extend its idea to general problems to derive a counterpart

[5] That explains why in Theorem 1, if we restrict our attention to binary problems, the preference over binary problems does admit a continuous cardinal representation $U(\{A, A^c\}) = \mu_*(A) + \mu_*(A^c)$, which is the probability of difficulty level 1 of binary problem $\{A, A^c\}$.

of small event continuity axiom in our setup.

Axiom 5^***. (Small Event Continuity)** For any problems $\hat{P} = (P_1, ..., P_n)$, $\hat{Q} = (Q_1, ..., Q_m)$, $\hat{R} = (R_1, ..., R_l) \in \mathbb{P}(\Omega)$ such that $\hat{P} \succ \hat{Q}$, there exists a finite partition $\{A_1, ..., A_s\}$ of Ω with $A_i \in \mathcal{T}$ for all i such that $\hat{P}^i \succ \hat{Q}$ and $\hat{P} \succ \hat{Q}^i$ where, for each $i = 1, 2, ..., s$,

$$\hat{P}^i = (P_1 \backslash A_i, ..., P_n \backslash A_i, R_1 \cap A_i, ..., R_l \cap A_i)$$

$$\hat{Q}^i = (Q_1 \backslash A_i, ..., Q_m \backslash A_i, R_1 \cap A_i, ..., R_l \cap A_i)$$

Recall that in the intuition of proof for Theorem 1, we showed that the for any $\hat{P}$, the refined problem induced by A_i $\Phi(\hat{P}, A_i) = (P_1 \backslash A_i, ..., P_n \backslash A_i, P_1 \cap A_i, ..., P_n \cap A_i)$ is indifferent with $\hat{P}$. We denote the partition $(P_1 \cap A_i, ..., P_n \cap A_i)$ as the *reduced problem of* $\Phi(\hat{P}, A_i)$ *restricted to* A_i. Then $\hat{P}^i$ is simply derived from $\Phi(\hat{P}, A_i)$ by replacing the reduced problem restricted to A_i with $\Phi(\hat{R}, A_i)$. Similarly, $\hat{Q}^i$ is derived from $\Phi(\hat{Q}, A_i)$ by replacing the reduced problem restricted to A_i with $\Phi(\hat{R}, A_i)$. The Axiom 5^* simply states that, if the likelihood of A_i is low enough, then changing the reduced problem (and thus the difficulty level) on it will not reverse the original strict ranking.

To briefly see why Axiom 5^* is stronger than Axiom 5 given that other axioms hold, we can consider the special case with $n = m = 2$, $\hat{R} = \Omega$ and refer to the observation that the reduced problem of $(Q_1 \cup A_i, Q_1^c \cap A_i^c)$ induced by A_i is indifferent to $(Q_1 \cup A_i, Q_1^c \cap A_i^c)$. Then the following theorem shows that strengthening Axiom 5 to Axiom 5^* is sufficient to guarantee a mixture continuous utility representation of $\succsim$.

Theorem 2. $\succsim$ *over* $\mathbb{P}(\Omega)$ *satisfies Axioms 0-4, Axiom* 5^* *and Axiom 6 if and only if* $\succsim$ *admits a.mixture continuous utility representation* (Σ, μ, U)*. Moreover,* (Σ, μ) *is unique and* U *is unique up to monotone transformations.*

4.2 Bounded EU Representation

Suppose that $\succsim$ admits a mixture continuous utility representation (Σ, μ, U). Denote $\delta^n \in \mathcal{L}(\mathbb{N})$ as the lottery that assigns probability 1 to difficulty level n. Then the utility of a problem with the induced lottery δ^n is given by $U(\delta^n)$, which represents the DM's subjective evaluation of difficulty level n. Consider a general problem $\hat{P}$ which induces lottery $\hat{p}$, then a natural question is when its utility equals the expected utility over difficulty levels, that is, $U(\hat{p}) = \sum_{n=1}^{\infty} p_n U(\delta^n)$.

Formally, suppose that the utility function $V : \mathbb{P}(\Omega) \to \mathbb{R}$ represents $\succsim$, then V is called an *expected utility (EU)* representation for $\succsim$ if $V(\hat{P}) = \sum_{n=1}^{\infty} p_n\theta_n$ where $\theta_n > \theta_{n'}$ if $n < n'$ and $\hat{p} = \Gamma_\mu(\hat{P}) = (p_1, p_2, \cdots)$. The decreasing order of coefficients $\{\theta_n\}_{n=1}^{\infty}$ reflects the monotonicity of $\succsim^l$.

One intuitive interpretation of the EU representation is that $V(\hat{P})$ represents the DM's subjective probability of correctly solving the problem. For any given problem $\hat{P} = (P_1, ..., P_K)$, the DM first transforms $\hat{P}$ into K reduced problems. Each reduced problem is indexed by its difficulty level $n = 1, ..., K$ and has probability p_n where $\hat{p} = \Gamma_\mu(\hat{P})$. In other words, for the reduced problem with difficulty level n, the DM can restrict the correct choice to n alternatives abut knows nothing more. Denote $\theta_n \in [0, 1]$ as the likelihood that the DM thinks she can correctly solve this reduced problem. Then by Bayes' rule, the probability of answering $\hat{P}$ correctly is exactly the weighed average of the probabilities to answer each reduced problem correctly, that is, $\sum_{n=1}^{\infty} p_n\theta_n$. Notice that the requirement $\theta_n \in [0, 1]$ for all n is not without loss of generality as it requires the utility function to uniformly bounded. This leads to the following stronger notion of representation.

Definition 7. *The preference* $\succsim$ *admits a* ***bounded EU representation*** *if there exists* $(\Sigma, \mu, \{\theta_n\}_{n=1}^{\infty})$ *such that i)* μ *is a prior over* (Ω, Σ)*; ii)* $\{\theta_n\}_{n=1}^{\infty}$ *is a strictly decreasing sequence in* $[0, 1]$ *and* $\theta_1 = 1$*; iii)* $\succsim$ *is represented by* $V(\hat{P}) = \sum_{n=1}^{\infty} \theta_n p_n$*, where* $\hat{p} = \Gamma_\mu(\hat{P})$*.*

Similar to Savage (1954), expected utility representation requires stronger independence conditions. Thus, we modify Axiom 4 (Monotone Independence) into the following two axioms. First, Axiom 4.1 resembles the sure thing principle axiom in Savage theory.

Axiom 4.1 (Sure Thing Principle) Suppose that $\hat{P} \approx^n \hat{Q}, \hat{R} \approx^n \hat{S}$, $\cup_{i=1}^n P_i = \cup_{i=1}^n Q_i \in \mathcal{T}$ and $Q_k = S_k, P_k = R_k, k = 1, ..., n$, then

$$\hat{P} \succsim \hat{Q} \Rightarrow \hat{R} \succsim \hat{S}$$

Denote $A := \cup_{i=1}^n P_i = \cup_{i=1}^n Q_i \in \mathcal{T}$. Since $\cup_{i=1}^n P_i = \cup_{i=1}^n Q_i \in \mathcal{T}$ and $Q_k = S_k, P_k = R_k$, for$k = 1, ..., n$, $\hat{P}$ and $\hat{R}$, $\hat{Q}$ and $\hat{S}$ agree on the reduced problem restricted to A. Also $\hat{P} \approx^n \hat{Q}, \hat{R} \approx^n \hat{S}$ implies that $\hat{P}$ and $\hat{Q}$, $\hat{R}$ and $\hat{S}$ agree on the reduced problem restricted to A^c. Then Axiom 4.1 simply means when comparing two problems, it suffices to compare their reduced problem restricted to the trivial event on which they disagree, that is, A in this case.

By comparison, Axiom 4.2 focuses on the other part of Axiom 4 and establishes the existence of a consistent assessment of probability.

Axiom 4.2 (Weak Comparative Probability) Take $m > n, v > t$, suppose that

$$\hat{P} =^{m,n} (A_1, B_1, \emptyset), \hat{R} =^{v,t} (A_1, B_1, \emptyset), \hat{Q} =^{m,n} (A_2, B_2, \emptyset), \hat{S} =^{v,t} (A_2, B_2, \emptyset)$$

then

$$\hat{P} \succsim \hat{Q} \Rightarrow \hat{R} \succsim \hat{S}$$

Clearly, Axiom 4.2 is a special case of Axiom 4 where C is always $\emptyset$ and thus the same intuition applies that comparing $\hat{P}$ and $\hat{Q}$ ($\hat{R}$ and $\hat{S}$) is simply comparing the likelihood of A_1 and A_2. It is not straightforward that Axiom 4.1 and Axiom 4.2 are stronger than Axiom 4. However, the following lemma validates this relationship given other axioms.

Lemma 2. *Suppose that Axioms 1, 2, 3, 5, 6 hold, then Axioms 4.1 and 4.2 $\Rightarrow$ Axiom 4.*

Actually, Axioms 1, 2, 3, 4.1, 4.2, 5^*, 6 are enough to guarantee the existence of an EU representation, but not a bounded EU representation. Intuitively, in Axiom 5^* (Small Event Continuity), we are allowed to choose different partitions $\{A_j\}$ for different problems $\hat{R}$. Thus, even if a sequence of problems whose utility approaches (negative) infinity, the axiom can still be satisfied if the maximum likelihood of blocks in $\{A_j\}$ converges to 0 at a higher rate. Thus, uniform boundedness requires a stronger version of continuity which excludes such scenarios.

Axiom 5^{**}**. (Uniform Small Event Continuity)** For any problems $\hat{P} = (P_1, ..., P_n)$, $\hat{Q} = (Q_1, ..., Q_m) \in \mathbb{P}(\Omega)$ such that $\hat{P} \succ \hat{Q}$, there exists a finite partition of Ω as $\{A_1, ..., A_s\}$ with $A_i \in \mathcal{T}$ for all i such that, for each $i = 1, 2, ..., s$, for any $\hat{R} = (R_1, ..., R_l) \in \mathbb{P}(\Omega)$ with

$$\hat{P}^i = (P_1 \backslash A_i, ..., P_n \backslash A_i, R_1 \cap A_i, ..., R_l \cap A_i)$$

$$\hat{Q}^i = (Q_1 \backslash A_i, ..., Q_m \backslash A_i, R_1 \cap A_i, ..., R_l \cap A_i)$$

we have $\hat{P}^i \succ \hat{Q}$ and $\hat{P} \succ \hat{Q}^i$.

Compared to Axiom 5^*, we add the requirement that the partition $\{A_1, ..., A_s\}$ should be the same for any problem $\hat{R}$, that is, changing the reduced problem reduced to each A_i to any other problem will not reverse the original strict ranking. Since the likelihood of each A_i is fixed as we vary $\hat{R}$, the utility of $\hat{R}$ must uniformly bounded to avoid preference reversal. Then the following characterization theorem for the bounded EU representation holds.

Theorem 3. *$\succsim$ satisfies Axioms 0-3, Axioms 4.1-4.2, Axiom 5^{**} and Axiom 6 if and only if it admits a bounded EU representation.*

A special case of bounded EU representation is that $\theta_n = 1/n$, which means that the DM thinks that the probability of choosing the correct alternative when she is uncertain bout n options is exactly $1/n$. However, we allow for other possibilities since θ_n is a subjective assessment and represents the attitude towards simplicity (see the next section for more details). For instance, one may be pessimistic in the sense that she might think the nature will always go against her and thus $\theta_n < 1/n$. By comparison, $\theta_n > 1/n$ is also possible when the DM is optimistic and regards herself as a lucky person. Thus, we allow for the variation and put no restrictions other than monotonicity.

5 Perception and Attitude towards Simplicity

Although the objects on which the preference is defined are problems (partitions of the state space) instead of Savage acts (mappings from the state space to some exogenous outcome space), the prior plays the same role as in the traditional expected utility theory and the EUU model in Gul and Pesendorfer (2014). Specifically, it transforms each problem to a finite lottery over difficulty levels. Moreover, this mapping is onto, that is, for any lottery $\hat{p} \in \mathcal{L}(\mathbb{N})$, there exists a problem $\hat{P} \in \mathbb{P}(\Omega)$ that induces p. Thus the prior serves to determine how the DM *perceives* the simplicity of a problem.

By comparison, the induced preference over lotteries on difficulty levels measures the *attitude* towards the simplicity of a problem. Formally, following Gul and Pesendorfer (2014), two preferences for simplicity $\succsim_1$ with representation $(\Sigma_1, \mu_1, \succsim_1^l)$ and $\succsim_2$ with representation $(\Sigma_2, \mu_2, \succsim_2^l)$ have the *same attitude towards simplicity* if $\Gamma_{\mu^1}(\hat{P}^1) = \Gamma_{\mu^2}(\hat{P}^2)$, $\Gamma_{\mu^1}(\hat{Q}^1) = \Gamma_{\mu^2}(\hat{Q}^2)$ implies that

$$\hat{P}^1 \succsim_1 \hat{Q}^1 \Leftrightarrow \hat{P}^2 \succsim_2 \hat{Q}^2$$

The following Lemma 3 shows that the preference for simplicity model achieves the separation between simplicity perception and attitude. Notice that once controlling for the perception of simplicity in $\mathcal{L}(\mathbb{N})$, the prior of the agent becomes irrelevant. In this way, we can isolate the simplicity attitude from the simplicity level the agent perceives.

Lemma 3. *The preferences for simplicity* $(\Sigma_1, \mu_1, \succsim_1^l)$ *and* $(\Sigma_2, \mu_2, \succsim_2^l)$ *have the same attitude towards simplicity if and only if* $\succsim_1^l = \succsim_2^l$.

Given the separation lemma, we can study the comparative statics of the perception of simplicity (Σ, μ) and the attitude towards simplicity $\succsim^l$ separately. First, we call a perception of simplicity to be *more accurate* than another if it always transforms a given problem to a more preferred lottery no matter what the attitude is. For now, we just require the attitude $\succsim^l$ to be a preference relation on $\mathcal{L}(\mathbb{N})$ that respects monotonicity. Later we will briefly discuss the situation where the attitude $\succsim^l$ need to admit a mixture continuous utility representation or bounded EU representation.

Definition 8. *A perception* (Σ_1, μ_1) *is* ***more accurate*** *than* (Σ_2, μ_2) *if* $\Gamma_{\mu^1}(\hat{P}) \succsim^l \Gamma_{\mu^2}(\hat{P})$ *for any* $\hat{P} \in \mathbb{P}(\Omega)$ *and any attitude* $\succsim^l$.

For $\hat{p}, \hat{q} \in \mathcal{L}(\mathbb{N})$, we call $\hat{p}$ *weakly FOSD* $\hat{q}$ if $\forall n, F_n(\hat{p}) \leq F_n(\hat{q})$ where $F_n(\hat{p}) = \sum_{i=1}^n p_n$ is the cumulative distribution function of $\hat{p}$. Moreover, $\hat{p}$ *FOSD* $\hat{q}$ if $\hat{p}$ weakly FOSD $\hat{q}$ and at least one above inequality is strict. Recall that the attitude $\succsim^l$ is required to be monotone, that is, $\hat{q} \succsim^l \hat{p}$ if $\hat{p}$ weakly FOSD $\hat{q}$. Thus, if for any $\hat{P} \in \mathbb{P}(\Omega)$, $\Gamma_{\mu^2}(\hat{P})$ weakly FOSD $\Gamma_{\mu^1}(\hat{P})$, then (Σ_1, μ_1) must be more accurate than (Σ_2, μ_2). The following Lemma establishes the validity of the reverse.

Lemma 4. *A perception* (Σ_1, μ_1) *is more accurate than* (Σ_2, μ_2) *if and only if* $\Gamma_{\mu^2}(\hat{P})$ *weakly FOSD* $\Gamma_{\mu^1}(\hat{P})$, *for any* $\hat{P} \in \mathbb{P}(\Omega)$.

Lemma 4 ensures that we can focus on the incomplete order *weak FOSD* on $\mathcal{L}(\mathbb{N})$ when comparing the accuracy of different perceptions of simplicity. Recall for each perception (Σ, μ), μ is a prior (complete, convex-ranged and countably additive probability measure) on the measurable space (Ω, Σ). We follow the standard definition that perception (Σ_1, μ_1) is an *extension* of (Σ_2, μ_2) if $\Sigma_2 \subseteq \Sigma_1$ and $\mu_1(A) = \mu_2(A)$ for all $A \in \Sigma_2$. Intuitively, if an agent's perception is extended, then her information structure becomes finer and she is better informed of where the true state lies when it is realized. In this way, the agent can perceive and thus solve the problem more accurately. This observation leads to the following characterization of comparison of accuracy.

Theorem 4. *A perception* (Σ_1, μ_1) *is more accurate than* (Σ_2, μ_2) *if and only if* (Σ_1, μ_1) *is an extension of* (Σ_2, μ_2)*.*

There is another possible definition of comparing accuracy by focusing on preferences that admit a bounded EU representation.[6] A perception (Σ_1, μ_1) is defined to be *weakly more accurate* than (Σ_2, μ_2) if $\Gamma_{\mu^1}(\hat{P}) \succsim^l \Gamma_{\mu^2}(\hat{P})$ for any $\hat{P} \in \mathbb{P}(\Omega)$ and any attitude $\succsim^l$ that admits a bounded EU representation. Apparently, this notion is weaker than Definition 8 as the comparison of induced lotteries are only required to hold for a strictly smaller set of attitudes. However, the following corollary shows that the two definitions are actually equivalent, reflecting the robustness of the comparison of accuracy to some extent.

Corollary 1. *A perception* (Σ_1, μ_1) *is more accurate than* (Σ_2, μ_2) *if and only if* (Σ_1, μ_1) *is weakly more accurate than* (Σ_2, μ_2)*.*

Next, we turn to the comparative statics of the attitudes towards simplicity. To begin with, since all attitudes $\succsim^l$ are monotone, we know that $\hat{q} \succsim^l \hat{p}$ if $\hat{p}$ weakly FOSD $\hat{q}$ and $\hat{q} \succ^l \hat{p}$ if $\hat{p}$ FOSD $\hat{q}$. For any probability distributions with finite support on natural numbers $\hat{p}$ and $\hat{q}$, $\hat{p}$ is defined as a *mean-preserving spread* of $\hat{q}$ if $\sum_{k\geq 1} k \cdot p_k = \sum_{k \geq 1} k \cdot q_k$ and

$$\int_{-\infty}^{x} (F_{\hat{q}}(t) - F_{\hat{p}}(t))dt \geq 0, \quad \forall x \in \mathbb{R}$$

and strict inequality holds for at least some x, where $F_{\hat{p}}(\cdot)$ is the cumulative function of $\hat{p}$ defined over $\mathbb{R}$. Similarly we can define *weak mean-preserving spread* if no strict inequality is required. For instance, $(1/2, 0, 1/2)$ is a mean-preserving of $(0, 1, 0)$. Recall that $k \in \mathbb{N}$ can be interpreted as the difficulty level, which means the number of options the agent is uncertain about after making full use of prior knowledge. Just like prizes as the exogenous outcomes in the expected utility theory, the difficulty levels play the role of endogenous outcomes in our model. Then each lottery in $\mathcal{L}(\mathbb{N})$ is exactly the subjective distribution of endogenous outcomes. For instance, $(1/2, 0, 1/2)$ means there is a half probability that the

[6] The definition where we focus on preferences that admit a mixture continuous utility representation is an intermediate case and Corollary 1 implies that it is equivalent to the other two definitions discussed in the main text.

difficulty level is 1 and a half probability that the difficulty level is 3, while $(0,1,0)$ means that the difficulty level is always 2. Thus, $\hat{p}$ is a mean-preserving spread of $\hat{q}$ exactly when the variation of difficulty levels under $\hat{p}$ is larger than that under $\hat{q}$. In this way, by considering mean-preserving spreads, we can characterize the agent's attitude towards variation in the difficulty levels.

Definition 9. *An attitude $\succsim^l$ is* ***variation-averse*** *if $\hat{p}$ is a mean-preserving spread of $\hat{q}$ implies $\hat{q} \succ^l \hat{p}$. Similarly, an attitude $\succsim^l$ is* ***variation-loving*** *if $\hat{p}$ is a mean-preserving spread of $\hat{p}$ implies $\hat{q} \succ^l \hat{q}$.*

Due to the separation lemma, an attitude is a monotone preference over finite lotteries on natural numbers, regardless of the perception. Then variation aversion (loving) exactly corresponds to risk aversion (loving) in the expected utility theory. Thus, we can state the following characterization lemma without proof.

Lemma 5. *Suppose that the attitude $\succsim^l$ is represented by $U(\hat{p}) = \sum_{n=1}^{\infty} p_n \theta(n)$ where $\theta : \mathbb{N} \to [0,1]$ is strictly decreasing. Then $\succsim^l$ is variation-averse if and only if θ is strictly concave* [7]*.*

We conclude this section by relating to the measure for uncertainty of events in Gul and Pesendorfer (2014). They define A as *more uncertain* than B if the less uncertainty averse agent prefers to betting on A than B whenever the more uncertainty averse agent has the same ranking. Then they show that A is more uncertain than B if and only if $\mu_*(A) < \mu_*(B)$ and $\mu_*(A^c) < \mu_*(B^c)$. Similarly, we can use the preference for simplicity restricted on binary problems to derive a comparative measure for the *difficulty of events*.

Definition 10. *Event A is* ***more difficult than*** *B under perception (Σ, μ) if $(B, B^c) \succ (A, A^c)$ for any $\succsim$ with the preference for simplicity representation $(\Sigma, \mu, \succsim^l)$ for some $\succsim^l$.*

In other words, A is *more difficult than* B under some perception when the binary problem of B is perceived to be simpler than the binary problem of A for any agent with the same perception. By the monotonicity of $\succsim^l$ and the observation that the lottery induced by a binary problem can be represented by one number, it is easy to derive the following characterization of the measure of difficulty and its relationship with the measure of uncertainty.

[7] We call θ over $\mathbb{N}$ is strictly concave if $\theta((1-\lambda)x + \lambda y) > (1-\lambda)\theta(x) + \lambda\theta(y)$ for any $\lambda \in [0,1]$, $x, y, (1-\lambda)x + \lambda y \in \mathbb{N}$ and $x \neq y$.

Lemma 6. *Given perception* (Σ, μ)*, then (1)* A *is more difficult than* B *if and only if* $\mu_*(A) + \mu_*(A^c) \leq \mu_*(B) + \mu_*(B^c)$*; (2) if* A *is more uncertain than* B*, then* A *is more difficult than* B*.*

Intuitively, with probability $\mu_*(A) + \mu_*(A^c)$, the agent can exactly know which alternative in the binary problem (A, A^c) is correct after the state is realized and then the lower the probability is, the harder the binary problem (A, A^c) is. Thus $1-(\mu_*(A)+\mu_*(A^c))$ is a reasonable measure of the difficulty of the event A. Moreover, suppose A is more uncertain than B, then $\mu_*(A) < \mu_*(B)$ and $\mu_*(A^c) < \mu_*(B^c)$, which directly implies $\mu_*(A) + \mu_*(A^c) \leq \mu_*(B) + \mu_*(B^c)$ and thus A is more difficult than B.

6 Discussion

In this section, we compare our model for simplicity with models about ambiguity. A possible concern is that the preference for simplicity combines menu preference and ambiguity as a problem can be interpreted as a menu of acts, and thus our model hedges in the ambiguity literature. For instance, consider the Anscombe Aumann setting with two exogenous prizes 1 and 0. The problem with the partition $(P_1, ..., P_n)$ is associated with menu $\{f_1, ..., f_n\}$ where f_k takes the degenerate lottery at the superior outcome 1 when P_k is realized and takes the degenerate lottery at the inferior outcome 0 when P_k^c is realized. Then the preference over problems can be considered as a preference over menu of acts.

Recall that there are two important aspects of decision making we want to address with the preference for simplicity. First, the agent should rationally expect that she would have exhausted her prior knowledge to solve the problem. As a result, any two states with the same number of possibly correct alternatives ex post should be taken as identical regarding their simplicity. Second, the number of possibly correct ex post alternatives matters. That is, the more choices the DM cannot exclude in the reduced problem, the more difficult the problem is. However, it is hard to find out a natural model of menu preference and ambiguity to accommodate the following two basic properties of preference for simplicity. Roughly speaking, the difference is that ambiguity concerns with uncertainty about exogenous outcomes within an act, while simplicity measures an aspect cross acts. In the following, we illustrate the difference in details with two examples.

We first consider the maxmin expected utility (MEU) in Gilboa and Schmeidler (1989) with the typical urn example.

Example 2. *There are two urns. Urn 1 contains either 100 white balls or 100 black balls. Urn 2 contains 100 white and black balls in total while the component is unknown. The agent knows nothing more and needs to make a choice between the two urns, and guess the true component of the urn. Once his guess is correct, he is awarded 100 dollars. Otherwise, he is paid nothing.*

With preference for simplicity, the agent would choose urn 1 since the induced lottery of urn 1 is degenerate at difficulty level 2, while that of urn 2 is degenerate at difficulty level 101. However, by transferring each alternative into an act, it turns out each act in both urns yields a degenerate lottery at prize 0 for states in some unmeasurable event and a degenerate lottery at prize \$100 for states in its complement. As only the entire state space and the empty set are measurable, any prior over the state space should lie in the set of distributions Δ in the MEU representation[8]:

$$W(f) = \min_{\mu \in \Delta} \int_{\Omega} (U \circ f) d\mu$$

where U is a vNM utility function. Thus, if randomization is precluded as is usual in the literature, then under MEU, any act in both urns should deliver the utility of the degenerate lottery with prize 0 dollar. This implies that it is indifferent for the agent to guess from urn 1 or urn 2. In other words, the number of possibly correct choices is irrelevant in the MEU model. This is the main difference between MEU (Gilboa and Schmeidler, 1989) and the preference for simplicity. A similar argument also accounts for our difference from EUU in Gul and Pesendorfer (2014).

If instead, randomization is allowed in the MEU model and the agent believes that nature moves before she randomizes, then hedging might help to render ambiguity irrelevant. For instance, in Urn 1, suppose that the agent randomizes between the two alternatives with equal probability, then no matter which component is correct, she will face the same lottery with probability 1/2 for 100 dollars and 1/2 for 0 dollar. Similarly, in Urn 2, the agent can randomize uniformly to guarantee the same lottery with probability 1/101 for 100 dollars and 100/101 for 0 dollar for all states. By monotonicity, the agent will prefer the lottery associated with the Urn 2, which means that the number of possible choices does matter,

[8]One can refer to Section 4 in Gul and Pesendorfer (2014) for more details.

although in a different way from the preference for simplicity. This implies an implicit link between our model and the MEU model through randomization and hedging among acts. However, randomization is not yet well understood in the presence of ambiguity and is usually ruled out in many applications[9] , while our model provides an intuitive and well-founded justification for the relevance of the number of ex post possible choices in terms of simplicity of problems.

Another typical ambiguity model relaxes additivity and introduces capacity over all events as in Schmeidler (1989). Given there are two exogenous outcomes and *succsim* is represented by a Choquet EU with capacity η, we can normalize the utility of degenerate lottery with outcome 1 to be 1 and the utility of degenerate lottery with outcome 0 to be 0. Then for each act f_A which takes 1 on set A and 0 on set A^c, the utility of f is given by $\eta(A)$. Suppose that $\succsim$ also admits a preference for simplicity representation $(\Sigma, \mu, \succsim^l)$, then by Lemma 6, the utility of problem (A, A^c) can be represented by $\mu_*(A) + \mu_*(A^c)$. Since problem (A, A^c) can be interpreted as a menu of acts $\{f_A, f_{A^c}\}$ and $\mu_*(\emptyset) = \eta(\emptyset) = 0, \mu_*(\Omega) = \eta(\Omega) = 1$, generically we should have $\eta = \mu_*$, that is, the capacity is the inner measure of the prior. Then for a general problem $\hat{P} = (P_1, ..., P_K)$ with K partitions, then under Choquet EU, the inner measures of each block $\{\mu_*(P_1), ..., \mu_*(P_K)\}$ should be a sufficient statistic for the utility over $\hat{P}$, which contradicts with the preference for simplicity. Briefly, the main difference is that Choquet EU only distinguishes between the situation where the DM knows the correct alternative for sure and the situation where she is still uncertain, while our model of preference for simplicity takes into account the number of uncertain alternatives to measure simplicity.

7 Conclusion

In this paper, we study the preference over problems defined as finite partitions of the state space. The decision maker wants to solve the problem by identifying the correct alternative after the true state is realized. We define and characterize the preference for simplicity as a subjective assessment of the simplicity of problems. The basic representation consists of a prior μ on the measurable space (Ω, Σ) and a monotone preference $\succsim^l$ over finite lotteries defined on natural numbers. Each natural number k is the endogenous outcome interpreted

[9]See Saito (2015) and Ke and Zhang (2018) for a more detailed discussion.

as the difficulty level and it represents the situation where k choices remain to be possibly correct after the agent has exhausted all her knowledge. Formally, the decision maker assesses the difficulty of each problem by first transforming it into a finite lottery using the prior μ and then evaluating the induced lottery with the induced preference over lotteries $\succsim^l$. Although the two-step evaluation is standard in the literature (Savage (1954), Gul and Pesendorfer (2014)), the objectives are no longer acts and there are no well-defined exogenous outcomes. Thus, new set of axioms are required to characterize the preference for simplicity.

By definition, the preference for simplicity may fail to admit a utility representation since the induced preference on lotteries is only required to be monotone. As a result, stronger axioms are necessary for more structures in the representation. For example, a counterpart of the small event continuity axiom in Savage theory can help to guarantee the existence of a mixture continuous utility representation. If we further strengthen the independence axiom and incorporate a modified version of the sure thing principle axiom, then the preference for simplicity will admit an expected utility representation.

Our model also maintains the separation between the perception of simplicity (i.e. the knowledge algebra and prior (Σ, μ)) and the attitude towards simplicity (i.e. the induced preference over lotteries). We characterize the accuracy of perceptions and variation-aversion of attitudes respectively. We also define a measure of difficulty for events and compare it with the measure of uncertainty introduced by Gul and Pesendorfer (2014).

Finally, we want to point out some potential extensions of our model. One direction is to allow the DM's prior to be defined over a σ-algebra different from her knowledge algebra and then introduce costly information acquisition. The technical obstacle is to derive a sensational measure for difficulty under such cases. Another possible extension is to consider general menus of acts and define the problem as choosing the optimal act from the menu. However, it should be noted that "simplicity" would be tortured by the difference in payoffs of acts. Thus, one possible representation is to separate the utility of the chosen act from the (dis)utility associated with the difficulty of the problem.

8 Appendix A: Technical Results on Diffuse Events

For a probability space (Ω, Σ, μ), we define an event A to be *μ-diffuse* or *diffuse in* (Ω, Σ, μ) if the inner measures of both A and A^c are 0 under μ, that is, $\mu_*(A) = \mu_*(A^c) = 0$. . The next lemma establishes the existence and properties for diffuse events for a countably additive and convex ranged probability measure.

Lemma 7. *Assume the continuum hypothesis holds and μ is a countably additive and convex ranged probability measure on (Ω, Σ). Then there exists a μ-diffuse set D. Moreover, for any n, there exists $\{D_1, ..., D_n\}$ as a partition of Ω such that the union of any proper subcollection of $\{D_1, ..., D_n\}$ is μ-diffuse, that is, for any $\{D_{k_1}, ..., D_{k_m}\} \subset \{D_1, ..., D_n\}$ with $m < n$, $\cup_{i=1}^m D_{k_i}$ is μ-diffuse.*

Proof of Lemma 7. Lemma A2 in Gul and Pesendorfer (2014) shows that if μ is a prior (complete, convex ranged and countably additive), then the set of all μ-diffuse sets $\mathcal{D}_\mu$ is nonempty and every $D \in \mathcal{D}_\mu$ can be partitioned into $D_1, D_2 \in \mathcal{D}_\mu$.

Now consider (Ω, Σ, μ), μ is convex ranged and countably additive, we could complete the measure space in the following way. Let $\mathcal{N}_0$ denote the null events in Σ, that is, $N \in \mathcal{N}_0 \Leftrightarrow \mu(N) = 0$. Let $\mathcal{N}^* := \{E : \exists N \in \mathcal{N}_0, E \subseteq N\}$. Define $\bar{\Sigma}$ as $A \in \bar{\Sigma} \Leftrightarrow \exists E \in \mathcal{N}^*, B \in \Sigma$ such that $A = B \cup E$. Define $\bar{\mu}$ as $\bar{\mu}(B \cup E) = \mu(B)$. To see why $\bar{\mu}$ is well defined, suppose $A = B_1 \cup E_1 = B_2 \cup E_2$ with $B_1, B_2 \in \Sigma$ and $E_1, E_2 \in \mathcal{N}^*$ and $\mu(B_1) > \mu(B_2)$ (without loss of generality), then $B_1 \backslash B_2 \subseteq E_2$ and $\mu(B_1 \backslash B_2) > 0$, which is impossible as E_2 is a subset of some null set.

Then it is easy to show that the space $(\Omega, \bar{\Sigma}, \bar{\mu})$ is an extension of (Ω, Σ, μ) and is complete. Moreover, μ is convex ranged and countably additive. Thus, μ is a prior and by Lemma A2 in Gul and Pesendorfer (2014), we could find $D \subset \Omega$ that is diffuse in $(\Omega, \bar{\Sigma}, \bar{\mu})$ and every $D \in \mathcal{D}_{\bar{\mu}}$ can be partitioned into $D_1, D_2 \in \mathcal{D}_{\bar{\mu}}$. Easy to show that D, D_1, D_2 are diffuse in (Ω, Σ, μ) since $(\Omega, \bar{\Sigma}, \bar{\mu})$ extends (Ω, Σ, μ) and $\mu_* = \bar{\mu}_*$. Since $D \in \mathcal{D}_\mu \Leftrightarrow D^c \in \mathcal{D}_\mu$, we can repeat the division so that for any n, Ω can be partitioned into n μ-diffuse sets $\{D_1, ..., D_n\}$. Then for any proper subcollection $\{D_{k_1}, ..., D_{k_m}\} \subset \{D_1, ..., D_n\}$ with $m < n$, choose $D \in \{D_1, ..., D_n\} - \{D_{k_1}, ..., D_{k_m}\}$, we know that $\cup_{i=1}^m D_{k_i} \in D^c$, $(\cup_{i=1}^m D_{k_i})^c \in D_{k_1}^c$. By monotonicity of inner measure μ_* and the fact that D_{k_1}, D are μ-diffuse,

$$\mu_*(\cup_{i=1}^m D_{k_i}) = \mu_*(D^c) = 0, \quad \mu_*((\cup_{i=1}^m D_{k_i})^c) = \mu_*(D_{k_1}^c) = 0$$

Thus, $\cup_{i=1}^m D_{k_i}$ is μ-diffuse and this completes the proof. $\square$

9 Appendix B: Proofs

9.1 Proofs in Section 2

Proof of Lemma 1. Given a problem $\hat{P} = \{P_i\}_{i=1}^K$ and any corresponding exhaustive information partition $\mathcal{I}(\hat{P})$, we want to show that the induced lottery is independent of $\mathcal{I}(\hat{P})$. We will focus on $n \leq K$ as the probability of difficulty level n is 0 for $n > K$.

Step 1: First, for any $n \leq K$ and $A = \cup_{i=1}^n P_{k_i} \in \mathcal{D}_n(\hat{P})$ as the union of n alternatives $\{P_{k_i}\}_{i=1}^n$, we define $E[A] \subseteq A$ as the set of states in A where the DM knows that the correct alternative is among $\{P_{k_i}\}_{i=1}^n$:

$$E[A] := \{w \in A : \{P_n | I(w) \cap P_n \notin \mathcal{N}\} \subseteq \{P_{k_i}\}_{i=1}^n\}$$

It is clear that $E[A] = \{w \in A : \mu_*(A \cap I(w)) = \mu(I(w))\}$, that is, $w \in E[A]$ if and only if the DM essentially knows that A occurs if w is realized. In other words, $E[A]$ is the set of states where the difficulty level is less than or equal to n and the DM can exclude any alternative outside of A. Denote $G(n) := \{w : K_{\hat{P},\mathcal{I}(\hat{P})}(w) \leq n\}$ as the set of states with difficulty level no larger than n. Then clearly,

$$\mu(G(n)) = \mu(\cup_{A \in \mathcal{D}_n(\hat{P})} E[A]) \tag{2}$$

Step 2: Now we claim that $\mu(E[A]) = \mu_*(A)$, which implies that $\mu(E[A])$ is independent of the chosen exhaustive information partition $\mathcal{I}(\hat{P})$. First, notice that the difference between $E[A]$ and the union of the collection of sets $\{I_m : \mu_*(P_1 \cap I_m) = \mu(I_m)\}$ is a measure 0 set, we have

$$\mu(E[A]) = \mu\left(\cup_{\{m:\mu_*(A\cap I_m)=\mu(I_m)\}} I_m\right) = \sum_{\{m:\mu_*(A\cap I_m)=\mu(I_m)\}} \mu(I_m)$$

On the one hand, denote $B = \cup_{\{m:\mu_*(A\cap I_m)=\mu(I_m)\}} I_m$, then $\mu_*(A \cap B) = \mu(B)$, which implies $B \backslash A$ has measure 0 and $\mu(A \cap B) = \mu(B)$. By definition of inner measure, $\mu(E[A]) = \mu(B) \leq \mu_*(A)$. One the other hand, suppose that $\mu(E[A]) = \mu(B) < \mu_*(A)$, then as $\mathcal{I}(\hat{P})$ is a countable partition of Ω consisting of non-null sets, there exists $I^* \in \mathcal{I}(\hat{P})$ such that $0 < \mu_*(A \cap I^*) < \mu(I^*)$. This contradicts with the definition of exhaustive information partition.

Step 3: By definition, it is clear that $E[\cdot] : \cup_{n=1}^K \mathcal{D}_n(\hat{P}) \to \Sigma$ is a well-defined operator such that $\forall A \in \cup_{n=1}^K \mathcal{D}_n(\hat{P})$, $\mu_*(A) = \mu(E[A])$ and $\forall A, B \in \cup_{n=1}^K \mathcal{D}_n(\hat{P}), A \subseteq B \Rightarrow E[A] \subseteq E[B]$.

Then, for any finite sequence of events $\{A_k\}_{k=1}^m \subset \cup_{n=1}^K \mathcal{D}_n(\hat{P})$, we want to show that

$$\mu(\cup_{k=1}^m E[A_k]) = \sup_{\forall k, F_k \subseteq A_k : F_k \in \Sigma} \mu(\cup_{k=1}^m F_k) \tag{3}$$

By definition of $E[A_k]$, it is clear that LHS $\leq$ RHS. For the other direction, suppose by contradiction that LHS $<$ RHS. Then, there exists a sequence of $\{F_k\}_{k=1}^m$ such that $F_k \subseteq A_k$ for all k and $\mu(\cup_{k=1}^m E[A_k]) < \mu(\cup_{k=1}^m F_k)$. This implies $\mu(\cup_{k=1}^m (F_k \backslash E[A_k])) > 0$ and thus $\exists j$, s.t. $\mu(F_j \backslash E[A_j]) > 0$. Now, consider $F_j \cup E[A_j]$, we know that $F_j \cup E[A_j] \subseteq A_j$, $F_j \cup E[A_j] \in \Sigma$ and $\mu(F_j \cup E[A_j]) = \mu(F_j \backslash E[A_j]) + \mu(E[A_j]) > \mu(E[A_j]) = \mu_*(A_j)$, which contradicts with the definition of $\mu_*(A_j)$. This proves the equality (3).

Now, we can simplify the expression for p_k in equation (1) as

$$\begin{aligned} p_n &= \sum_{A \in \mathcal{D}_k(\hat{P})} \Big(\mu(E[A]) - \mu(\cup_{i=1}^n E[A^i])\Big) \\ &= \sum_{A \in \mathcal{D}_n(\hat{P})} \mu(E[A] \backslash (\cup_{i=1}^n E[A^i])) \end{aligned} \tag{4}$$

The first equality comes from the definition of $E[\cdot]$ and equation (3) while the second equality is correct since $A^i \subseteq A \Rightarrow E[A^i] \subseteq E[A]$ for each i. For simplicity, denote $F[A] \equiv E[A] \backslash (\cup_{i=1}^n E[A^i])$ for each A.

Further, for any $n = 1, \ldots, K$, $A, B \in \mathcal{D}_n(\hat{P})$ and $A \neq B$, we want to show $\mu(F[A] \cap F[B]) = 0$. If $A \cap B = \emptyset$, then this is trivial as $F[A] \cap F[B] = \emptyset$. If instead $A \cap B \neq \emptyset$, then $A \cap B \subseteq A^j$ or B^j for some j. WLOG, suppose $A \cap B \subseteq A^j$, then $E[A \cap B] \subseteq \cup_{i=1}^n E[A^i]$. This implies that $E[A \cap B] \cap F[A] = \emptyset$. Now suppose by contradiction that $\mu(F[A] \cap F[B]) > 0$. Then $F[A] \cap F[B] \subseteq A \cap B$ and $(F[A] \cap F[B]) \cap E[A \cap B] = \emptyset$. So $\mu((F[A] \cap F[B]) \cup E[A \cap B]) > \mu(E[A \cap B]) = \mu_*(A \cap B)$, contradicting with definition of $\mu_*(A \cap B)$.

Thus, for any $n = 1, \ldots, K$, $A, B \in \mathcal{D}_n(\hat{P})$ and $A \neq B$, $\mu(F[A] \cap F[B]) = 0$ and

$$\begin{aligned} p_n &= \sum_{A \in \mathcal{D}_n(\hat{P})} \mu(E[A] \backslash (\cup_{i=1}^n E[A^i])) = \sum_{A \in \mathcal{D}_n(\hat{P})} \mu(F[A]) \\ &= \mu(\cup_{A \in \mathcal{D}_n(\hat{P})} E[A]) - \mu(\cup_{A \in \mathcal{D}_{n-1}(\hat{P})} E[A]) \\ &= \mu(G(n)) - \mu(G(n-1)) \end{aligned} \tag{5}$$

The last equality comes from equation (2). Thus, p_n given in equation (1) exactly represents the probability of states with difficulty level n, which completes the proof. $\square$

9.2 Proofs in Section 3

Proof of Theorem 1. **Proof for Necessity**: Suppose now that $\succsim$ is a preference for simplicity with representation $(\Sigma, \mu, \succsim^l)$. The following lemma identifies the collection of trivial events ($\mathcal{T}$), intractable events ($\mathcal{D}$), null events ($\mathcal{N}$) and unverifiable events ($\mathcal{V}$) using $(\Sigma, \mu, \succsim^l)$.

Lemma 8. *The following statements are true:*

1. $A \in \mathcal{T} \Leftrightarrow A \in \Sigma$;
2. $D \in \mathcal{D} \Leftrightarrow D$ *is diffuse in* (Ω, Σ, μ);
3. $N \in \mathcal{N} \Leftrightarrow N \in \Sigma$ *and* $\mu(N) = 0$;
4. $A \in \mathcal{V} \Leftrightarrow \mu_*(A) = 0$;

Proof of Lemma 8. Note that $A \in \mathcal{T} \Leftrightarrow (A, A^c) \sim \Omega$. Then this is equivalent to $\mu_*(A) + \mu_*(A^c) = 1$. Since μ is a prior (countably additive), we know that there is $E \subseteq A, F \subseteq A^c$ such that $E, F \in \Sigma$ and $\mu(E)+\mu(F) = 1$ since Σ is a σ-algebra. Notice that this implies that $E^c \cap F^c \in \Sigma$ and that $\mu(E^c \cap F^c) = 0$. By completeness, we know that $A \backslash E \subseteq E^c \cap F^c \Rightarrow A \backslash E \in \Sigma$. Thus, $A = E \cup (A \backslash E) \in \Sigma$. The inverse part is trivial since $A \in \Sigma$ implies

$$\mu_*(A) + \mu_*(A^c) = \mu(A) + \mu(A^c) = 1$$

which means $(A, A^c) \sim \Omega$ and $A \in \mathcal{T}$ by definition.

For the second statement, given any μ-diffuse event D, we know $\mu_*(D) + \mu_*(D^c) = 0$. Then the induced lottery of (D, D^c) is $\Gamma_\mu(D, D^c) = (0, 1)$, while for any $A \subseteq \Omega$, $\Gamma_\mu(A, A^c) = (a, 1-a)$ with $a \in [0, 1]$. Since $\succsim^l$ is monotone, we know that $(A, A^c) \succsim (D, D^c)$ and thus $D \in \mathcal{D}$. Inversely, since $D \in \mathcal{D}$ implies (D, D^c) is the hardest binary problem, it must be the case that $\mu_*(D) = \mu_*(D^c) = 0$, otherwise any binary problem generated by a diffuse event would be harder than (D, D^c). Thus, D is diffuse.

For the third statement, let $N \in \mathcal{N}$. By definition, $N \in \mathcal{T} \subset \Sigma$ and $(N \cup D, N^c \cap D^c) \sim (D, D^c)$ for some $D \in \mathcal{D}$. As a result, if $\mu(N) > 0$, then

$$\mu_*(N \cup D) + \mu_*(N^c \cap D^c) \geq \mu(N) > 0 = \mu_*(D) + \mu_*(D^c)$$

This contradicts with $(N \cup D, N^c \cap D^c) \sim (D, D^c)$. For the other direction, suppose $N \in \Sigma$ and $\mu(N) = 0$, then $(N, N^c) \sim \Omega$. Also for any $D \in \mathcal{D}$, since $\mu(N) = 0$ and μ is complete,

$\mu_*(N \cup D) = \mu_*(N^c \cap D^c) = 0$. This implies $N \cup D$ is diffuse and $(N \cup D, N^c \cap D^c) \sim (D, D^c)$. Thus $N \in \mathcal{N}$.

For the last statement, $A \in \mathcal{V} \Leftrightarrow [B \subseteq A, B \in \mathcal{T} \Rightarrow B \in \mathcal{N}]$. This is equivalent to $A \in \mathcal{V} \Leftrightarrow [B \subseteq A, B \in \Sigma \Rightarrow \mu(B) = 0] \Leftrightarrow \mu_*(A) = 0$. This completes the proof. □

Now consider the axioms. Axiom 0 and Axiom 1 are trivially true.

For Axiom 2, as μ is a prior, Lemma 7 guarantees the existence of a μ-diffuse set D and monotonicity implies $(\Omega, \emptyset) \succ (D, D^c)$. For Axiom 5, since μ is convex ranged, we can construct a partition $\{P_k\}_{k=1}^n$ of Ω such that $P_k \in \Sigma$ and $\mu(P_k) = \frac{1}{n}$ for any k. Suppose $(A, A^c) \succ (B, B^c)$, then $\mu_*(A) + \mu_*(A^c) > \mu_*(B) + \mu_*(B^c)$. Denote the difference as δ. Then we can choose n large enough such that $\delta > \frac{1}{n}$. As a result, for each $k = 1, ..., n$,

$$\mu_*(B \cup P_k) + \mu_*(B^c \cap P_k^c) \leq \mu_*(B) + \mu_*(B^c) + \frac{1}{n} < \mu_*(A) + \mu_*(A^c)$$

The first inequality holds because for any $E \subseteq B \cup P_k$ and $E \in \Sigma$, $P_k \in \Sigma$ implies that both $E \cap P_n$ and $E \cap B \in \Sigma$. Thus $\mu_*(B \cup P_k) < \mu_*(B) + \mu(P_k)$. Thus, $(A, A^c) \succ (B \cup P_n, B^c \cap P_n^c)$ holds for each n.

For Axiom 3 and 4, we need some notations for the following lemmas. For any $B \in \Sigma$ such that $\mu(B) \in (0, 1]$, we can define a *probability space restricted to* B as (B, Σ^B, μ^B) where $\Sigma^B = \{A \cap B : A \in \Sigma\}$ and $\mu^B(A) = \mu(A)/\mu(B)$ for all $A \in \Sigma^B$. Then for any problem $\hat{P} = (P_1, ..., P_n)$, the problem restricted to B is defined as $\hat{P}^B = (P_1 \cap B, ..., P_n \cap B)$. Corresponding to Γ_μ, we can define the operator Γ_μ^B where $\Gamma_\mu^B(\hat{P})$ is the induced lottery of $\hat{P}^B$, the problem restricted to B, in the probability space restricted to B. Moreover, when $\mu(B) = 0$, Γ_μ^B is not well-defined and we follow the convention that $\mu(B)\Gamma_\mu^B(\hat{P}) = 0$ for all $\hat{P}$.

Recall that $\Phi(\hat{P}, B) = (P_1 \backslash B, ..., P_n \backslash B, B \cap P_1, ..., B \cap P_n)$ denotes the refined problem of $\hat{P}$ induced by B, we have the following lemma.

Lemma 9. *For any any $B \in \Sigma$, $\hat{P} \sim \Phi(\hat{P}, B)$ and $\Gamma_\mu(\hat{P}) = \mu(B)\Gamma_\mu^B(\hat{P}) + \mu(B^c)\Gamma_\mu^{B^c}(\hat{P})$.*

Proof of Lemma 9. Denote $\hat{Q} = \Phi(\hat{P}, B)$, $\hat{p} = \Gamma_\mu(\hat{P})$, $\hat{q} = \Gamma_\mu(\hat{Q})$. Without loss of generality, assume $B \cap P_i \neq \emptyset$ and $P_i \backslash B \neq \emptyset$ for all i[10]. For the first part, it suffices to show that $\hat{p} = \hat{q}$.

[10]This simplification makes the definition of $\mathcal{D}_k(\hat{Q})$ straightforward. For more general cases, we just need to take care of $\mathcal{D}_k(\hat{Q})$ and the same argument goes through.

By equations (5), we need to show for any $k = 1, ...n$, $\mu(\cup_{A\in\mathcal{D}_k(\hat{P})} E[A]) = \mu(\cup_{A\in\mathcal{D}_k(\hat{Q})} E[A])$ and for $k \geq n+1$, $q_k = 0$.

For $k = 1, ...n$, by definition of $\hat{Q}$, for any $A_Q \in \mathcal{D}_k(\hat{Q})$, there exists some $A_P \in \mathcal{D}_k(\hat{P})$ where $A_Q \subseteq A_P$. By our construction of $E[\cdot]$ (note that $A_P, A_Q \in \cup_{i=1}^{2n}\mathcal{D}_i(\hat{Q})$), $E[A_Q] \subseteq E[A_P]$. Thus, $\cup_{A\in\mathcal{D}_k(\hat{Q})} E[A] \subseteq \cup_{A\in\mathcal{D}_k(\hat{P})} E[A]$ and $\mu(\cup_{A\in\mathcal{D}_k(\hat{Q})} E[A]) \leq \mu(\cup_{A\in\mathcal{D}_k(\hat{P})} E[A])$. One the other hand, suppose by contradiction that the inequality is strict. That is, there exists $A_P \in \mathcal{D}_k(\hat{P})$ such that $\mu(E[A_P]\backslash(\cup_{A_Q\in\mathcal{D}_k(\hat{Q})} E[A_Q])) > 0$. WLOG, suppose $A_p = \cup_{i=1}^{k} P_{n_i} \in \mathcal{D}_k(\hat{P})$, then $A_{Q1} \equiv A_P \cap B = \cup_{i=1}^{k}(P_{n_i} \cap B) \in \mathcal{D}_k(\hat{Q})$ and $A_{Q2} \equiv A_P \cap B^c = \cup_{i=1}^{k}(P_{n_i} \cap B^c) \in \mathcal{D}_k(\hat{Q})$. Moreover, as $B \in \Sigma$ and $E[A_{Q1}] \subseteq E[A_P] \cap B$, $E[A_{Q2}] \subseteq E[A_P] \cap B^c$, the definition of $E[\cdot]$ requires that $\mu(E[A_P]\backslash(E[A_{Q1}] \cup E[A_{Q2}])) = 0$. This contradicts with $\mu(E[A_P]\backslash(\cup_{A_Q\in\mathcal{D}_k(\hat{Q})} E[A_Q])) > 0$.

Now we have shown that for any $k = 1, ...n$, $\mu(\cup_{A\in\mathcal{D}_k(\hat{P})} E[A]) = \mu(\cup_{A\in\mathcal{D}_k(\hat{Q})} E[A])$. By equation (5), $p_k = q_k$. This implies that $\sum_{k=1}^{n} p_k = \sum_{k=1}^{n} q_k = 1$ and $k \geq n+1$, $p_k = q_k = 0$. To conclude, we get $\Gamma_\mu(\hat{P}) = \Gamma_\mu(\hat{Q})$ and thus $\hat{P} \sim \hat{Q}$.

For the second part, it is trivial if $\mu(B) = 0$ or 1. For $\mu(B) \in (0, 1)$, we need to prove $\Gamma_\mu(\hat{Q}) = \mu(B)\Gamma_\mu^B(\hat{P}) + \mu(B^c)\Gamma_\mu^{B^c}(\hat{P}) = \mu(B)\Gamma_\mu^B(\hat{Q}) + \mu(B^c)\Gamma_\mu^{B^c}(\hat{Q})$. By definition of Γ_μ^B and $\Gamma_\mu^{B^c}$, it suffices to prove that for any $k = 1, ...n$, $\mu(\cup_{A\in\mathcal{D}_k(\hat{Q})} E[A]) = \mu(\cup_{A\in\mathcal{D}_k^B(\hat{Q})} E[A]) + \mu(\cup_{A\in\mathcal{D}_k^{B^c}(\hat{Q})} E[A]) = \mu(\cup_{A\in(\mathcal{D}_k^B(\hat{Q})\cup\mathcal{D}_k^{B^c}(\hat{Q}))} E[A])$, where $\mathcal{D}_k^B(\hat{Q})$ denotes the set of union of any k blocks in problem $\hat{Q}$ restricted to B. It is clear that $\mathcal{D}_k^B(\hat{Q}) \cup \mathcal{D}_k^{B^c}(\hat{Q}) \subset \mathcal{D}_k(\hat{Q})$ and by the argument for the first part of the lemma, restricting attention to sets in $\mathcal{D}_k^B(\hat{Q}) \cup \mathcal{D}_k^{B^c}(\hat{Q})$ (like A_{Q1} and A_{Q2} above) will not change the measure of the union. Thus, $\Gamma_\mu(\hat{P}) = \Gamma_\mu(\hat{Q}) = \mu(B)\Gamma_\mu^B(\hat{P}) + \mu(B^c)\Gamma_\mu^{B^c}(\hat{P})$. □

Intuitively, Lemma 9 implies that for any trivial event B, any problem can be interpreted as a lottery over the reduced problems restricted to B and B^c. In other words, refining the original problem with respect to a trivial event will not change its difficulty.

This following lemma holds for a general refinement. Formally, given a partition $\hat{P} = (P_1, ..., P_n)$ over Ω, the partition $\hat{Q} = (Q_1, ..., Q_m)$ over Ω is called a *refinement* of $\hat{P}$ if for any $k = 1, ..., m$, there exists some j such that $Q_k \subseteq P_j$.

Lemma 10. *Suppose that $\hat{Q}$ is a refinement of $\hat{P}$, then $\hat{P} \succsim \hat{Q}$.*

Proof of Lemma 10. Suppose that $\hat{Q}$ is a refinement of $\hat{P} = (P_1, ..., P_n)$. WLOG, we can write $\hat{Q} = (Q_{11}, ..., Q_{1k_1}, Q_{21}, ..., Q_{2k_2}, ..., Q_{n1}, ..., Q_{nk_n})$ such that for any $i = 1, ..., n$, $P_i =$

$\cup_{j=1}^{k_i} Q_{ij}$. To show that $\hat{P} \succsim \hat{Q}$, by monotonicity, it suffices to show $\forall n$, $F_n(\hat{p}) \geq F_n(\hat{q})$ where $\hat{p} = \Gamma_\mu(\hat{P})$ and $\hat{q} = \Gamma_\mu(\hat{Q})$. By equation (5), it is equivalent to $\forall n$, $\mu(\cup_{A \in \mathcal{D}_n(\hat{P})} E[A]) \geq \mu(\cup_{A \in \mathcal{D}_n(\hat{Q})} E[A])$. This is correct since for any $A_Q \in \mathcal{D}_n(\hat{Q})$, by definition of refinement, there exists some $A_P \in \mathcal{D}_n(\hat{P})$ where $A_Q \subseteq A_P$ and thus $E[A_Q] \subseteq E[A_P]$. □

Now we could show Axiom 3 and Axiom 4. For Axiom 3, consider $\hat{P} = (P_1, ..., P_n)$, then $\hat{Q} \equiv (P_1 \backslash B, ..., P_n \backslash B, B \cap P_1, ..., B \cap P_n) \sim \hat{P}$ by Lemma 9. Also, for $\hat{P}' = (P_1 \cup (P_2 \cap B), P_2 \backslash B, P_3, ..., P_n)$ $\hat{Q}' \equiv (P_1 \backslash B, ..., P_n \backslash B, B \cap (P_1 \cup P_2), B \cap P_3 ..., B \cap P_n) \sim \hat{P}'$. Note that $\hat{Q}$ is a refinement of $\hat{Q}'$, then by lemma 10, $\hat{P}' \succsim \hat{P}$ and Axiom 3 holds.

For Axiom 4, take $m > n, w > t$ let $\hat{P} =^{m,n} (A_1, B_1, C), \hat{R} =^{w,t} (A_1, B_1, C), \hat{Q} =^{m,n} (A_2, B_2, C), \hat{S} =^{w,t} (A_2, B_2, C)$, $\hat{P} \approx^{m+n} \hat{Q}$ and $\hat{R} \approx^{w+t} \hat{S}$. By Lemma 10, then $\Gamma_\mu(\hat{P}) = \mu(A_1)\Gamma_\mu^{A_1}(\hat{P}) + \mu(B_1)\Gamma_\mu^{B_1}(\hat{P}) + \mu(C)\Gamma_\mu^{C}(\hat{P})$. Since $\hat{P}$ includes a nowhere verifiable m-partition of A_1 and nowhere verifiable n-partition of B_1, $\Gamma_\mu^{A_1}(\hat{P})_m = 1$ and $\Gamma_\mu^{B_1}(\hat{P})_n = 1$. Similarly, $\Gamma_\mu(\hat{Q}) = \mu(A_2)\Gamma_\mu^{A_2}(\hat{Q}) + \mu(B_2)\Gamma_\mu^{B_2}(\hat{Q}) + \mu(C)\Gamma_\mu^{C}(\hat{Q})$ and $\Gamma_\mu^{A_2}(\hat{Q})_m = 1$ and $\Gamma_\mu^{B_2}(\hat{Q})_n = 1$. Moreover, $\hat{P} \approx^{m+n} \hat{Q}$ implies that $\Gamma_\mu^C(\hat{Q}) = \Gamma_\mu^C(\hat{P})$. By monotonicity and $m > n$, $\hat{P} \succ \hat{Q} \Leftrightarrow \mu(A_1) < \mu(A_2)$. By the same argument, we know $\hat{R} \succ \hat{S} \Leftrightarrow \mu(A_1) < \mu(A_2)$ as $w > t$. Thus, $\hat{P} \succ \hat{Q} \Rightarrow \hat{R} \succ \hat{S}$ and Axiom 4 holds.

Now we prove Axiom 6 with the previous 5 axioms. For the first part, $(E \cup \lim_n A_n, E^c \cap \lim_n A_n^c) = (E \cup A_k \cup ((\lim_n A_n) \backslash A_k), E^c \cap A_k^c \cap ((\lim_n A_n) \backslash A_k)^c) \succsim (E \cup A_k, E^c \cap A_k^c)$ by Axiom 3 as $(\lim_n A_n) \backslash A_k \in \Sigma$. By transitivity, the result is true. For the second part, there exists $\delta > 0$ such that

$$\left(\mu_*(E \cup \lim_n A_n) + \mu_*(E^c \cap \lim_n A_n^c)\right) - (\mu_*(B) + \mu_*(B^c)) > \delta$$

Apply the same argument as the proof of axiom 5 and we could find k large enough so that

$$\mu_*(E \cup A_k) + \mu_*(E^c \cap A_k^c) > \left(\mu_*(E \cup \lim_n A_n) + \mu_*(E^c \cap \lim_n A_n^c)\right) - \delta > \mu_*(B) + \mu_*(B^c)$$

This completes the proof for necessity.

Proof for Sufficiency:

Step 1: Identify the knowledge set Σ and prior μ. First, we identify the σ-algebra $\Sigma = \mathcal{T}$.

Lemma 11. *$\mathcal{T}$ is a σ-algebra.*

Proof of Lemma 11. Notice that $\Omega \in \mathcal{T}$ and $A \in \mathcal{T} \Rightarrow A^c \in \mathcal{T}$ due to our definition. By **Axiom 3 (Triviality Improvement)** , $(A \cup B, A^c \cap B^c) \succsim (A, A^c) \sim (\Omega, \emptyset)$. Thus, it must be that $(A \cup B, A^c \cap B^c) \sim (\Omega, \emptyset)$ by **Axiom 0**, which implies $A \cup B \in \mathcal{T}$.

Finally, we want to show that $\mathcal{T}$ is a σ-algebra. Take a sequence of increasing sets $\{A_1, A_2, ...\}$ in $\mathcal{T}$, it suffices to show that $\lim_n A_n \in \mathcal{T}$. Suppose not, then $\Omega \succ (\lim_n A_n, \lim_n A_n^c)$. However, by **Axiom 6 (Dominance Continuity)**, there exists k such that $\Omega \succ (A_k, A_k^c)$, which is a contradiction. Thus, $\mathcal{T}$ is a σ-algebra. □

Now we want to construct a qualitative probability measure on $\mathcal{T}$ based on the preference restricted to binary problems. The first observation is that there is essentially no difference among intractable events, which is summarized in Lemma 14, whose proof depends on the following two lemmas.

Lemma 12. $D \in \mathcal{D} \Rightarrow D, D^c \in \mathcal{V}$

Proof of Lemma 12. Suppose $D \notin \mathcal{V}$, then there is some $A \in \mathcal{T}$ such that $A \notin \mathcal{N}$, $A \subseteq D$, and $(A, A^c) \sim \Omega$. However, this implies $(D, D^c) = (A \cup D, A^c \cap D^c) \succ (D, D^c)$ according to the definition of $\mathcal{N}$. As a result, $D \in \mathcal{V}$. Also, $D \in \mathcal{D} \Rightarrow D^c \in \mathcal{D} \Rightarrow D^c \in \mathcal{V}$. □

Lemma 13. *For $\hat{P}, \hat{Q}$, $A \in \mathcal{T}$, both $\{P_1, ..., P_n\}$ and $\{Q_1, ..., Q_n\}$ are nowhere verifiable n-partition of A, and $P_{n+k} = Q_{n+k}, \forall k \geq 1$, then $\hat{P} \sim \hat{Q}$.*

Proof of Lemma 13. If $A = \emptyset$, then $\hat{P} = \hat{Q}$, and the conclusion is trivial. Otherwise, we have $n \geq 1$, $\hat{P} =^{n+1,n} (\emptyset, A, A^c)$, $\hat{Q} =^{n+1,n} (\emptyset, A, A^c)$. By **Axiom 4 (Monotone Independence)**, we know that $\hat{P} \succ \hat{Q} \Leftrightarrow \hat{Q} \succ \hat{P}$ which is impossible. Thus, $\hat{Q} \sim \hat{P}$. □

Now we can prove that for the binary problem of the union of a trivial event and an intractable event, its difficulty is independent of the intractable event.

Lemma 14. $A \in \mathcal{T} \Rightarrow (A \cup D_1, A^c \cap D_1^c) \sim (A \cup D_2, A^c \cap D_2^c)$ *for any* $D_1, D_2 \in \mathcal{D}$

Proof of Lemma 14. First, as $A \in \mathcal{T}$, by applying **Axiom 3 (Triviality Improvement)** twice, we get $(A \cup D_1, A^c \cap D_1^c) \succsim (A, D_1 \backslash A, A^c \cap D_1^c) \succsim (A \cup D_1, A^c \cap D_1^c)$. Thus $(A \cup D_1, A^c \cap D_1^c) \sim (A, D_1 \backslash A, D_1^c \backslash A)$. Similarly, $(A \cup D_2, A^c \cap D_2^c) \sim (A, D_2 \backslash A, D_2^c \backslash A)$. Moreover, by Lemma 12, $D_1, D_2 \in \mathcal{D} \Rightarrow D_1, D_2, D_1^c, D_2^c \in \mathcal{V} \Rightarrow D_1 \backslash A, D_2 \backslash A, D_1^c \backslash A, D_2^c \backslash A \in \mathcal{V}$, we know $(D_1 \backslash A, D_1^c \backslash A)$ and $(D_2 \backslash A, D_2^c \backslash A)$ are nowhere verifiable 2-partitions of $A^c \in \mathcal{T}$. Then by 13, $(A, D_1 \backslash A, D_1^c \backslash A) \sim (A, D_2 \backslash A, D_2^c \backslash A)$, which completes the proof. □

Then we define a binary relation $\succsim^*$ on $\mathcal{T}$, where $A \succsim^* B$ if and only if $(A \cup D, A^c \cap D^c) \succsim (B \cup D, B^c \cap D^c)$ for any $D \in \mathcal{D}$. By Lemma 14, $\succsim^*$ is well-defined and we can focus on one fixed $D \in \mathcal{D}$. Also, since $\succsim$ is complete and transitive, so is $\succsim^*$. To show $\succsim^*$ is a qualitative probability, we need the following lemma.

Lemma 15. *If $A, B, C \in \mathcal{T}$ and $(A \cup B) \cap C = \emptyset$, then $A \succ^* B \Leftrightarrow A \cup C \succ^* B \cup C$.*

Proof of Lemma 15. Fix $D \in \mathcal{D}$. By **Axiom 3 (Triviality Improvement)** and a similar argument in Lemma 14, as $A, B \in \mathcal{T}$, we know that $(A \cup D, A^c \cap D^c) \sim (A, D \backslash A, D^c \backslash A) \sim (A \backslash B, A \cap B, D \backslash A, D^c \backslash A) \sim (A \backslash B, A \cap B, D \cap (B \backslash A), D^c \cap (B \backslash A), D \backslash (A \cup B), D^c \backslash (A \cup B))$. Similarly, we have $(B \cup D, B^c \cap D^c) \sim (B \backslash A, D \cap (A \backslash B), D^c \cap (A \backslash B), A \cap B, D \backslash (A \cup B), D^c \backslash (A \cup B))$. Then, notice that $A \backslash B$ is a nowhere verifiable 1-partition of itself, $B \backslash A$ is a nowhere verifiable 1-partition of itself, $(D \cap (B \backslash A), D^c \cap (B \backslash A))$ is a nowhere verifiable 2-partitions of $B \backslash A$ and $(D \cap (A \backslash B), D^c \cap (A \backslash B))$ is a nowhere verifiable 2-partitions of $A \backslash B$, by **Axiom 4 (Monotone Independence)** , we know that

$$(A \backslash B, D \cap (B \backslash A), D^c \cap (B \backslash A), R_1, R_2, R_3) \succ (B \backslash A, D \cap (A \backslash B), D^c \cap (A \backslash B), R_1, R_2, R_3)$$

$$\Leftrightarrow (A \backslash B, D \cap (B \backslash A), D^c \cap (B \backslash A), S_1, S_2, S_3) \succ (B \backslash A, D \cap (A \backslash B), D^c \cap (A \backslash B), S_1, S_2, S_3)$$

where $R_1 = A \cap B, R_2 = D \backslash (A \cup B), R_3 = D^c \backslash \{A \cup B\}, S_1 = C \cup (A \cap B), S_2 = D \backslash (A \cup B \cup C)$ and $S_3 = D^c \backslash (A \cup B \cup C)$. However, again by **Axiom 3 (Triviality Improvement)**, we know that $(A \backslash B, D \cap (B \backslash A), D^c \cap (B \backslash A), S_1, S_2, S_3) \sim (A \cup C \cup D, A^c \cap C^c \cap D^c)$ and $(B \backslash A, D \cap (A \backslash B), D^c \cap (A \backslash B), S_1, S_2, S_3) \sim (B \cup C \cup D, B^c \cap C^c \cap D^c)$ since $(A \cup B) \cap C = \emptyset$. Thus $(A \cup D, A^c \cap D^c) \succ (B \cup D, B^c \cap D^c) \Leftrightarrow (A \cup C \cup D, A^c \cap C^c \cap D^c) \succ (B \cup C \cup D, B^c \cap C^c \cap D^c)$ and we are done by definition of $\succsim^*$. $\square$

The next lemma further shows that $\succsim^*$ defined above actually admits a unique well-behaved quantitative probability measure.

Lemma 16. *There is a unique countable additive convex ranged probability measure μ defined on $\mathcal{T}$ such that $A, B \in \mathcal{T}$, $\forall D \in \mathcal{D}$, $(A \cup D, A^c \cap D^c) \succsim (B \cup D, B^c \cap D^c) \Leftrightarrow \mu(A) \geq \mu(B)$.*

Proof of Lemma 16. Note that $\succsim^*$ is a preference on $\mathcal{T}$. First we show that $\succsim^*$ is a qualitative probability. To see this, $(\Omega, \emptyset) \succ (D, D^c)$ for $D \in \mathcal{D}$, which implies $\Omega \succ^* \emptyset$. Also by definition of $\mathcal{D}$, we know that $\Omega \succsim^* A \succsim^* \emptyset$ for any $A \in \mathcal{T}$. Furthermore, by Lemma 15,

$A \succ^* B \Leftrightarrow A \cup C \succ^* B \cup C$ if $(A \cup B) \cap C = \emptyset$ and $A, B, C \in \mathcal{T}$. Therefore, $\succsim^*$ is a qualitative probability on $\mathcal{T}$.

Next, we want to show that $\succsim^*$ satisfies the condition that $A \succ^* B \Rightarrow$ there is a partition $\{P_n\}_{n=1}^m \subset \mathcal{T}$ such that $A \succ^* B \cup P_n$ for each n. This is immediately implied by **Axiom 5 (Triviality Continuity)**. Thus, by Savage (1954), there is a unique convex ranged finitely additive probability measure μ defined over $\mathcal{T}$ that represents $\succsim^*$.

Finally, we want to show that μ is countably additive. Consider $\{A_n\}_{n \geq 1} \subseteq \mathcal{T}$ such that $A_n \subseteq A_{n+1}$ and $\lim_n A_n \in \mathcal{T}$. It suffices to show that $\lim_n \mu(A_n) = \mu(\lim_n A_n)$. Easy to see that $\lim_n \mu(A_n) \leq \mu(\lim_n A_n)$ since $A_n \subseteq \lim_n A_n$ for all n. Suppose now that $\lim_n \mu(A_n) < \mu(\lim_n A_n)$. Then there exists $\epsilon > 0$ small enough such that $\lim_n \mu(A_n) < \mu(\lim_n A_n) - \epsilon$. For one case, suppose $\mu(A_n) = 0$ for all n, then $\mu(\lim_n A_n) > 0$ implies that $(\lim_n A_n \cup D, \lim_n A_n^c \cap D) \succ (D, D^c)$ for any $D \in \mathcal{D}$ since μ represents $\succsim^*$. Thus, by **Axiom 6 (Dominance Continuity)**, there is some A_k such that $(A_k \cup D, A_k^c \cap D^c) \succ (D, D^c)$. Then $A_k \succ^* \emptyset$ and thus $\mu(A_k) > 0$, which is a contradiction. Now consider the other case where for some n, $\mu(A_n) > 0$. Pick ϵ small enough such that there is some A_k with $\mu(A_k) > \epsilon$. Since μ is convex ranged, we can find $B \subseteq A_k$ with $\mu(B) = \epsilon$. As a result, $\mu((\lim_n A_n) \backslash B) = \mu(\lim_{n \geq k}(A_n \backslash B)) = \mu(\lim_n A_n) - \epsilon > \lim_n \mu(A_n) \geq \mu(A_n)$ for all n. Again by **Axiom 6 (Dominance Continuity)**, for each n, there exists some $m \geq k$ with $A_m \backslash B \succ^* A_n$. Pick A_n such that $\mu(A_n) > \lim_t \mu(A_t) - \frac{\epsilon}{2}$. Thus $A_m \backslash B \succ^* A_n \Rightarrow \mu(A_m) - \epsilon > \mu(A_n) > \lim_t \mu(A_t) - \frac{\epsilon}{2} \geq \mu(A_m) - \frac{\epsilon}{2}$, which is a contradiction. To conclude, $\lim_n \mu(A_n) = \mu(\lim_n A_n)$ and μ is countably additive. This completes the proof. □

Making use of the representation lemma of qualitative probability, we can characterize the collections of null events, trivial events and intractable events, just like in Lemma 8.

Lemma 17. *(1) $N \in \mathcal{N} \Leftrightarrow N \in \mathcal{T}$ and $\mu(N) = 0$; (2) $A \in \mathcal{V} \Leftrightarrow \mu_*(A) = 0$; (3) $D \in \mathcal{D} \Leftrightarrow D, D^c \in \mathcal{V}$.*

Proof of Lemma 17. For (1), by Lemma 16, $N \in \mathcal{N} \Leftrightarrow (N, N^c) \sim \Omega$, $N \sim^* \emptyset \Leftrightarrow N \in \mathcal{T}$ and $\mu(N) = 0$. (2) is trivial by definition and (1). The necessity of (3) is given in Lemma 12.

For the sufficiency of (3), by **Axiom 2 (Strictness)**, there exists $D \in \mathcal{D}$ and thus $D, D^c \in \mathcal{V}$. Suppose $D_1, D_1^c \in \mathcal{V}$, then both $\{D, D^c\}$ and $\{D_1, D_1^c\}$ are nowhere verifiable 2-partition of Ω. By Lemma 13, $(D, D^c) \sim (D_1, D_1^c)$ and by definition $D_1 \in \mathcal{D}$. □

Step 2: Represent the preference over binary problems. Now we construct a representation for the preference restricted to binary problems. Lemma 18 guarantees that each binary problem can be identified with a trivial event.

Lemma 18. *For any $A \subseteq \Omega$ and $B \in \mathcal{T}$ such that $\mu(B) = \mu_*(A) + \mu_*(A^c)$, then $(A, A^c) \sim (B \cup D, B^c \cap D^c)$ for any $D \in \mathcal{D}$.*

Proof of Lemma 18. First, since μ is convex ranged and countably additive, we can define $E[\cdot]$ for all subsets of Ω where $E[A] \subseteq A$, $E[A] \in \mathcal{T}$ and $\mu(E[A]) = \mu_*(A)$. By **Axiom 3 (Triviality Improvement)** and $E[A], E[A^c] \in \mathcal{T}$, we know $(A, A^c) \sim (A \cup E[A^c], A^c \backslash E[A^c])$. Then $A^c \backslash E[A^c] \in \mathcal{V}$ and $\mu_*(A) + \mu_*(A^c) = \mu_*(A \cup E[A^c])$. Thus, WLOG, we can focus on A where $A^c \in \mathcal{V}$ and $\mu_*(A) + \mu_*(A^c) = \mu_*(A)$.

If $\mu_*(A) = 0$, then $A \in \mathcal{D}$ and we can choose $B = \emptyset$ so that the result trivially holds.

Now we suppose $\mu_*(A) = \mu(E[A]) > 0$. Since $E[A] \in \mathcal{T}$, we can define a *probability space restricted to* $E[A]$ as $(E[A], \mathcal{T}^{E[A]}, \mu^{E[A]})$ where $\mathcal{T}^{E[A]} \equiv \{B \cap E[A] : B \in \mathcal{T}\}$ and $\mu^{E[A]}(C) = \mu(C)/\mu(E[A])$ for all $C \in \mathcal{T}^{E[A]}$. Clearly $\mathcal{T}^{E[A]}$ is a σ-algebra and $\mu^{E[A]}$ is convex ranged and countably additive. Then by Lemma 7, we can find D_A diffuse in $(E[A], \mathcal{T}^{E[A]}, \mu^{E[A]})$, that is $\mu_*^{E[A]}(D_A) + \mu_*^{E[A]}(D_A^c) = 0$. Then $\mu_*(D_A) + \mu_*(D_A^c) = 0$.

Define $D = D_A \cup (A \backslash E[A])$. We claim that $D \in \mathcal{D}$. By Lemma 17, it suffices to show $\mu_*(D) + \mu_*(D^c) = 0$. For $\mu_*(D)$, notice that $\mu_*(D_A) = \mu_*(A \backslash E[A]) = 0$ and $D_A \subset E[A]$. For any $E \in \mathcal{T}$ and $E \subseteq D$, $E \cap D_A = E \cap E[A] \in \mathcal{T}$ and $E \cap (A \backslash E[A]) = E \backslash (E \cap D_A) \in \mathcal{T}$. Then $\mu(E) = \mu(E \cap D_A) + \mu(E \cap (A \backslash E[A])) = 0$ for all such E and thus $\mu_*(D)$. Similarly, $D^c = (E[A] \backslash D_A) \cup A^c$ and $\mu_*(A^c) = \mu_*(E[A] \backslash D_A) = 0$. Then repeat the argument above and we can get $\mu_*(D^c) = 0$. Thus $D \in \mathcal{D}$.

Finally, $(A, A^c) = (E[A] \cup D, E[A]^c \cap D^c)$ and $\mu(B) = \mu_*(A) + \mu_*(A^c) = \mu(E[A])$. Since $B, E[A] \in \mathcal{T}$, by Lemma 16, $(E[A] \cup D, E[A]^c \cap D^c) \sim (B \cup D, B^c \cap D^c)$. Thus $(A, A^c) \sim (B \cup D, B^c \cap D^c)$. This completes the proof. □

Thus, we a corollary of Lemma 18, which states that to compare two binary problems, it suffices to compare the sum of inner measures of the two alternatives.

Corollary 2. $(A, A^c) \succsim (B, B^c) \Leftrightarrow \mu_*(A) + \mu_*(A^c) \geq \mu_*(B) + \mu_*(B^c)$.

We close this step by showing the completeness of the probability measure μ.

Lemma 19. *μ is complete. That is, $A \in \mathcal{T}$ and $\mu(A) = 0 \Rightarrow C \in \mathcal{T}, \forall C \subseteq A$.*

Proof of Lemma 19. Suppose that $A \in \mathcal{T}$ and $\mu(A) = 0$, then $(A, A^c) \sim \Omega$ and $\mu_*(A) + \mu_*(A^c) = 1$. Then for any $C \subseteq A$, $\mu_*(C) = 0$ and $\mu_* C^c \geq \mu_*(A^c) = 1$. Thus, by Corollary2, $(C, C^c) \sim (A, A^c) \sim \Omega$ and $C \in \mathcal{T}$. This implies μ is complete. □

Step 3: Extend the representation to general problems.

Definition 11. *A problem $\hat{P} = (P_1, ..., P_{\frac{n(n+1)}{2}})$ is said to be a **fundamental representation** of the problem associated with $(A_1, ..., A_n)$ if for each $1 \leq k \leq n$, $\{P_{\frac{k(k-1)}{2}+1}, ..., P_{\frac{k(k+1)}{2}}\}$ is a nowhere verifiable k-partition of some A_k, where $(A_1, ..., A_n)$ is a n-partition of Ω with $(A_1, ..., A_n) \subset \mathcal{T}$ and for $k > 1$, $A_k \in \mathcal{N} \Rightarrow A_k = \emptyset$.*

One advantage of fundamental representation is the simplicity in the expression of the induced lottery, as is shown in Lemma 20.

Lemma 20. *Suppose that $\hat{P} = (P_1, ..., P_{\frac{n(n+1)}{2}})$ is a fundamental representation associated with $(A_1, ..., A_n)$, then $\Gamma_\mu(\hat{P}) = (\mu(A_1), ..., \mu(A_n))$.*

Proof of Lemma 20. By definition of nowhere verifiable partitions, for $k = 1, ..., n$, $\cup_{A \in \mathcal{D}_k(\hat{P})} E[A] = \cup_{i=1}^{k} A_i$. Then by Equation (5), $\Gamma_\mu(\hat{P})_k = \mu(\cup_{A \in \mathcal{D}_k(\hat{P})} E[A]) - \mu(\cup_{A \in \mathcal{D}_{k-1}(\hat{P})} E[A]) = \mu(A_k)$. Thus, $\Gamma_\mu(\hat{P}) = (\mu(A_1), ..., \mu(A_n))$. □

Then we show that each problem is indifferent to some problem with fundamental representation. Before that, we need some lemmas about the robustness of the indifference relation and Γ_μ with respect to certain operations over problems. The following lemma states that the refined problem induced by a trivial event is indifferent to the original problem.

Lemma 21. *For problem $\hat{P} = (P_1, ..., P_n)$ and $B \in \mathcal{T}$ and $\hat{Q} = \Phi(\hat{P}, B)$, then $\hat{P} \sim \hat{Q}$ and $\Gamma_\mu(\hat{P}) = \Gamma_\mu(\hat{Q})$.*

Proof of Lemma 21. By the same argument as Lemma 9, we know $\Gamma_\mu(\hat{P}) = \Gamma_\mu(\hat{Q})$. Then it suffices to show $\hat{P} \sim \hat{Q}$. By **Axiom 3 (Triviality Improvement)** and $B \in \mathcal{T}$, $\hat{P} = (P_1, ..., P_n) \sim (P_1 \backslash B, B \cap P_1, P_2, ..., P_n) \sim (P_1 \backslash B, P_2 \backslash B, B \cap P_1, B \cap P_2, P_3, ..., P_n) \sim \cdots \sim (P_1 \backslash B, ..., P_n \backslash B, B \cap P_1, ..., B \cap P_n) = \hat{Q}$. This completes the proof. □

The following lemma states that invariance also holds for combining two nowhere-verifiable n-partitions element by element.

Lemma 22. *For two problems $\hat{P} = (P_1, ..., P_n)$ and $\hat{Q} = (Q_1, ...Q_k, P_{2k+1}, ..., P_n)$, where $Q_i = P_i \cup P_{k+i}$ for $i = 1, ..., k$, $(P_1, ..., P_k)$ is a nowhere-verifiable k-partition of $A_1 \in \mathcal{T}$ and $(P_{k+1}, ..., P_{2k})$ is a nowhere-verifiable k-partition of $A_2 \in \mathcal{T}$ with $A_1 \cap A_2 = \emptyset$, then (1). $(Q_1, ..., Q_k)$ is a nowhere-verifiable k-partition of $A_1 \cup A_2$; (2). $\hat{P} \sim \hat{Q}$; (3). $\Gamma_\mu(\hat{P}) = \Gamma_\mu(\hat{Q})$.*

Proof of Lemma 22. (1). To prove that $(Q_1, ..., Q_k)$ is a nowhere-verifiable k-partition of $A_1 \cup A_2$, it suffices to show that for any $i < k$, for any A as the union of some i alternatives among $(Q_1, ..., Q_k)$, $\mu_*(A) = 0$. Suppose by contradiction that $\exists B = \cup_{l=1}^{i} Q_{t_l}$ where $\{Q_{t_1}, ..., Q_{t_i}\} \subset \{Q_1, ..., Q_k\}$, $i < k$ and $\mu(E[B]) > 0$. Since $A_1, A_2 \in \mathcal{T}$, $E[B] \cap A_1, E[B] \cap A_2 \in \mathcal{T}$. WLOG, suppose $\mu(E[B] \cap A_1) > 0$. Notice that $\mu(E[B] \cap A_1) \subseteq \cup_{j=1}^{i} P_{t_j}$ for some $t_1, ..., t_i \in 1, 2, ..., k$ and $i < k$. This means that $\sup_{E \subseteq \cup_{j=1}^{i} P_{t_j} : E \in \mathcal{T}} \mu(E) > 0$ and this leads to a contradiction. Thus, $(Q_1, ..., Q_k)$ is a nowhere-verifiable k-partition of $A_1 \cup A_2$.

(2). For $i = 1, ..., k$, $P_i = Q_i \cap A_1$ and $P_{k+i} = Q_i \backslash A_1$. Since $A_1 \in \mathcal{T}$, by **Axiom 3 (Triviality Improvement)**, $\hat{Q} = (Q_1, ...Q_k, P_{2k+1}, ..., P_n) \sim (Q_1 \cap A_1, Q_1 \backslash A_1, Q_2 \cap A_1, Q_2 \backslash A_1, ..., Q_k \cap A_1, Q_k \backslash A_1, P_{2k+1}, ..., P_n) = (P_1, ..., P_n) = \hat{P}$.

(3). By the same argument as Lemma 9, for $B \in \mathcal{T}$, $\Gamma_\mu(\hat{P}) = \mu(B)\Gamma_\mu^B(\hat{P}) + \mu(B^c)\Gamma_\mu^{B^c}(\hat{P})$. Since $A_1 \cup A_2 \in \mathcal{T}$ and $\hat{P}$, $\hat{Q}$ agree on $(A_1 \cup A_2)^c$, $\Gamma_\mu^{A_1^c \cap A_2^c}(\hat{P}) = \Gamma_\mu^{A_1^c \cap A_2^c}(\hat{Q})$. If $\mu(A_1) + \mu(A_2) = 0$, then clearly $\Gamma_\mu(\hat{P}) = \Gamma_\mu^{A_1^c \cap A_2^c}(\hat{P}) = \Gamma_\mu^{A_1^c \cap A_2^c}(\hat{Q}) = \Gamma_\mu(\hat{Q})$. If instead $\mu(A_1) + \mu(A_2) > 0$, then by definition of Γ_μ and $(P_1, ..., P_k)$, $(P_{k+1}, ..., P_{2k})$, $(Q_1, ..., Q_k)$ are a nowhere-verifiable k-partition of $A_1, A_2, A_1 \cup A_2$ respectively, $\Gamma_\mu^{A_1 \cup A_2}(\hat{P})_k = 1 = \Gamma_\mu^{A_1 \cup A_2}(\hat{Q})_k$. Thus, $\Gamma_\mu(\hat{P}) = \mu(A_1 \cup A_2)\Gamma_\mu^{A_1 \cup A_2}(\hat{P}) + (1 - \mu(A_1 \cup A_2))\Gamma_\mu^{A_1^c \cap A_2^c}(\hat{P}) = \mu(A_1 \cup A_2)\Gamma_\mu^{A_1 \cup A_2}(\hat{Q}) + (1 - \mu(A_1 \cup A_2))\Gamma_\mu^{A_1^c \cap A_2^c}(\hat{Q}) = \Gamma_\mu(\hat{Q})$. We are done with the proof. □

Lemma 23. *For each problem $\hat{P} = (P_1, ..., P_n)$, there is a problem $\hat{Q}$ with fundamental representation such that $\hat{P} \sim \hat{Q}$ and $\Gamma_\mu(\hat{P}) = \Gamma_\mu(\hat{Q})$.*

Proof of Lemma 23. First, by Lemma 1, for any $k = 1, \dots, n$, $A, B \in \mathcal{D}_k(\hat{P})$ and $A \neq B$, denote $F_k[A] \equiv E[A] \backslash (\cup_{i=1}^{k} E[A^i])$, then we know $\mu(F_k[A] \cap F_k[B]) = 0$. We can remove $F_k[A] \cap F_k[B]$ from $E[B]$ to get $E[B']$ so that $\mu(F_k[B]) = \mu(F_k[B'])$ and $F_k[A] \cap F_k[B'] = \emptyset$. In this way, WLOG, we can assume for any $k = 1, \dots, n$, $A, B \in \mathcal{D}_k(\hat{P})$ and $A \neq B$, $F_k[A] \cap F_k[B] = \emptyset$. This can also hold for $A \in \mathcal{D}_k(\hat{P})$ and $B \in \mathcal{D}_{k'}(\hat{P})$ with $k \neq k'$.

Now, again by Lemma 1, set $A_k = \cup_{A \in \mathcal{D}_k(\hat{P})} F_k[A]$ for all $k = 1, ..., n$. . Then $\forall k, A_k \in \mathcal{T}$ and $(A_1, ..., A_n)$ is a partition of Ω. Then denote $\hat{P}^1 = \Phi(\hat{P}, A_1)$, $\hat{P}^2 = \Phi(\hat{P}^1, A_2), \cdots, \hat{P}^n = \Phi(\hat{P}^{n-1}, A_n)$. Apply Lemma 21 repeatedly and we know $\hat{P}^n \sim \hat{P}$ and $\Gamma_\mu(\hat{P}^n) = \Gamma_\mu(\hat{P})$.

Notice that for any $A \in \mathcal{D}_k(\hat{P})$ and $B \in \mathcal{D}_{k'}(\hat{P})$ and $A \neq B$, we set $F_k[A] \cap F'_k[B] = \emptyset$. By definition of $F[\cdot]$, for each $A \in \mathcal{D}_k(\hat{P})$, denote $A = \cup_{j=1}^k P_{t_j}$ for some $(t_1, ..., t_k)$, then $(F_k[A] \cap P_{t_j})_{j=1}^k$ must be a nowhere verifiable k-partition of $F_k[A]$ and by our construction, $(F_k[A] \cap P_{t_j})_{j=1}^k \subset \hat{P}^n$. In other words, $\{F_k(A)\}_{k=1,...,n, A \in \mathcal{D}_k(\hat{P})}$ forms a partition of Ω with each block in $\mathcal{T}$ (some of the blocks might be empty) and for each such $F_k(A)$, there exists a subcollection of $\hat{P}^n$ as a nowhere verifiable k-partition of $F_k[A]$. Moreover, for $k > 1$, if $\mu(F_k[A]) = 0$, then $F_k[A] = \emptyset$.

Now the only difference between $\hat{P}^n$ and a fundamental representation is that for each k, there might be multiple nowhere verifiable k-partitions in $\hat{P}^n$, while this number is limited to at most 1 in the fundamental representation. This is where Lemma 22 kicks in. By applying Lemma 22 repeatedly for each $k = 1, ..., n$, we can combine all possible nowhere verifiable k-partitions element by element to get $\hat{Q}$ such that for each $k = 1, ..., n$, there is only one nowhere verifiable k-partition of $A_k = \cup_{A \in \mathcal{D}_k(\hat{P})} F_k[A]$ in $\hat{Q}$ and for $k > 1$, $A_k \in \mathcal{N} \Rightarrow \mu(F_k[A]) = 0, \forall A \in \mathcal{D}_k(\hat{P}) \Rightarrow F_k[A] = \emptyset, \forall A \in \mathcal{D}_k(\hat{P}) \Rightarrow A_k = \emptyset$. That is, $\hat{Q} = (Q_1, .., Q_{\frac{n(n+1)}{2}})$ is a **fundamental representation** of the problem associated with $(A_1, ..., A_n)$. Moreover, by Lemma 22, $\hat{Q} \sim \hat{P}^n \sim \hat{P}$ and $\Gamma_\mu(\hat{Q}) = \Gamma_\mu(\hat{P}^n) = \Gamma_\mu(\hat{P})$. This completes the proof. □

Now we want to show that the the original preferences can be represented by an induced preferences over $\mathcal{L}(\mathbb{N})$. The basic idea is similar to the probabilistic sophistication in Savage (1954) and Machina and Schmeidler (1992). Concretely, one could consider the outcome space as the set of difficulty levels $\mathbb{N}$ such that $k \succ k' \Leftrightarrow k < k'$. In this way, our **Axiom 4 (Monotone Independence)** reduces to Axiom P4* in Machina and Schmeidler (1992). The following lemmas implement the idea formally.

Lemma 24 states the inverse of the first part in Lemma 22. That is, dividing a nowhere-verifiable $k-$partition with respect to a trivial event will result in two nowhere-verifiable $k-$partitions. In this way, k is a good measure of difficulty for $A \in \mathcal{T}$ if A admits a nowhere-verifiable $k-$partition.

Lemma 24. *For a problem $\hat{P} = (P_1, ..., P_n)$ where $(P_1, ..., P_k)$ is a nowhere-verifiable $k-$partition of $A \in \mathcal{T}$, then for any $B \in \mathcal{T}$, $(P_1 \cap B, ..., P_k \cap B)$ is a nowhere-verifiable $k-$partition of $A \cap B$.*

Proof of Lemma 24. Suppose by contradiction that $(P_1 \cap B, ..., P_k \cap B)$ is not a nowhere-verifiable $k-$partition of $A \cap B$, then there exists a proper subcollection $\{P_{t_j} \cap B\}_{j=1}^{i}, i < k$ such that $\cup_{j=1}^{i}(P_{t_j} \cap B) \notin \mathcal{V}$. This implies that $\cup_{j=1}^{i} P_{t_j} \notin \mathcal{V}$ as $\mu_*(\cup_{j=1}^{i} P_{t_j}) \geq \mu_*(\cup_{j=1}^{i}(P_{t_j} \cap B)) > 0$. Since $\{P_{t_j}\}_{j=1}^{i}$ is a proper subcollection of $(P_1, ..., P_k)$, this contradicts with $(P_1, ..., P_k)$ is a nowhere-verifiable $k-$partition of A. We are done. □

The following lemma shows that exchanging difficulty level on on equally likely events leaves the individual indifferent.

Lemma 25. *For $A, B \in \mathcal{T}$, $\mu(A) = \mu(B)$ and $A \cap B = \emptyset$, $\hat{P} =^{m,n} (A, B, C)$, $\hat{Q} =^{n,m} (A, B, C)$ and $\hat{P} \approx^{m+n} \hat{Q}$, then $\hat{P} \sim \hat{Q}$ and $\Gamma_\mu(\hat{P}) = \Gamma_\mu(\hat{Q})$.*

Proof of Lemma 25. First, suppose $m = n$, then by Lemma 22, we can combine the nowhere verifiable $m-$partitions of A and B block by block to get a nowhere verifiable $m-$partition of $A \cup B$. Denote the new problem as $\hat{P}^1$ and $\hat{Q}^1$ separately. Then $\hat{P}^1 \sim \hat{P}$ and $\hat{Q}^1 \sim \hat{Q}$. Now for $\hat{P}^1$ and $\hat{Q}^1$, both $(P_1^1, ..., P_m^1)$ and $(Q_1^1, ..., Q_m^1)$ are nowhere verifiable $n-$partitions of $A \cup B$ and $\hat{P}^1 \approx^n \hat{Q}^1$. By Lemma 13, $\hat{P}^1 \sim \hat{Q}^1$ and thus $\hat{P} \sim \hat{Q}$.

Now, WLOG, suppose $m > n$. Consider $\hat{P}^2 = (P_1, ..., P_{m+n}, C)$, $\hat{Q}^2 = (Q_1, ..., Q_{m+n}, C)$, then $\hat{P} =^{m,n} (A, B, C) =^{m,n} \hat{P}^2, \hat{Q} =^{m,n} (B, A, C) =^{m,n} \hat{Q}^2$. and $\hat{P} \approx^{m+n} \hat{Q}$, $\hat{P}^2 \approx^{m+n} \hat{Q}^2$. By **Axiom 4 (Monotone Independence)**, $\hat{P} \sim \hat{Q} \Leftrightarrow \hat{P}^2 \sim \hat{Q}^2$.

Since $A, B \in \mathcal{T}$, by Lemma 7, there exist a nowhere verifiable $2-$partition (R_1, R_2) of A and a nowhere verifiable $2-$partition (S_1, S_2) of B. Define $\hat{R} = (R_1, R_2, B, C)$, $\hat{S} = (S_1, S_2, A, C)$. Then $\hat{P}^2 =^{m,n} (A, B, C), \hat{R} =^{2,1} (A, B, C), \hat{Q}^2 =^{m,n} (B, A, C), \hat{S} =^{2,1} (B, A, C)$ and $\hat{P}^2 \approx^{m+n} \hat{Q}^2$, $\hat{R} \approx^{2+1} \hat{S}$. Again by **Axiom 4 (Monotone Independence)**, $\hat{P}^2 \sim \hat{Q}^2 \Leftrightarrow \hat{R} \sim \hat{S}$. Now it suffices to show $\hat{R} \sim \hat{S}$.

Since $A, B, C \in \mathcal{T}$, by **Axiom 3 (Triviality Improvement)**, $\hat{R} \sim (R_1, R_2 \cup B \cup C) = (R_1, R_1^c)$ and $\hat{S} \sim (S_1, S_2 \cup A \cup C) = (S_1, S_1^c)$. By definition, $R_1, R_2, S_1, S_2 \in \mathcal{V}$. Then $\mu_*(R_1) = \mu_*(S_1) = 0$ and $\mu_*(R_1^c) = \mu(B) + \mu(C) = \mu(A) + \mu(C) = \mu_*(S_1^c)$. This implies $\mu_*(R_1) + \mu_*(R_1^c) = \mu_*(S_1) + \mu_*(S_1^c)$. By Lemma 2, $(S_1, S_1^c) \sim (R_1, R_1^c)$. Thus $\hat{R} \sim \hat{S}$.

Finally, to see $\Gamma_\mu(\hat{P}) = \Gamma_\mu(\hat{Q})$, just notice that we have $\mu(A) = \mu(B)$, $\Gamma_\mu^A(\hat{P}) = \Gamma_\mu^B(\hat{Q})$, $\Gamma_\mu^B(\hat{P}) = \Gamma_\mu^A(\hat{Q})$ and $\Gamma_\mu^C(\hat{P}) = \Gamma_\mu^C(\hat{Q})$. Then by a similar argument in Lemma 9, we know $\Gamma_\mu(\hat{P}) = \Gamma_\mu(\hat{Q})$. This completes the proof. □

Lemma 26 shows that problems inducing the same lottery are indifferent.

Lemma 26. *Suppose that $\hat{P}$ and $\hat{Q}$ with $\Gamma_\mu(\hat{P}) = \Gamma_\mu(\hat{Q})$, then $\hat{P} \sim \hat{Q}$*

Proof of Lemma 26. By Lemma 23, we know that there exist problems $\hat{P}_F$ and $\hat{Q}_F$ with fundamental representations associated with $(A_1, ..., A_n)$ and $(B_1, ..., B_n)$ respectively such that $\hat{P} \sim \hat{P}_F$, $\hat{Q} \sim \hat{Q}_F$ and $\Gamma_\mu(\hat{P}_F) = \Gamma_\mu(\hat{P}) = \Gamma_\mu(\hat{Q}) = \Gamma_\mu(\hat{Q}_F)$. By Lemma 20, $\mu(A_k) = \mu(B_k)$ for all k. Now it suffices to show $\hat{P}_F \sim \hat{Q}_F$.

For each problem $\hat{R}$ with fundamental representation, by definition, we can identify exactly a partition $(C_1, ..., C_n)$ of Ω such that $\hat{R}$ is a fundamental representation associated with $(C_1, ..., C_n)$. Formally, we denote $\Theta_{\hat{R}}(i) = C_i$ for $i = 1, ..., n$, which means the event has a nowhere verifiable $i-$partition within $\hat{R}$. Intuitively, the problem restricted to $\Theta_{\hat{R}}(i)$ has difficulty level i.

Now we want to construct a sequence of problems with fundamental representation $\hat{P}_F = \hat{S}^0, \hat{S}^1, ..., \hat{S}^{n-1}$ such that for $1 \leq i \leq n-1$, $\hat{S}^i$ is obtained from $\hat{S}^{i-1}$ by exchanging difficulty levels on equally likely events so that $\hat{S}^{i-1} \sim \hat{S}^i$, $\Gamma_\mu(\hat{S}^{i-1}) = \Gamma_\mu(\hat{S}^i)$, and such that $\Theta_{\hat{S}^i}(j) = B_j$ for all $1 \leq j \leq i$. Notice that $\Theta_{\hat{S}^{n-1}}(j) = B_j$ for all $1 \leq j \leq n-1$ naturally implies $\Theta_{\hat{S}^{n-1}}(n) = B_n$. Moreover, by applying Lemma 13 repeatedly, we know if two problems with fundamental representation are associated with the same partition $(C_1, ..., C_n)$, then the two problems are indifferent. That is, $\hat{S}^{n-1} \sim \hat{Q}_F$. In the following, we will prove that such inductive construction is feasible.

Construction of $\hat{S}^1$ from $\hat{P}_F = \hat{S}^0$: Since $\mu(A_1) = \mu(\Theta_{\hat{S}^0}(1)) = \mu(B_1)$, $\mu(\Theta_{\hat{S}^0}(1) - B_1) = \mu(B_1 - \Theta_{\hat{S}^0}(1))$. Consider the refined problem induced by $\hat{S}^0$ on B_1 as $\hat{R}^0 = \Phi(\hat{S}^0, B_1)$, by Lemma 21, $\hat{R}^0 \sim \hat{S}^0$ and $\Gamma_\mu(\hat{S}^0) = \Gamma_\mu(\hat{R}^0)$. Moreover, by Lemma 24, for each $k = 1, ..., n$, $\{S^0_{\frac{k(k-1)}{2}+1} \cap B_1, ..., S^0_{\frac{k(k+1)}{2}} \cap B_1\}$ is a a nowhere verifiable k-partition of $\Theta_{\hat{S}^0}(k) \cap B_1$ and $\{S^0_{\frac{k(k-1)}{2}+1} \cap B_1^c, ..., S^0_{\frac{k(k+1)}{2}} \cap B_1^c\}$ is a a nowhere verifiable k-partition of $\Theta_{\hat{S}^0}(k) \cap B_1^c$. Actually, all those blocks are elements in $\hat{R}^0$ by definition of $\Phi(\cdot)$.

Now, $\{\Theta_{\hat{S}^0}(k) \cap B_1\}_{k=2}^n$ is a partition of $B_1 - \Theta_{\hat{S}^0}(1)$ and we want to construct a corresponding partition in $\Theta_{\hat{S}^0}(1) - B_1$. Since μ is convex-ranged, we can find $\{E^1_k\}_{k=2}^n \subset \mathcal{T}$ as a partition of $\Theta_{\hat{S}^0}(1) - B_1$ such that for each $2 \leq k \leq n$, $\mu(E^1_k) = \mu(\Theta_{\hat{S}^0}(k) \cap B_1)$. Again, by Lemma 7, we can construct nowhere verifiable $k-$ partition of each E^1_k for each $2 \leq k \leq n$. In this way, we constructed events within $\Theta_{\hat{S}^0}(1) - B_1$ with the same likelihood and the same difficulty as $\{\Theta_{\hat{S}^0}(k) \cap B_1\}_{k=2}^n$ within $B_1 - \Theta_{\hat{S}^0}(1)$. Then, we can apply Lemma 25 repeatedly to exchange blocks within E^1_k and $\Theta_{\hat{S}^0}(k) \cap B_1$ for each $k = 2, ..., n$. In each step, the problems are indifferent and they induce the same lottery. Now consider the ultimate problem as $\hat{T}^0$. Clearly, for $\hat{T}^0$, each event in $\{\Theta_{\hat{S}^0}(k) \cap B_1\}_{k=1}^n$ now has difficulty level 1 (that is, admits nowhere-verifiable $1-$partition within $\hat{T}^0$), since their difficulty has been exchanged with $\{E_k^1\}_{k=2}^n \cup \{\Theta_{\hat{S}^0}(1) \cap B_1\}$, which originally has difficulty level 1 within $\hat{S}^0$.

By Lemma 23, we can find a problem $\hat{H}$ with fundamental representation associated with $\{\Theta_{\hat{H}}(k)\}_{k=1}^n$ such that $\hat{H} \sim \hat{T}^0 \sim \hat{S}^0$ and $\Gamma_\mu(\hat{H}) = \Gamma_\mu(\hat{T}^0) = \Gamma_\mu(\hat{S}^0)$. Moreover, by the argument above, we know $\Theta_{\hat{H}}(1) = \cup_{k=1}^n \Theta_{\hat{S}^0}(k) \cap B_1 = B_1$. In this way, define $\hat{S}^1 = \hat{H}$, we get $\hat{S}^0 \sim \hat{S}^1$, $\Gamma_\mu(\hat{S}^0) = \Gamma_\mu(\hat{S}^1)$, and such that $\Theta_{\hat{S}^1}(1) = B_1$. This completes the construction for $i = 1$.

Construction of $\hat{S}^i$ from $\hat{S}^{i-1}$, $i = 2, ..., n-1$: From the previous construction, we have: (i) $\hat{S}^{i-1} \sim \hat{P}_F$; (ii) $\Gamma_\mu(\hat{S}^{i-1}) = \Gamma_\mu(\hat{P}_F) = \Gamma_\mu(\hat{Q}_F)$; (iii) $\Theta_{\hat{S}^{i-1}}(j) = B_j$ for all $1 \leq j \leq i-1$. Since $\mu(\Theta_{\hat{S}^{i-1}}(i)) = \mu(B_i)$, $\mu(\Theta_{\hat{S}^{i-1}}(i) - B_i) = \mu(B_i - \Theta_{\hat{S}^{i-1}}(i))$. Consider the problem induced by $\hat{S}^{i-1}$ on B_i as $\hat{R}^{i-1} = \Phi(\hat{S}^{i-1}, B_i)$, by Lemma 21, $\hat{R}^{i-1} \sim \hat{S}^{i-1}$ and $\Gamma_\mu(\hat{S}^{i-1}) = \Gamma_\mu(\hat{R}^{i-1})$. Also, since $\{B_k\}_{k=1}^n$ mutually disjoint, the manipulation will not affect the reduced problem restricted to $\cup_{j=1}^{i-1} B_i$. Moreover, by Lemma 24, for each $k = i, ..., n$, $\{S^{i-1}_{\frac{k(k-1)}{2}+1} \cap B_i, ..., S^{i-1}_{\frac{k(k+1)}{2}} \cap B_i\}$ is a a nowhere verifiable k-partition of $\Theta_{\hat{S}^{i-1}}(k) \cap B_i$ and $\{S^{i-1}_{\frac{k(k-1)}{2}+1} \cap B_i^c, ..., S^{i-1}_{\frac{k(k+1)}{2}} \cap B_i^c\}$ is a a nowhere verifiable k-partition of $\Theta_{\hat{S}^{i-1}}(k) \cap B_i^c$. Actually, all those blocks are elements in $\hat{R}^{i-1}$ by definition of $\Phi(\cdot)$.

Now, $\{\Theta_{\hat{S}^{i-1}}(k) \cap B_i\}_{k=i+1}^n$ is a partition of $B_i - \Theta_{\hat{S}^{i-1}}(i)$ and we want to construct a corresponding partition in $\Theta_{\hat{S}^{i-1}}(i) - B_i$. Since μ is convex-ranged, we can find $\{E_k^i\}_{k=i+1}^n \subset \mathcal{T}$ as a partition of $\Theta_{\hat{S}^{i-1}}(i) - B_i$ such that for each $i+1 \leq k \leq n$, $\mu(E_k^i) = \mu(\Theta_{\hat{S}^{i-1}}(k) \cap B_i)$. Now the new partition will induce a finer problem on $\Theta_{\hat{S}^{i-1}}(i) - B_i$, without changing the DM's preference and the induced lottery. That is, all those events $\{E_k^i\}_{k=i+1}^n$ now admits a nowhere-verifiable $i-$partition and thus has difficulty level i. Again, by Lemma 7, we can refine the original nowhere-verifiable $i-$partition to get a nowhere verifiable $k-$ partition of each E_k^i for each $i+1 \leq k \leq n$. In this way, we constructed events within $\Theta_{\hat{S}^{i-1}}(i) - B_i$ with the same likelihood and the same difficulty as $\{\Theta_{\hat{S}^{i-1}}(k) \cap B_i\}_{k=i+1}^n$ within $B_i - \Theta_{\hat{S}^{i-1}}(i)$. Then, we can apply Lemma 25 repeatedly to exchange blocks within E_k^i and $\Theta_{\hat{S}^{i-1}}(k) \cap B_i$ for each $k = i+1, ..., n$. In each step, the problems are indifferent and they induce the same lottery. Now consider the ultimate problem as $\hat{T}^{i-1}$. Clearly, for $\hat{T}^i - 1$, each event in $\{\Theta_{\hat{S}^{i-1}}(k) \cap B_i\}_{k=i}^n$ now has difficulty level k (that is, admits nowhere-verifiable $k-$partition within $\hat{T}^{i-1}$), since their difficulty has been exchanged with $\{E_k^i\}_{k=i+1}^n \cup \{\Theta_{\hat{S}^{i-1}}(i) \cap B_i\}$, which originally has difficulty level i within $\hat{S}^{i-1}$. Just as a reminder, all the operations do not change the reduced problem restricted to $\cup_{j=1}^{i-1} B_i$.

By Lemma 23, we can find a problem $\hat{H}$ with fundamental representation associated with $\{\Theta_{\hat{H}}(k)\}_{k=1}^{n}$ such that $\hat{H} \sim \hat{T}^{i-1} \sim \hat{S}^{i-1}$ and $\Gamma_\mu(\hat{H}) = \Gamma_\mu(\hat{T}^{i-1}) = \Gamma_\mu(\hat{S}^{i-1})$. Moreover, by the argument above, we know $\Theta_{\hat{H}}(i) = \cup_{k=i}^{n} \Theta_{\hat{S}^{i-1}}(k) \cap B_i = B_i$. Also, $\Theta_{\hat{H}}(j) = B_j$ for $1 \leq j \leq i-1$. In this way, define $\hat{S}^i = \hat{H}$, then we get $\hat{S}^{i-1} \sim \hat{S}^i \sim \hat{P}_F$, $\Gamma_\mu(\hat{S}^{i-1}) = \Gamma_\mu(\hat{S}^i) = \Gamma_\mu(\hat{P}_F)$, and such that $\Theta_{\hat{S}^i}(j) = B_j$ for all $1 \leq j \leq i$. This completes the construction for i.

By induction, for any $\hat{P}$ and $\hat{Q}$ with $\Gamma_\mu(\hat{P}) = \Gamma_\mu(\hat{Q})$, we can find a sequence of problems with fundamental representation $\hat{P}_F = \hat{S}^0, \hat{S}^1, ..., \hat{S}^{n-1}$ such that $\hat{P} \sim \hat{P}_F \sim \hat{S}^1 \sim \cdots \sim \hat{S}^{n-1} \sim \hat{Q}_F \sim \hat{Q}$. By transitivity, $\hat{P} \sim \hat{Q}$ and we complete the proof. $\square$

Lemma 27. Γ_μ *is an onto mapping.*

Proof of Lemma 27. Since μ is convex-ranged, for any lottery over natural numbers $\hat{p} \in \mathcal{L}(\mathbb{N})$ with $p_m = 0$ for all $m > n$, there exists a partition $(A_1, ..., A_n)$ of Ω where $\mu(A_k) = p_k$ for all $k = 1, ..., n$. Since $A_k \in \mathcal{T}$, by Lemma 7, we can find a nowhere verifiable $k-$partition of A_k for each k. Denote the collection of all sets in those partitions as $\hat{P}$ and then $\Gamma_\mu(\hat{P}) = \hat{p}$. Thus $\Gamma_\mu : \mathbb{P}(\Omega) \to \mathcal{L}(\mathbb{N})$ is onto. $\square$

Now we can define a binary relation $\succsim^l$ on $\mathcal{L}(\mathbb{N})$ as follows: $p \succsim^l q$ if and only if there exists problems $\hat{P}$ and $\hat{Q}$ such that $\hat{P} \succsim \hat{Q}$ and $\Gamma_\mu(\hat{P}) = p$, $\Gamma_\mu(\hat{Q}) = q$. By Lemma 27 and Lemma 26, $\succsim^l$ on $\mathcal{L}(\mathbb{N})$ is well-defined. It is clear to see that $\succsim^l$ is complete and transitive. Thus, to show that $\succsim$ is a preference for simplicity, it suffices to show that $\succsim^l$ is monotone.

Lemma 28. $\succsim^l$ *on* $\mathcal{L}(\mathbb{N})$ *is monotone.*

Proof of Lemma 28. To begin with, we need to prove a similar result to Lemma 25. Suppose $m > n$, for $A_1, A_2, B_1, B_2 \in \mathcal{T}$, $\hat{P} =^{m,n} (A_1, B_1, C)$, $\hat{Q} =^{m,n} (A_2, B_2, C)$ and $\hat{P} \approx^{m+n} \hat{Q}$, we claim that $\hat{P} \succ \hat{Q} \Leftrightarrow \mu(A_1) < \mu(A_2)$. To see why this is true, as $A_1, A_2 \in \mathcal{T}$, we can construct $\hat{R} = (R_1, R_2, B_1, C) =^{2,1} (A_1, B_1, C)$ and $\hat{S} = (S_1, S_2, B_2, C) =^{2,1} (A_2, B_2, C)$. Also, $\hat{R} \approx^{2+1} \hat{S}$. Then by **Axiom 4 (Monotone Independence)**, $\hat{P} \succ \hat{Q} \Rightarrow \hat{R} \succ \hat{S}$. Moreover, $\hat{R} = (R_1, R_2, B_1, C) \sim (R_1, R_1^c)$, $\hat{S} = (S_1, S_2, B_2, C) \sim (S_1, S_1^c)$ and by Lemma 2, $\hat{R} \succ \hat{S} \Leftrightarrow \mu_*(R_1) + \mu_*(R_1^c) > \mu_*(S_1) + \mu_*(S_1^c) \Leftrightarrow \mu(B_1) + \mu(C) > \mu(B_2) + \mu(C) \Leftrightarrow \mu(A_1) < \mu(A_2)$. This completes the proof for the claim.

Now consider $\hat{p}, \hat{q} \in \mathcal{L}(\mathbb{N})$ with $\forall n, F_n(\hat{p}) \geq F_n(\hat{q})$. First, if equalities hold for all n, then $\hat{p} = \hat{q}$ and $\hat{p} \sim^l \hat{q}$. Thus, we assume that there exists at least one strict inequality.

By definition of $\succsim^l$, we can find a problem with fundamental representation $\hat{P}$ associated with $(A_1, ..., A_n)$ and $\hat{Q}$ associated with $(B_1, ..., B_n)$ such that $\Gamma_\mu(\hat{P}) = \hat{p}$, $\Gamma_\mu(\hat{Q}) = \hat{q}$. Then to show that $\hat{p} \succ^l \hat{q}$, it suffices to show $\hat{P} \succ \hat{Q}$. Since μ is convex-ranged and $\forall n, F_n(\hat{p}) \geq F_n(\hat{q})$, WLOG, we can assume that $\cup_{j=1}^i B_j \subseteq \cup_{j=1}^i A_j$ for $i = 1, ..., n$. With at least some i, $\mu(\cup_{j=1}^i A_j - \cup_{j=1}^i B_j) > 0$.

Denote $\hat{S}^0 = \hat{P}$. Since $A_1 - B_1 \in \mathcal{T}$, we can construct a nowhere verifiable 2−partition (C_1, C_2) of $A_1 - B_1$. Define $\hat{S}^1$ such that $\hat{S}^1$ agrees with $\hat{S}^0$ over A_1^c while over A_1, $\hat{S}^1$ admits a partition (B_1, C_1, C_2). To simplify the expression, we write $\hat{S}^0$ restricted to A_1 as $(A_1, \emptyset, \emptyset)$ and relabel the partition over A_1^c with subscript from 4 on. In this way, $\hat{S}^0 \approx^3 \hat{S}^1$ and $\hat{S}^0 =^{2,1} (\emptyset, A_1, A_1^c)$, $\hat{S}^1 =^{2,1} (A_1 - B_1, B_1, A_1^c)$. By the above claim, as $\mu(A_1 - B_1) \geq \mu(\emptyset) = 0$, then $\hat{S}^0 \succsim \hat{S}^1$. Strictness holds if $\mu(A_1 - B_1) > 0$. Denote $\hat{S}_F^1$ as the fundamental representation of $\hat{S}^1$. Then $\hat{S}_F^1$ is associated with $(B_1, A_2 \cup (A_1 - B_1), A_3, ..., A_n)$.

By induction, suppose that $\hat{P} \succsim \hat{S}_F^i$, where strictness holds if there exists some $j \leq i$ such that $\mu(\cup_{t=1}^j A_t - \cup_{t=1}^j B_t) > 0$ and $\hat{S}_F^i$ is a fundamental representation associated with $(B_1, B_2, ...B_i, \cup_{j=1}^{i+1} A_j - \cup_{j=1}^i B_j, ..., A_n)$. Since $\cup_{j=1}^i B_j \subseteq \cup_{j=1}^i A_j$ for $i = 1, ..., n$, $B_{i+1} \subseteq \cup_{j=1}^{i+1} A_j - \cup_{j=1}^i B_j$. Since $\cup_{j=1}^{i+1} A_j - \cup_{j=1}^{i+1} B_j \in \mathcal{T}$, and $\hat{S}_F^i$ admits a nowhere verifiable $(i+1)$-partition on $\cup_{j=1}^{i+1} A_j - \cup_{j=1}^{i+1} B_j \in \mathcal{T}$, then by Lemma 7, we can refine this partition to get a nowhere verifiable $(i+2)$-partition on $\cup_{j=1}^{i+1} A_j - \cup_{j=1}^{i+1} B_j \in \mathcal{T}$. Keep every block else fixed and denote the new problem as $\hat{S}^{i+1}$. Then we can add empty sets into the partition and relabeling to get $\hat{S}_F^i =^{i+2,i+1} (\emptyset, \cup_{j=1}^{i+1} A_j - \cup_{j=1}^i B_j, C)$ and $\hat{S}^{i+1} =^{i+2,i+1} (\cup_{j=1}^{i+1} A_j - \cup_{j=1}^{i+1} B_j, B_{i+1}, C)$ and $\hat{S}^i \approx^{2i+3} \hat{S}^{i+1}$. Again by the previous claim and $\mu(\cup_{t=1}^{i+1} A_t - \cup_{t=1}^{i+1} B_t) \geq \mu(\emptyset) = 0$, then $\hat{S}^i \succsim \hat{S}^{i+1}$. Strictness holds if $\mu(\cup_{t=1}^{i+1} A_t - \cup_{t=1}^{i+1} B_t) > 0$. Denote $\hat{S}_F^{i+1}$ as the fundamental representation of $\hat{S}^{i+1}$. Then $\hat{S}_F^1$ is associated with $(B_1, B_2, ...B_i + 1, \cup_{j=1}^{i+2} A_j - \cup_{j=1}^{i+1} B_j, ..., A_n)$.

Continue the induction to get $\hat{S}_F^n$ as a fundamental representation associated with $(B_1, ..., B_n)$. Then $\hat{S}_F^n \sim \hat{Q}$. Thus, $\hat{P} \succsim \hat{S}_F^1 \succsim \cdots \succsim \hat{S}_F^n \sim \hat{Q}$. Since with at least some i, $\mu(\cup_{j=1}^i A_j - \cup_{j=1}^i B_j) > 0$, at least one of the $\succsim$ is strict. Then, $\hat{P} \succ \hat{Q}$, which implies $p \succ^l q$. □

Finally, for the uniqueness, notice that $\mathcal{T}$ is uniquely identified as $A \in \mathcal{T}$ if and only if $(A, A^c) \sim \Omega$. Furthermore, we know that μ is uniquely defined over $\mathcal{T}$ as is standard in Savage (1954). This completes the proof of the theorem. □

9.3 Proofs in Section 4

Proof of Theorem 2. **Sufficiency:** By Theorem 1, it suffices to show that $\succsim^l$ can be represented by a non-constant, mixture continuous function U which respects monotonicity. By **Axiom 2 (Strictness)** and monotonicity of $\succsim^l$, U is naturally non-constant, mixture continuous and respects monotonicity if U represents $\succsim^l$ and $\succsim^l$ is mixture continuous.

Lemma 29. *$\succsim^l$ on $\mathcal{L}(\mathbb{N})$ is mixture continuous.*

Proof of Lemma 29. For any $p, q, r \in \mathcal{L}(\mathbb{N})$, denote $\Lambda_{\succ^l} \equiv \{\lambda \in [0,1] : \lambda p + (1-\lambda)q \succ^l r\}$, $\Lambda_{\prec^l} \equiv \{\lambda \in [0,1] : \lambda p + (1-\lambda)q \prec^l r\}$. It suffices to show both $\Lambda_{\succ^l}$ and $\Lambda_{\prec^l}$ are open in $[0,1]$. WLOG, we can denote $p = (p_1, ..., p_n), q = (q_1, ..., q_n)$ and assume $\Lambda_{\succ^l}$ and $\Lambda_{\prec^l}$ are nonempty as empty set is naturally open.

For all $\lambda^* \in \Lambda_{\succ^l}$, we need to show that there exists $\epsilon > 0$ such that $(\lambda^* - \epsilon, \lambda^* + \epsilon) \cap [0,1] \subseteq \Lambda_{\succ^l}$. By definition, $\lambda^* p + (1-\lambda^*)q \succ^l r$. Since Γ_μ is onto, we can find fundamental representations $\hat{S}$ and $\hat{R}$ such that $\Gamma_\mu(\hat{S}) = \lambda^* p + (1-\lambda^*)q$ and $\Gamma_\mu(\hat{R}) = r$. Moreover, we denote that $\hat{S}$ is a fundamental representation associated with a partition of Ω $(A_1, ..., A_n) \subset \mathcal{T}$ and $\lambda^* p + (1-\lambda^*)q = (\mu(A_1), ..., \mu(A_n))$. By definition of $\succsim^l$, we know $\hat{S} \succ \hat{R}$. Again as Γ_μ is onto, there exist $H \in \mathbb{P}(\Omega)$ such that $\Gamma_\mu(H) = e^n$, where $e^n_n = 1$ and $e^n_{n'} = 0$ for all $n' \neq n$. For simplicity, denote $(\hat{P}, A^c; \hat{R}, A) \equiv (P_1 \backslash A, ..., P_n \backslash A, R_1 \cap A, ..., R_l \cap A)$ as the problem that agrees with $\hat{P}$ on A^c and agrees with $\hat{R}$ on A.

Since $\hat{S} \succ \hat{R}$, by **Axiom Small Event Continuity**, there exists a finite partition of Ω as $\{B_{1,1}, ..., B_{1,n_1}\} \subseteq \mathcal{T}$ such that $(\hat{S}, B^c_{1,i}; \hat{H}, B_{1,i}) \succ \hat{R}$ for all $i = 1, ..., n_1$. If $\mu(A_1) > 0$, then choose B^*_1 from $\{B_{1,1}, ..., B_{1,n_1}\}$ such that $\mu(A_1 \cap B^*_1) > 0$. If $\mu(A_1) = 0$, choose any B^*_1. Denote $\hat{S}^1 = (\hat{S}, B^{*c}_1; \hat{H}, B^*_1)$, then $\hat{S}^1 \succ \hat{R}$.

Again, by **Axiom Small Event Continuity**, there exists a finite partition of Ω as $\{B_{2,1}, ..., B_{2,n_2}\} \subseteq \mathcal{T}$ such that $(\hat{S}^1, B^c_{2,i}; \hat{H}, B_{2,i}) \succ \hat{R}$ for all $i = 1, ..., n_2$. If $\mu(A_2) > 0$, then choose B^*_2 from $\{B_{2,1}, ..., B_{2,n_2}\}$ such that $\mu(A_2 \cap B^*_2) > 0$. If $\mu(A_2) = 0$, choose any B^*_2. Denote $\hat{S}^2 = (\hat{S}^1, B^{*c}_2; \hat{H}, B^*_2) = (\hat{S}, (B^*_1 \cup B^*_2)^c; \hat{H}, B^*_1 \cup B^*_2)$, then $\hat{S}^2 \succ \hat{R}$.

Continue the process, we end up with a sequence $\{B^*_1, ..., B^*_{n-1}\}$ such that $\hat{S}^{n-1} = (\hat{S}, (\cup_{i=1}^{n-1} B^*_i)^c; \hat{H}, \cup_{i=1}^{n-1} B^*_i) \succ \hat{R}$. Now denote $s^{n-1} = \Gamma_\mu(\hat{S}^{n-1})$. Then $s^{n-1} \succ^l r$. Also, for any $i = 1, 2, ..., n-1$, if $\mu(A_i) > 0$, then $s^{n-1}(i) \leq \mu(A_i - B^*_i) < \mu(A_i) = \lambda^* p_i + (1-\lambda^*)q_i$. Otherwise $s^{n-1}(i) = \mu(A_i) = 0$. Choose $\epsilon > 0$ such that $\epsilon < \min_{\{i:\mu(A_i)>0\}}(p_i + (1-\lambda^*)q_i -$

$s^{n-1}(i))$. Consider $\forall\lambda \in (\lambda^*-\epsilon, \lambda^*+\epsilon)\cap[0,1]$, then for any $i = 1, 2, ..., n-1$, when $\mu(A_i) = 0$, $\lambda p_i + (1-\lambda)q_i \geq s^{n-1}(i) = 0$. When $\mu(A_i) > 0$, we have

$$\lambda p_i + (1-\lambda)q_i \geq \lambda^* p_i + (1-\lambda^*)q_i - \epsilon|p_i - q_i| \geq \lambda^* p_i + (1-\lambda^*)q_i - \epsilon > s^{n-1}(i)$$

In this way, we know that $\forall j = 1, ..., n$, $\sum_{i=1}^{j}(\lambda p_i+(1-\lambda)q_i) \geq \sum_{i=1}^{j} s^{n-1}(i)$. By monotonicity and transitivity of $\succsim^l$, $\lambda p + (1-\lambda)q \succsim^l s^{n-1} \succ^l r$. This implies that $\lambda \in \Lambda_{\succ^l}$ and thus $\Lambda_{\succ^l}$ is open in $[0,1]$. A similar argument making use of the other half of **Axiom Small Event Continuity** can show that $\Lambda_{\prec^l}$ is open. This completes the proof. □

Now we just need to show that there exists a utility representation of $\succsim^l$. [11] First, by monotonicity, recall the definition of e^i, we have $e^1 \succ^l e^2 \succ^l \cdots \succ^l e^n \succ^l \cdots$. Define $U(e^i) = -i$ for all $i = 1, 2, ...$ Then for any $p \in \mathcal{L}(\mathbb{N})$, there exists a unique $k \in \mathbb{N}$ such that $e^k \succsim^l p \succ^l e^{k+1}$. To see why this is correct, just notice that if $p_m = 0$ for all $m > n$, then by monotonicity, $e^1 \succsim^l p \succ^l e^{n+1}$ and we can simply identify k as the smallest integer in [1, n] such that $p \succ^l e^{k+1}$.

Consider $0 \leq \lambda < \lambda' \leq 1$, by monotonicity, $\lambda e^{k+1} + (1-\lambda)e^k \succ^l \lambda' e^{k+1} + (1-\lambda')e^k$. Since $\succsim^l$ is mixture continuous, $\Lambda_{\succsim^l} \equiv \{\lambda \in [0,1] : \lambda e^{k+1} + (1-\lambda)e^k \succsim^l p\}$ is closed and bounded. Then we can define the supremum of this set as $\lambda(p)$ and it is clear that $\lambda(p) < 1$. Then $\lambda(p) \in \Lambda_{\succsim^l}$, that is, $\lambda(p)e^{k+1} + (1-\lambda(p))e^k \succsim^l p$. Suppose by contradiction that $\lambda(p)e^{k+1} + (1-\lambda(p))e^k \succ^l p$, then $\lambda(p) \in \Lambda_{\succ^l}$, which is open. As $\lambda(p) < 1$, we can find $\lambda' > \lambda(p)$ and $\lambda' \in \Lambda_{\succ^l} \succ \Lambda_{\succsim^l}$, which contradicts with the definition of $\lambda(p)$. Thus, $\lambda(p)e^{k+1} + (1-\lambda(p))e^k \sim^l p$ with $\lambda(p) \in [0,1)$. Again by monotonicity, such $\lambda(p)$ is unique. In this way, we will define $U(p) = -k - \lambda$. This completes the construction of U.

Then we need to show U represents $\succsim^l$. On the one hand, suppose $p \succsim^l q$, and the above construction gives $p \sim \lambda(p)e^{k(p)+1} + (1-\lambda(p))e^{k(p)}$, $q \sim \lambda(q)e^{k(q)+1} + (1-\lambda(q))e^{k(q)}$. Then either $k(p) < k(q)$ or $k(p) = k(q)$ and $\lambda(p) \leq \lambda(q)$. Since $k(p), k(q) \in \mathbb{N}$ and $\lambda(p), \lambda(q) \in [0,1)$, $U(p) = -k(p) - \lambda(p) \geq -k(q) - \lambda(q) = U(q)$. On the other hand, suppose $p \succ^l q$ and the above construction gives $p \sim \lambda(p)e^{k(p)+1}+(1-\lambda(p))e^{k(p)}$, $q \sim \lambda(q)e^{k(q)+1}+(1-\lambda(q))e^{k(q)}$. Then Then either $k(p) < k(q)$ or $k(p) = k(q)$ and $\lambda(p) < \lambda(q)$. Since $k(p), k(q) \in \mathbb{N}$ and $\lambda(p), \lambda(q) \in [0,1)$, $U(p) = -k(p) - \lambda(p) > -k(q) - \lambda(q) = U(q)$. Thus, U represents $\succsim^l$.

[11] Note that Eilenberg-Debreu Theorem cannot be directly applied here as mixture continuity does not imply continuity. Actually, without strong enough independence axiom, $\succsim^l$ is not necessarily continuous.

Since $\succsim^l$ is non-degenerate, mixture continuous and monotonic, U is non-constant, mixture continuous and respecting monotonicity.

Necessity: Again by Theorem 1, it suffices to show that **Axiom Small Event Continuity** is satisfied. Consider $\hat{P}$, $\hat{Q}$, $\hat{R} \in \mathbb{P}(\Omega)$ such that $\hat{P} \succ \hat{Q}$. WLOG, we assume that $\hat{P}$, $\hat{Q}$, $\hat{R}$ are fundamental representations associated with $\{C_i^P\}_{i=1}^n$, $\{C_i^Q\}_{i=1}^n$ and $\{C_i^R\}_{i=1}^n$ respectively (some of the sets can be empty). Denote $\Gamma_\mu(\hat{P}) = p$, $\Gamma_\mu(\hat{Q}) = q$, $\Gamma_\mu(\hat{R}) = r$. Then by monotonicity, $e^1 \succsim^l p \succ^l q \succsim^l e^n$. Since $\succsim^l$ is mixture continuous as U is mixture continuous, there exists a N large enough such that $(1-1/N)p+(1/N)e^n \succ^l q$ and $p \succ^l (1-1/N)q+(1/N)e^1$. Now, $\{A_i^P \cap A_j^Q\}_{1\le i,j\le n}$ forms a partition of Ω. For each $C_i^P \cap C_j^Q$, since μ is convex-valued, we can find a partition $\{B_{i,j,k}\}_{k=1}^N$ of $C_i^P \cap C_j^Q$ such that $\mu(B_{i,j,k}) = \frac{1}{N}\mu(C_i^P \cap C_j^Q)$ for each k. Define $A_k = \cup_{1\le i,j\le n} B_{i,j,k}$, then $\mu(A_k) = 1/N$ and $\{A_k\}_{k=1}^N$ is a partition of Ω. Then for any $\hat{P}' \in \mathbb{P}(\Omega)$ and $\hat{P}'' = (\hat{P}, A_k^c; \hat{P}', A_k)$, $\hat{Q}'' = (\hat{Q}, A_k^c; \hat{P}', A_k)$, we have $\Gamma_\mu(\hat{P}'') = \mu(A_k^c)\Gamma_\mu^{A_k^c}(\hat{P}) + \mu(A_k)\Gamma_\mu^{A_k}(\hat{P}') = (1-1/N)p + \mu(A_k)\Gamma_\mu^{A_k}(\hat{P}')$, $\Gamma_\mu(\hat{Q}'') = (1-1/N)q + \mu(A_k)\Gamma_\mu^{A_k}(\hat{P}')$.

For the given $\hat{R}$, $\forall k = 1, ..., N$, denote $p^k = \Gamma_\mu((\hat{P}, A_k^c; \hat{R}, A_k))$, $q^k = \Gamma_\mu((\hat{Q}, A_k^c; \hat{R}, A_k))$, then clearly, for any $i = 1, ..., n$, we have

$$\sum_{t=1}^{i} p_t^k = \sum_{t=1}^{i}((1-1/N)p_t + (1/N)\Gamma_\mu^{A_k}(\hat{R})_t) \ge \sum_{t=1}^{i}((1-1/N)p_t + (1/N)e_t^n)$$

$$\sum_{t=1}^{i} q_t^k = \sum_{t=1}^{i}((1-1/N)q_t + (1/N)\Gamma_\mu^{A_k}(\hat{R})_t) \le \sum_{t=1}^{i}((1-1/N)q_t + (1/N)e_t^1)$$

By monotonicity of $\succsim^l$, $p^k \succsim^l (1-1/N)p + (1/N)e^n \succ^l q$ and $p \succ^l (1-1/N)q + (1/N)e^1 \succsim^l q^k$. By transitivity and definition of $\succsim^l$, this implies for any $k = 1, ..., N$, $(\hat{P}, A_k^c; \hat{R}, A_k) \succ \hat{Q}$ and $\hat{P} \succ (\hat{Q}, A_k^c; \hat{R}, A_k)$. Thus **Axiom Small Event Continuity** holds. This completes the proof. □

Proof of Lemma 2. To begin with, we need to argue that Lemmas 5 - 19 still hold when we replace **Axiom 4** by **Axiom 4.1** and **Axiom 4.2**. Actually, among those, only proofs for Lemma 13 and Lemma 15 directly utilize **Axiom 4**. That is, it suffices to show those two lemmas.

For Lemma 13, construct $\hat{P}' = (P_1, ..., P_n, A^c)$, $\hat{Q}' = (Q_1, ..., Q_n, A^c)$. Then $\hat{P} \approx^n \hat{Q}, \hat{P}' \approx^n \hat{Q}'$ and $Q_k = Q_k', P_k = P_k', k = 1, ..., n$, then by **Axiom 4.1 (Sure Thing**

Principle), $\hat{P} \succsim \hat{Q} \Leftrightarrow \hat{P}' \succsim \hat{Q}'$. For $n = 1$, $\hat{P}' = (A, A^c) = \hat{Q}'$, then clearly $\hat{P} \sim \hat{Q}$. For $n > 1$, $\hat{P}' =^{n,1} (A, A^c, \emptyset) =^{n,1} \hat{Q}'$. By **Axiom 4.2 (Weak Comparative Probability)**, we know that $\hat{P}' \succ \hat{Q}' \Leftrightarrow \hat{Q}' \succ \hat{P}'$ which is impossible. Thus, $\hat{Q}' \sim \hat{P}'$, which implies $\hat{Q} \sim \hat{P}$. This completes the proof for Lemma 13.

For Lemma 15, repeat the original argument, we just need to show

$$(A\backslash B, D \cap (B\backslash A), D^c \cap (B\backslash A), R_1, R_2, R_3) \succ (B\backslash A, D \cap (A\backslash B), D^c \cap (A\backslash B), R_1, R_2, R_3)$$

$$\Leftrightarrow (A\backslash B, D \cap (B\backslash A), D^c \cap (B\backslash A), S_1, S_2, S_3) \succ (B\backslash A, D \cap (A\backslash B), D^c \cap (A\backslash B), S_1, S_2, S_3)$$

where $R_1 = A \cap B, R_2 = D\backslash(A \cup B), R_3 = D^c\backslash\{A \cup B\}, S_1 = C \cup (A \cap B), S_2 = D\backslash(A \cup B \cup C)$ and $S_3 = D^c\backslash(A \cup B \cup C)$. This is directly implied by **Axiom 4.1 (Sure Thing Principle)**.

Now we will prove that **Axiom 4** holds. For $m > n, w > t$, suppose that

$$\hat{P} =^{m,n} (A_1, B_1, C) =^{w,t} \hat{R}, \hat{Q} =^{m,n} (A_2, B_2, C) =^{w,t} \hat{S}$$

and

$$\hat{P} \approx^{m+n} \hat{Q}, \hat{R} \approx^{w+t} \hat{S}$$

We need to show $\hat{P} \succeq \hat{Q} \Rightarrow \hat{R} \succeq \hat{S}$. First, by Lemma 16, we can make use of Lemma 7 to construct a nowhere verifiable m-partition of C as $(C_1^1, ..., C_m^1)$ and a nowhere verifiable w-partition of C as $(C_1^2, ..., C_w^2)$. Define $\hat{P}' = (P_1, ..., P_{m+n}, C_1^1, ..., C_m^1)$, $\hat{R}' = (R_1, ..., R_{w+t}, C_1^2, ..., C_w^2)$, $\hat{Q}' = (Q_1, ..., Q_{m+n}, C_1^1, ..., C_m^1)$, $\hat{S}' = (S_1, ..., S_{w+t}, C_1^2, ..., C_w^2)$. Then $\hat{P} \approx^{m+n} \hat{Q}, \hat{P}' \approx^{m+n} \hat{Q}', \hat{R} \approx^{w+t} \hat{S}, \hat{R}' \approx^{w+t} \hat{S}'$ and $Q_k = Q_k', P_k = P_k', k = 1, ..., n+m$, $S_j = S_j', R_j = R_j', j = 1, ..., w+t$. By **Axiom 4.1 (Sure Thing Principle)**, $\hat{P} \succeq \hat{Q} \Leftrightarrow \hat{P}' \succeq \hat{Q}'$ and $\hat{R} \succeq \hat{S} \Leftrightarrow \hat{R}' \succeq \hat{S}'$. Thus, we just need to show $\hat{P}' \succeq \hat{Q}' \Leftrightarrow \hat{R}' \succeq \hat{S}'$.

Moreover, by Lemma 22], $(P_1, ..., P_m)$, $(Q_1, ..., Q_m)$ and $(C_1^1, ..., C_m^1)$ are nowhere verifiable n-partitions of A_1, A_2 and C respectively, $(R_1, ..., R_w)$, $(S_1, ..., S_w)$ and $(C_1^2, ..., C_w^2)$ are nowhere verifiable w-partitions of A_1, A_2 and C respectively. Denote $\hat{P}'' = (P_1 \cup C_1^1, ..., P_m \cup C_m^1, P_{m+1}, ..., P_{m+n}) \sim \hat{P}'$, $\hat{Q}'' = (Q_1 \cup C_1^1, ..., Q_m \cup C_m^1, Q_{m+1}, ..., Q_{m+n}) \sim \hat{Q}'$, $\hat{R}'' = (R_1 \cup C_1^2, ..., R_w \cup C_w^2, R_{w+1}, ..., R_{w+t}) \sim \hat{R}'$ and $\hat{S}'' = (S_1 \cup C_1^2, ..., S_w \cup C_w^2, S_{w+1}, ..., S_{w+t}) \sim \hat{S}'$. Then $\hat{P}'' =^{m,n} (A_1 cupC, B_1, \emptyset) =^{w,t} \hat{R}''$, $\hat{Q}'' =^{m,n} (A_2 cupC, B_2, \emptyset) =^{w,t} \hat{S}''$. By **Axiom 4.2 (Weak Comparative Probability)**, we know $\hat{P}'' \succeq \hat{Q}'' \Leftrightarrow \hat{R}'' \succeq \hat{S}''$, which implies $\hat{P}' \succeq \hat{Q}' \Leftrightarrow \hat{R}' \succeq \hat{S}'$ and $\hat{P} \succeq \hat{Q} \Leftrightarrow \hat{R} \succeq \hat{S}$. This completes the proof. □

Proof of Theorem 3. **Sufficiency**. First **Axiom** 5^{**} implies **Axiom** 5^{*}. This is trivial by definition when we set $\hat{R} = \hat{R}'$. By **Theorem 2**, we know that $\succsim$ is a preference for simplicity with representation $(\Sigma, \mu, \succsim^l)$ and $\succsim^l$ is mixture continuous and monotone. To show that $\succsim^l$ has a expected utility representation, it suffices to show that $\succsim^l$ satisfies independence and vNM continuity as $\succsim^l$ is defined over finite lotteries over posiitve natural numbers.

vNM continuity is implied by mixture continuity. To see why this is true, consider $p \succ^l r \succ^l q$, then $\Lambda_{\succ^l}(r) := \{\lambda \in [0,1] : \lambda p + (1-\lambda)q \succ^l r\}$, $\Lambda_{\prec^l}(r) := \{\lambda \in [0,1] : r \succ^l \lambda p + (1-\lambda)q\}$ are open and $1 \in \Lambda_{\succ^l}(r)$, $0 \in \Lambda_{\prec^l}(r)$. Thus, there exists $0 < a < 1$, $0 < b < 1$ such that $a \in \Lambda_{\succ^l}(r)$, $b \in \Lambda_{\prec^l}(r)$, that is, $ap + (1-a)q \succ^l r \succ^l bp + (1-b)q$. This is exactly vNM continuity.

For independence, suppose $p \succ^l q$ and $\lambda \in (0,1)$. Since μ is convex valued and Lemma 7 holds, for any sequence of lotteries $\{s^i\}_{i=1}^n \subset \mathcal{L}(\mathbb{N})$, for any $\{B_i\}_{i=l}^n \subset \mathcal{T}$ with strictly positive measure, we can always find a problem $\hat{S}$ such that the partition restricted to B_i will induce a conditional probability measure equal to s^i, that is, $\Gamma_\mu^{B_i}(\hat{S}) = s^i$ for all $i = 1, ..., n$. Also, by Lemma 9, $\Gamma_\mu(\hat{S}) = \sum_{i=1}^n \mu(B_i)s^i$. For simplicity, we denote such $\hat{S}$ as $\hat{\Phi}(s^1, B_1; ...; s^n, B_n)$

Lemma 30. $\forall \lambda \in (0,1)$,

$$p \succ^l q \Rightarrow \lambda p + (1-\lambda)q \succ^l q$$

.

Proof of Lemma 30. **Step 1. The lemma holds for $\lambda = 1/n$, $\forall n \in \mathbb{N}_+$.** Construct $\{B_i\}_{i=l}^n \subset \mathcal{T}$ such that $\mu(B_i) = 1/n$ for all i. $p \succ^l q \Leftrightarrow \hat{\Phi}(p, B_1; ...; p, B_n) \succ \hat{\Phi}(q, B_1; ...; q, B_n)$. Suppose by contradiction that $\lambda p + (1-\lambda)q \precsim^l q$, then we have $\hat{\Phi}(q, B_1; ...; q, B_n) \succsim \hat{\Phi}(p, B_1; q, B_2; ...; q, B_n) \sim \hat{\Phi}(q, B_1; ...; q, B_{n-1}; p, B_n)$. By **Axiom 4.1**, $\hat{\Phi}(q, B_1; ...; q, B_{n-1}; p, B_n) \succsim \hat{\Phi}(p, B_1; q, B_2; ...; q, B_{n-1}; p, B_n) \sim \hat{\Phi}(q, B_1; ...; q, B_{n-2}; p, B_{n-1}; p, B_n)$. This implies $q \succsim^l (2/n)p + (1-2/n)q$. Apply **Axiom 4.1** repeatedly, we can show $q \succsim^l p$, which contradicts with our primitive. Thus $(1/n)p + (1-1/n)q \succ^l q$ for all $n \in \mathbb{N}_+$.

Step 2. The lemma holds for $\lambda \in \mathbb{Q} \cap (0,1)$. Consider any element of $\mathbb{Q} \cap (0,1)$ in the form of m/n with $m < n$. By step 1, we know $(1/n)p + (1-1/n)q \succ^l q$. Again, Construct $\{B_i\}_{i=l}^n \subset \mathcal{T}$ such that $\mu(B_i) = 1/n$ for all i. Then $\hat{\Phi}(p, B_1; q, B_2; ...; q, B_n) \succ \hat{\Phi}(q, B_1; ...; q, B_n)$. By **Axiom 4.1**, $\hat{\Phi}(p, B_1; p, B_2; q, B_3 ...; q, B_n) \succ \hat{\Phi}(q, B_1; p, B_2; q, B_3 ...; q, B_n) \sim \hat{\Phi}(p, B_1; q, B_2; ...; q, B_n) \succ \hat{\Phi}(q, B_1; ...; q, B_n)$. That is, (2/

$n)p+(1-2/n)q \succ^l q$. Repeat the argument for $m-1$ times and we derive $(m/n)p+(1-m/n)q \succ^l q$.

Step 3. The lemma holds for $\lambda \in (0,1)$. By Step 2, for any $\lambda \in \mathbb{Q} \cap (0,1)$, $lambdap + (1-\lambda)q \succ^l q$. Since μ is convex valued, we can find B with $\mu(B) = \lambda$. Then $\hat{\Phi}(p, B; q, B^c) \succ \hat{\Phi}(q, B; q, B^c)$. By **Axiom** 5^{**}, we can identify a finite partition $\{A_1, ..., A_s\} \subset \mathcal{T}$. Then there exists some A^* such that $\mu(A^* \cap B) > 0$. Denote $\delta(\lambda) := \mu(A^* \cap B) > 0$. For any $\epsilon \in (0, \delta(\lambda))$, there is a set $C(\epsilon) \subset (A^* \cap B)$ with $\mu(C(\epsilon)) = \epsilon$.

Notice that two partitions are indifferent if they induce the same probability distribution. Then, WLOG, we assume that $\Gamma_\mu^{C(\epsilon)}(\hat{\Phi}(p, B; q, B^c)) = p$, $\Gamma_\mu^{A^* \cap B - C(\epsilon)}(\hat{\Phi}(p, B; q, B^c)) = p$ and $\Gamma_\mu^{A^*-B}(\hat{\Phi}(p, B; q, B^c)) = q$. Construct $\hat{R}$ such that $\Gamma_\mu^{C(\epsilon)}(\hat{R}) = q$, $\Gamma_\mu^{A^* \cap B - C(\epsilon)}(\hat{R}) = p$ and $\Gamma_\mu^{A^*-B}(\hat{R}) = q$. By **Axiom** 5^{**}, we can replace the partition restricted to A^* of $\hat{\Phi}(p, B; q, B^c)$ by $\hat{R}$ without altering the preference order. This implies $(\lambda - \epsilon)p + (1 - \lambda + \epsilon)q \succ^l q$ for all $\epsilon \in (0, \delta(\lambda))$. A similar argument can be used to show that there exists $\delta'(\lambda) > 0$ such that $(\lambda + \epsilon)p + (1 - \lambda + \epsilon)q \succ^l q$ for all $\epsilon \in (0, \delta'(\lambda))$. Since $\mathbb{Q} \cap (0,1)$ is dense in $(0,1)$, this completes the proof for any $\lambda \in (0,1)$, $\lambda p + (1-\lambda)q \succ^l q$. □

Suppose $p \succ^l q$, $r \in \mathcal{L}(\mathbb{N})$ and $\lambda \in (0,1)$. By **Lemma 30**, $\lambda p + (1-\lambda)q \succ^l q$. Now construct $\mu(B) = \lambda$ and $\hat{P} = \hat{\Phi}(p, B; q, B^c)$, $\hat{Q} = \hat{\Phi}(q, B; q, B^c)$ such that $\hat{P} \approx^n \hat{Q}$. Then $\hat{P} \succ \hat{Q}$. Then construct $\hat{R} = \hat{\Phi}(p, B; r, B^c)$ and $\hat{S} = \hat{\Phi}(q, B; r, B^c)$ such that $\hat{R} \approx^n \hat{S}$ and $Q_k = S_k, P_k = R_k, k = 1, ..., n$. By **Axiom 4.1**, $\hat{R} \succ \hat{S}$, that is, $\lambda p + (1-\lambda)r \succ^l \lambda q + (1-\lambda)r$. Thus $\succsim^l$ satisfies independence.

Now appeal to the Mixture Space Theorem, we know that $\succsim^l$ has an expected utility representation as $U(p) = \sum_{n=1}^\infty \theta_n p_n$ for $p \in \mathcal{L}(\mathbb{N})$, $\{\theta_n\}_{n \geq 1} \subset \mathbb{R}$. Since $\succsim^l$ is monotone, θ_n is decreasing in n. Then it remains to be shown that $\{\theta_n\}_{n \geq 1}$ are uniformly bounded. Suppose not, then for any partition $\{A_1, ..., A_n\}$ with $\mu(A_1) > 0$, we can always find N large enough so that θ_N is arbitrarily negative and thus altering partitions on A_1 so that the utility of $\hat{P}^1$ is less than that of $\hat{P}$ by an arbitrarily large amount. This contradicts with **Axiom** 5^{**}. Thus, WLOG, we can restrict $\theta_1 = 1$ and $\theta_n > 0$ for all n.

To conclude, $\succsim$ admits a bounded expected utility representation, that is, $V(\hat{P}) = \sum_{n=1}^\infty \theta_n p_n$ where (i) $p = \Gamma_\mu(\hat{P})$; (ii) $\{\theta_n\}_{n=1}^\infty$ is a decreasing sequence in $[0,1]$ and $\theta_1 = 1$. This completes the proof for sufficiency.

Necessity. Again by **Theorem 2**, we just need to show Axioms 4.1, 4.2 and 5^{**} hold.

For **Axiom 4.1**, $\hat{P} \approx^n \hat{Q}, \hat{R} \approx^n \hat{S}$, $\cup_{i=1}^n P_i, \cup_{i=1}^n Q_i \in \mathcal{T}$ and $Q_k = S_k, P_k = R_k, k = 1, ..., n$. Denote $\cup_{i=1}^n P_i = \cup_{i=1}^n Q_i = A$, then $\Gamma_\mu(\hat{P}) = \mu(B)\Gamma_\mu^B(\hat{P}) + \mu(B^c)\Gamma_\mu^B(\hat{P})$ and similar decompositions hold for $\hat{Q}, \hat{R}, \hat{S}$. Since V has a EU representation and thus is linear, if $\mu(B) = 0$, then clearly $\hat{P} \sim \hat{Q}$ and $\hat{S} \sim \hat{R}$. If $\mu(B) = 0$, then

$$\hat{P} \succsim \hat{Q} \Leftrightarrow \sum_n \theta_n \Gamma_\mu^B(\hat{P})_n \geq \sum_n \theta_n \Gamma_\mu^B(\hat{Q})_n \Leftrightarrow \hat{R} \succsim \hat{S}$$

Thus, **Axiom 4.1** holds.

Axiom 4.2 is a weaker form of **Axiom 4** so that it holds by **Theorem 1**.

For **Axiom** 5^{**}, for any problems $\hat{P} = (P_1, ..., P_n)$, $\hat{Q} = (Q_1, ..., Q_m) \in \mathbb{P}(\Omega)$ such that $\hat{P} \succ \hat{Q}$, choose N large enough so that $1/N < (V(\hat{P}) - V(\hat{Q}))/2$ and $\{A_1, ..., A_N\}$ as a partition of Ω with $\mu(A_i) = 1/N$ for all i. For any $\hat{R}, \hat{R}' \in \mathbb{P}(\Omega)$, $V(\hat{P}^i) \geq V(\hat{P}) - 1/N$, $V(\hat{Q}^i) \leq V(\hat{Q}) + 1/N$. Then, $V(\hat{P}^i) - V(\hat{Q}^i) \geq V(\hat{P}) - V(\hat{Q}) - 2/N > 0$. Thus we have $\hat{P}' \succ \hat{Q}'$ and **Axiom** 5^{**} is satisfied. This completes the proof for necessity.

□

9.4 Proofs in Section 5

Proof of Lemma 3. The result directly follows from the definitions. Given $\Gamma_{\mu^1}(\hat{P}^1) = \Gamma_{\mu^2}(\hat{P}^2)$, $\Gamma_{\mu^1}(\hat{Q}^1) = \Gamma_{\mu^2}(\hat{Q}^2)$, $\hat{P}^1 \succsim_1 \hat{Q}^1 \Leftrightarrow \hat{P}^2 \succsim_2 \hat{Q}^2$ if and only if $\Gamma_{\mu^1}(\hat{P}^1) \succsim_1^l \Gamma_{\mu^1}(\hat{Q}^1) \Leftrightarrow \Gamma_{\mu^1}(\hat{P}^1) \succsim_2^l \Gamma_{\mu^1}(\hat{Q}^1)$. Recall that Γ_{μ^1} is onto, this is equivalent to $\succsim_1^l = \succsim_2^l$. □

Proof of Lemma 4. The sufficiency is given in the main text. Suppose the necessity dose not hold, then there exists some $\hat{P}$ such that $\Gamma_{\mu^2}(\hat{P})$ does not weakly FOSD $\Gamma_{\mu^1}(\hat{P})$. Either $\Gamma_{\mu^1}(\hat{P})$ FOSD $\Gamma_{\mu^2}(\hat{P})$ or they are not ordered by the binary relation weak FOSD. In the former case, $\Gamma_{\mu^2}(\hat{P}) \succ^l \Gamma_{\mu^1}(\hat{P})$ for any $\succsim^l$, which is a contradiction. In the latter case, recall that there is no restriction on ranking of lotteries unordered under weak FOSD in the definition, then we can always find an attitude $\succsim^l$ such that $\Gamma_{\mu^2}(\hat{P}) \succ^l \Gamma_{\mu^1}(\hat{P})$ and there is a contradiction again. Thus the necessity holds. □

Proof of Theorem 4. To begin with, by equation 5 in the proof of Lemma 1, we can simplify the expression of the cumulative distribution function of each $\hat{p}$ given perception (Σ, μ) as

$$F_k(\hat{p}) = \mu(\cup_{A \in \mathcal{D}_k(\hat{P})} E[A]) \tag{6}$$

where $\mathcal{D}_k$ is defined by $A \in \mathcal{D}_n(\hat{P}) \Leftrightarrow \exists\{P_{k_1}, ..., P_{k_n}\} \in \mathcal{C}_n(\hat{P})$ such that $A = \cup_{i=1}^n P_{k_n}$, and $E[\cdot] : 2^\Omega \to \Sigma$ is an operator such that $\forall A \subset \Omega$, $\mu_*(A) = \mu(E[A])$ and $\forall A \subseteq B \Rightarrow E[A] \subseteq E[B]$.

For sufficiency, suppose that (Σ_1, μ_1) is an extension of (Σ_2, μ_2). We use superscript 1 and 2 to distinguish the two different perceptions. First, for any A and given operators E^1 and E^2, we have $E^2[A] \in \Sigma_2 \subseteq \Sigma_1$ and $\mu_1(E^2[A]) = \mu_2^*(A) \leq \mu_1(E^1[A]) = \mu_1^*(A)$. Suppose by contradiction that $\mu_1(E^2[A] - E^1[A]) > 0$, then $\mu_1(E^1[A] \cup E^2[A]) > \mu_1(E^1[A]) = \mu_1^*(A)$. Also $E^1[A] \cup E^2[A] \subseteq A$ and $E^1[A] \cup E^2[A] \in \Sigma_1$, which contradicts with the definition of E^1. Thus, $\mu_1(E^2[A] - E^1[A]) = 0$ and by completeness of μ, without loss of generality, we can assume $E^2[A] \subseteq E^1[A]$ for all A. Then for any $\hat{P} \in \mathbb{P}(\Omega)$ and any k:

$$\begin{aligned} F_k(\Gamma_{\mu^2}(\hat{P})) = \mu_2(\cup_{A \in \mathcal{D}_k(\hat{P})} E^2[A]) &= \mu_1(\cup_{A \in \mathcal{D}_k(\hat{P})} E^2[A]) \\ &\leq \mu_1(\cup_{A \in \mathcal{D}_k(\hat{P})} E^1[A]) = F_k(\Gamma_{\mu^1}(\hat{P})) \end{aligned}$$

Thus, $\Gamma_{\mu^2}(\hat{P})$ weakly FOSD $\Gamma_{\mu^1}(\hat{P})$, for any $\hat{P} \in \mathbb{P}(\Omega)$. By Lemma 4, (Σ_1, μ_1) is more accurate than (Σ_2, μ_2).

Now we prove the necessity. Suppose (Σ_1, μ_1) is more accurate than (Σ_2, μ_2), then $\Gamma_{\mu^2}(\hat{P})$ weakly FOSD $\Gamma_{\mu^1}(\hat{P})$, for any $\hat{P} \in \mathbb{P}(\Omega)$. First, for any $A \in \Sigma_1 \cap \Sigma_2$, choose $B_1 \subset A^c$ diffuse within A^c under (Σ_1, μ_1) and denote $\hat{P}_1 = (A, B_1, A^c - B_1)$. Then $\Gamma_{\mu^1}(\hat{P}_1) = (\mu_1(A), 1 - \mu_1(A), 0)$ while $\Gamma_{\mu^2}(\hat{P}_1) = (\mu_2(A \cup C), 1 - \mu_2(A \cup C), 0)$ for some $C \subset A^c$ and $C \in \Sigma_2$. Since $\Gamma_{\mu^2}(\hat{P}_1)$ weakly FOSD $\Gamma_{\mu^1}(\hat{P}_1)$, we know $\mu_1(A) \geq \mu_2(A \cup C) \geq \mu_2(A)$. Similarly, $A^c \in \Sigma_1 \cap \Sigma_2$ and we can follow the same argument to get $\mu_1(A^c) \geq \mu_2(A^c)$. Since $\mu_1(A) + \mu_1(A^c) = 1 = \mu_2(A) + \mu_2(A^c)$, we know $\mu_1(A) = \mu_2(A)$ for all $A \in \Sigma_1 \cap \Sigma_2$.

Then we suppose by contradiction that there exists $A^* \in \Sigma_2 - \Sigma_1$. By definition, $E^1[A^*] \in \Sigma_1$. We can again choose $B_1 \subset E^1[A^*]^c$ diffuse within $E^1[A^*]^c$ under (Σ_1, μ_1). Denote $\hat{P}_2 = (A^*, A^{*c} \cap B_1, A^{*c} - B_1)$. Since B_1 is diffuse within $E^1[A^*]^c$, the inner measures under μ_1 are $\mu_1^*(A^{*c} \cap B_1) = \mu_1^*(A^{*c} - B_1) = 0$ and then $\Gamma_{\mu^1}(\hat{P}_2)(1) = \mu_1(E^1[A^*])$. By comparison, $\Gamma_{\mu^2}(\hat{P}_2)(1) \geq \mu_2(A^*)$. As $\Gamma_{\mu^2}(\hat{P}_2)$ weakly FOSD $\Gamma_{\mu^1}(\hat{P}_2)$, we know $\mu_1(E^1[A^*]) \geq \mu_2(A^*)$.

On the other hand, $A^* \in \Sigma_2 - \Sigma_1$ implies that $A^{*c} \in \Sigma_2 - \Sigma_1$ and we can repeat the same argument to get $\mu_1(E^1[A^{*c}]) \geq \mu_2(A^{*c})$. This means $\mu_1(E^1[A^*]) + \mu_1(E^1[A^{*c}]) = 1$. Since μ_1 is complete, $\mu_1(A^* - E^1[A^*]) = 0$ and $A^* - E^1[A^*] \in \Sigma_1$, which further implies $A^* \in \Sigma_1$ and leads to a contradiction.

To sum up, we have shown that $\Sigma_2 \subseteq \Sigma_1$ and $\mu_1(A) = \mu_2(A)$ for all $A \in \Sigma_1 \cap \Sigma_2 = \Sigma_2$. Thus (Σ_1, μ_1) is an extension of (Σ_2, μ_2) and this completes the proof for necessity. $\square$

Proof of Corollary 1. By Lemma 4 and its proof, it suffices to show that if $\Gamma_{\mu^1}(\hat{P})$ and $\Gamma_{\mu^2}(\hat{P})$ are not ordered by the binary relation weak FOSD, then there exists an attitude $\succsim^l$ which admits a bounded EU representation such that $\Gamma_{\mu^2}(\hat{P}) \succ^l \Gamma_{\mu^1}(\hat{P})$.

Denote $\Gamma_{\mu^1}(\hat{P}) = \hat{p^1}$, $\Gamma_{\mu^2}(\hat{P}) = \hat{p^2}$. Suppose that $\hat{p^1}$ and $\hat{p^2}$ are not ordered by weak FOSD, then there exists k such that $F_i(\hat{p^2}) \leq F_i(\hat{p^1})$ for all $i < k$ and $F_k(\hat{p^2}) - F_k(\hat{p^1}) = \delta > 0$. Choose $\theta : \mathbb{N} \to [0, 1]$ which is strictly decreasing and $\theta(k) = 1 - \epsilon$, $\theta(k+1) = \epsilon$ and $\succsim^l$ has a bounded EU representation $V : \mathcal{L}(\mathbb{N}) \to \mathbb{R}$ with $V(\hat{p}) = \sum_{n=1}^{+\infty} p_n \theta(n)$. Then we can always choose ϵ small enough such that

$$\sum_{n=1}^{k} p_n^2 \theta(n) - \sum_{n=1}^{k} p_n^1 \theta(n) > \frac{\delta}{2}, \quad \sum_{n=k+1}^{+\infty} p_n^1 \theta(n) - \sum_{n=k+1}^{+\infty} p_n^2 \theta(n) < \frac{\delta}{2}$$

This implies $V(\hat{p^2}) > V(\hat{p^1})$ and thus $\Gamma_{\mu^2}(\hat{P}) \succ^l \Gamma_{\mu^1}(\hat{P})$. $\square$

Proof of Lemma 6. (1) For any μ and A, $\Gamma_\mu((A, A^c)) = (\mu_*(A) + \mu_*(A^c), 1 - \mu_*(A) - \mu_*(A^c))$. For any preference for simplicity $(\Sigma, \mu, \succsim^l)$ with prior μ, $\succsim^l$ is monotone. Then

$$\begin{aligned}(B, B^c) \succ (A, A^c) &\Leftrightarrow (\mu_*(B) + \mu_*(B^c), 1 - \mu_*(B) - \mu_*(B^c)) \succ^l (\mu_*(A) + \mu_*(A^c), 1 - \mu_*(A) - \mu_*(A^c)) \\ &\Leftrightarrow \mu_*(A) + \mu_*(A^c) < \mu_*(B) + \mu_*(B^c)\end{aligned}$$

Thus, A is more difficult than B if and only if $\mu_*(A) + \mu_*(A^c) \leq \mu_*(B) + \mu_*(B^c)$.
(2) If A is more uncertain than B, then $\mu_*(A) < \mu_*(B)$ and $\mu_*(A^c) < \mu_*(B)$, which further implies $\mu_*(A) + \mu_*(A^c) \leq \mu_*(B) + \mu_*(B^c)$. By the first part, the proof is completed. $\square$

References

Borgers, T. and J. Li (2017): *Strategically Simple Mechanism.*

Ergin, H. and T. Sarver (2010): "A Unique Costly Contemplation Representation," *Econometrica*, 78, 1285–1339.

Gilboa, I. and D. Schmeidler (1989): "Maxmin expected utility with non-unique prior," *Journal of Mathematical Economics*, 18, 141–153.

Gul, F. and W. Pesendorfer (2014): "Expected Uncertain Utility Theory," *Econometrica*, 82, 1–39.

Hart, O. and J. Moore (1999): "Foundations of Incomplete Contracts," *The Review of Economic Studies*, 66, 115–138.

Ke, S. and Q. Zhang (2018): "Randomization and Ambiguity Aversion," *mimeo.*

Li, S. (2017): "Obviously Strategy-Proof Mechanisms," *American Economic Review*, 107, 3257–3287.

Machina, M. J. and D. Schmeidler (1992): "A more robust definition of subjective probability," *Econometrica*, 60, 745–780.

Ortoleva, P. (2013): "The Price of Flexibility: Towards a Theory of Thinking Aversion," *Journal of Economic Theory*, 148, 903 – 934.

Saito, K. (2015): "Preferences for Flexibility and Randomization under Uncertainty," *American Economic Review*, 105, 1246–1271.

Savage, L. J. (1954): *The Foundations of Statistics*, New York: Wiley.

Schmeidler, D. (1989): "Subjective Probability and Expected Utility without Additivity," *Econometrica*, 57, 571–587.

Segal, I. (1999): "Complexity and Renegotiation: A Foundation for Incomplete Contracts," *The Review of Economic Studies*, 66, 57–82.